Supplementary Material to "Optimal response-free cluster subsampling for longitudinal data under measurement constrains"

Junhao Shan[1], Lei Wang*[1] and Haiying Wang[2]

[1] *School of Statistics and Data Science, KLMDASR, LEBPS and LPMC, Nankai University, China*

[2] *Department of Statistics, University of Connecticut, U.S.A.*

S1 Unbalanced cluster-size scenario

To adapt the proposed method to unbalanced data, we apply a transformation matrix to each cluster (Zhou and Qu, 2012). Specifically, we define the largest cluster with size m, representing the maximum number of possible observation time points, and assume that fully observed clusters contain m measurements. For the i-th cluster, we define $\boldsymbol{A}_{0i}$ as the diagonal matrix of order m_i whose j-th diagonal entry is $\{\ddot{\psi}(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{x}_{ij})\}^{1/2}$. In addition, we construct a $m \times m_i$ transformation matrix $\boldsymbol{T}_i$ by removing from the $m \times m$ identity matrix the columns corresponding to time points that are not observed in the i-th cluster. Given consistent estimators $\{\tilde{\boldsymbol{A}}_i\}_{i=1}^n$ of $\{\boldsymbol{A}_{0i}\}_{i=1}^n$ and a working correlation matrix $\tilde{\boldsymbol{R}}$, the response-free weighted cluster

---

Corresponding author: Lei Wang, School of Statistics and Data Science, KLMDASR, LEBPS and LPMC, Nankai University, Tianjin, 300071, China. E-mail: lwangstat@nankai.edu.cn.

subsample estimator can be solved from:

$$\boldsymbol{S}_w^*(\boldsymbol{\beta};\tilde{\boldsymbol{R}}) = \frac{1}{n}\sum_{i\in\mathcal{I}}\frac{1}{\pi_i^*}\{\boldsymbol{T}_i^*\boldsymbol{X}_i^*\}^{\mathrm{T}}\{\tilde{\mathcal{A}}_i^*\}^{1/2}\tilde{\boldsymbol{R}}^{-1}\{\tilde{\mathcal{A}}_i^*\}^{-1/2}\boldsymbol{T}_i^*\{\boldsymbol{Y}_i^* - \boldsymbol{\mu}_i^*(\boldsymbol{\beta})\} = \boldsymbol{0},$$

where $\tilde{\mathcal{A}}_i^* = \boldsymbol{T}_i^*\tilde{\boldsymbol{A}}_i^*\boldsymbol{T}_i^{*\mathrm{T}}$. Since the zero values specified in and ensure that contributions from unobserved time points do not affect the estimating equation, we can set the corresponding variance components in $\tilde{\mathcal{A}}_i^*$ to zero for convenience.

For the initial estimators in step (i), denote the j-th diagonal element of $\tilde{\boldsymbol{A}}_i^{\mathrm{p}}$ as $(\tilde{\boldsymbol{A}}_i^{\mathrm{p}})_{jj}$ and it can be estimated by the sample variance of the residuals $\{\tilde{v}_{ij}^{*\mathrm{p}}\}_{i\in\mathcal{I}_j}$ for $i = 1, \dots, n$, where $\mathcal{I}_j$ denotes the set of clusters with available observations at time j. Under the IND working correlation structure, $\tilde{\boldsymbol{R}}^{\mathrm{p}} = \boldsymbol{I}_m$; under the CS and AR working correlation structures, all diagonal elements of $\tilde{\boldsymbol{R}}^{\mathrm{p}}$ are equal to 1, while the (j, j')-th off-diagonal elements of $\tilde{\boldsymbol{R}}^{\mathrm{p}}$ are equal to

$$\frac{1}{r_{\mathrm{p}}}\sum_{i=1}^{r_{\mathrm{p}}}\frac{1}{m_i(m_i-1)}\sum_{j\neq j'}\frac{\tilde{v}_{ij}^{*\mathrm{p}}\tilde{v}_{ij'}^{*\mathrm{p}}}{\sqrt{(\tilde{\boldsymbol{A}}_i^{\mathrm{p}})_{jj}(\tilde{\boldsymbol{A}}_i^{\mathrm{p}})_{j'j'}}},\ \Big(\frac{1}{r_{\mathrm{p}}}\sum_{i=1}^{r_{\mathrm{p}}}\frac{1}{2(m_i-1)}\sum_{|\iota-\iota'|=1}\frac{\tilde{v}_{i\iota}^{*\mathrm{p}}\tilde{v}_{i\iota'}^{*\mathrm{p}}}{\sqrt{(\tilde{\boldsymbol{A}}_i^{\mathrm{p}})_{\iota\iota}(\tilde{\boldsymbol{A}}_i^{\mathrm{p}})_{\iota'\iota'}}}\Big)^{|j-j'|},$$

respectively.

Under the unbalanced scenario, our proposed two-step subsampling estimator $\hat{\boldsymbol{\beta}}_w^{\boldsymbol{D}}$ is the solution to

$$\frac{1}{n}\sum_{i\in\mathcal{I}_{\boldsymbol{D}}}\frac{1}{\tilde{\pi}_i^{*\boldsymbol{D}}\wedge 1}(\boldsymbol{T}_i^{*\boldsymbol{D}}\boldsymbol{X}_i^{*\boldsymbol{D}})^{\mathrm{T}}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\boldsymbol{\beta})\}^{1/2}\tilde{\boldsymbol{R}}_{\mathrm{p}}^{-1}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\boldsymbol{\beta})\}^{-1/2}\boldsymbol{T}_i^{*\boldsymbol{D}}\{\boldsymbol{Y}_i^{*\boldsymbol{D}}-\boldsymbol{\mu}_i^{*\boldsymbol{D}}(\boldsymbol{\beta})\} = \boldsymbol{0}.$$

S2 Technical proofs

Proof of Theorem 1. (i) For consistency, the first step is to show that $\|\hat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}_0\| = o_P(1)$. For any $\boldsymbol{\beta} \in \mathbb{R}^p$, denote

$$\begin{aligned}
\boldsymbol{S}_w^*(\boldsymbol{\beta}; \boldsymbol{R}) &= \frac{1}{n}\sum_{i=1}^n \frac{\eta_i}{\pi_i} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}) \boldsymbol{R}^{-1} \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}) \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\}, \\
\boldsymbol{S}_w(\boldsymbol{\beta}; \boldsymbol{R}) &= \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}) \boldsymbol{R}^{-1} \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}) \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\}.
\end{aligned}$$

where $\boldsymbol{R}$ = the limit of $\tilde{\boldsymbol{R}}$, η_i is the indicator function taking value 1 if the i-th cluster $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$ is selected and 0 otherwise. By the law of iterated expectation,

$$\begin{aligned}
E[\boldsymbol{S}_w^*(\boldsymbol{\beta}; \boldsymbol{R})] &= E\{E[\boldsymbol{S}_w^*(\boldsymbol{\beta}; \boldsymbol{R}) | \mathcal{F}_n]\} \\
&= E\Big[\frac{1}{n}\sum_{i=1}^n \frac{E(\eta_i|\mathcal{F}_n)}{\pi_i} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}) \boldsymbol{R}^{-1} \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}) \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\}\Big] \\
&= E\Big[\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}) \boldsymbol{R}^{-1} \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}) \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\}\Big] \\
&= E[\boldsymbol{S}_w(\boldsymbol{\beta}; \boldsymbol{R})].
\end{aligned}$$

By the variance decomposition formula,

$$Var[\boldsymbol{S}_w^*(\boldsymbol{\beta}; \boldsymbol{R})] = E\{Var[\boldsymbol{S}_w^*(\boldsymbol{\beta}; \boldsymbol{R}) | \mathcal{F}_n]\} + Var\{E[\boldsymbol{S}_w^*(\boldsymbol{\beta}; \boldsymbol{R}) | \mathcal{F}_n]\} = O(r^{-1})$$

due to the fact that

$$\begin{aligned}
&E\{Var[\boldsymbol{S}_w^*(\boldsymbol{\beta}; \boldsymbol{R}) | \mathcal{F}_n]\} \\
=&E\Big[\frac{1}{n^2}\sum_{i=1}^n \frac{Var(\eta_i|\mathcal{F}_n)}{\pi_i^2} [\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}) \boldsymbol{R}^{-1} \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}) \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\}]^{\otimes 2}\Big]
\end{aligned}$$

$$
\begin{aligned}
&\leq E\Big[\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{\pi_i}[\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta})\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta})\}]^{\otimes 2}\Big]\\
&\leq E\Big[\Big(\max_{1\leq i\leq n}\frac{1}{n\pi_i}\Big)\frac{1}{n}\sum_{i=1}^{n}[\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta})\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta})\}]^{\otimes 2}\Big]\\
&=O(r^{-1}),
\end{aligned}
$$

where the last equality holds under assumption (A.3), and

$$
\begin{aligned}
&Var\{E[\boldsymbol{S}_w^*(\boldsymbol{\beta};\boldsymbol{R})|\mathcal{F}_n]\}\\
=&Var\Big[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta})\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta})\}\Big]\\
\leq&E\Big[\frac{1}{n^2}\sum_{i=1}^{n}[\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta})\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta})\}]^{\otimes 2}\Big]\\
=&O(n^{-1}).
\end{aligned}
$$

Thus, by Chebyshev's inequality,

$$\boldsymbol{S}_w^*(\boldsymbol{\beta};\boldsymbol{R})=E[\boldsymbol{S}_w(\boldsymbol{\beta};\boldsymbol{R})]+o_P(1).$$

By the consistency of $\tilde{\boldsymbol{R}}$ to $\boldsymbol{R}$,

$$\boldsymbol{S}_w^*(\boldsymbol{\beta};\tilde{\boldsymbol{R}})=\boldsymbol{S}_w^*(\boldsymbol{\beta};\boldsymbol{R})+o_P(1),$$

and therefore $\boldsymbol{S}_w^*(\boldsymbol{\beta};\tilde{\boldsymbol{R}})=E[\boldsymbol{S}_w(\boldsymbol{\beta};\boldsymbol{R})]+o_P(1)$. Applying Theorem 5.9 of van der Vaart (1998), we obtain $\|\hat{\boldsymbol{\beta}}_w-\boldsymbol{\beta}_0\|=o_P(1)$.

In the second step, using Taylor's expansion,

$$\mathbf{0}=\boldsymbol{S}_w^*(\hat{\boldsymbol{\beta}}_w;\hat{\boldsymbol{\beta}}_w,\tilde{\boldsymbol{R}})$$

$$= \boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \hat{\boldsymbol{\beta}}_w, \tilde{\boldsymbol{R}}) + \nabla_{\boldsymbol{\beta}} \boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \hat{\boldsymbol{\beta}}_w, \tilde{\boldsymbol{R}})(\hat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}_0) + O_P(\|\hat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}_0\|^2),$$

where

$$\boldsymbol{S}_w^*(\boldsymbol{\beta}; \hat{\boldsymbol{\beta}}_w, \tilde{\boldsymbol{R}}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\eta_i}{\pi_i} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\hat{\boldsymbol{\beta}}_w) \tilde{\boldsymbol{R}}^{-1} \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_w)\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\},$$

$$\nabla_{\boldsymbol{\beta}} \boldsymbol{S}_w^*(\boldsymbol{\beta}; \hat{\boldsymbol{\beta}}_w, \tilde{\boldsymbol{R}}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\eta_i}{\pi_i} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\hat{\boldsymbol{\beta}}_w) \tilde{\boldsymbol{R}}^{-1} \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_w) \boldsymbol{A}_i(\boldsymbol{\beta}) \boldsymbol{X}_i.$$

To obtain $\|\hat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}_0\| = O_P(r^{-1/2})$, we will show $\boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \hat{\boldsymbol{\beta}}_w, \tilde{\boldsymbol{R}}) = O_P(r^{-1/2})$ and $\nabla_{\boldsymbol{\beta}} \boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \hat{\boldsymbol{\beta}}_w, \tilde{\boldsymbol{R}}) - \boldsymbol{\Phi} = o_P(1)$ in the following proof, respectively.

(a) To show $\boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \hat{\boldsymbol{\beta}}_w, \tilde{\boldsymbol{R}}) = O_P(r^{-1/2})$, denote

$$\boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \boldsymbol{\beta}_0, \boldsymbol{R}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\eta_i}{\pi_i} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0) \boldsymbol{R}^{-1} \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}.$$

It is simple to obtain that

$$\begin{aligned}
E[\boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \boldsymbol{\beta}_0, \boldsymbol{R})] =& E\{E[\boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \boldsymbol{\beta}_0, \boldsymbol{R})|\mathcal{F}_n]\} = \boldsymbol{0}, \\
Var[\boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \boldsymbol{\beta}_0, \boldsymbol{R})] =& E\{Var[\boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \boldsymbol{\beta}_0, \boldsymbol{R})|\mathcal{F}_n] + Var\{E[\boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \boldsymbol{\beta}_0, \boldsymbol{R})|\mathcal{F}_n]\} \\
=& E\Big[\frac{1}{n^2}\sum_{i=1}^{n} \frac{1}{\pi_i}[\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0) \boldsymbol{R}^{-1} \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}]^{\otimes 2}\Big] \\
\leq& E\Big[\Big(\max_{1\leq i\leq n} \frac{1}{n\pi_i}\Big)\frac{1}{n}\sum_{i=1}^{n} [\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0) \boldsymbol{R}^{-1} \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}]^{\otimes 2}\Big] \\
=& O(r^{-1}).
\end{aligned}$$

Thus $\|\boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \boldsymbol{\beta}_0, \boldsymbol{R})\| = O_P(r^{-1/2})$. Next, notice that

$$\boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \hat{\boldsymbol{\beta}}_w, \tilde{\boldsymbol{R}}) - \boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \boldsymbol{\beta}_0, \boldsymbol{R})$$

$$
\begin{aligned}
=&\frac{1}{n}\sum_{i=1}^{n}\frac{\eta_i}{\pi_i}\boldsymbol{X}_i^{\mathrm{T}}\{\boldsymbol{A}_i^{1/2}(\hat{\boldsymbol{\beta}}_w)\tilde{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_w)-\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\}\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}\\
=&o_P(r^{-1/2}).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\|\boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\hat{\boldsymbol{\beta}}_w,\tilde{\boldsymbol{R}})\| &= \|\boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\boldsymbol{\beta}_0,\boldsymbol{R})+\boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\hat{\boldsymbol{\beta}}_w,\tilde{\boldsymbol{R}})-\boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\boldsymbol{\beta}_0,\boldsymbol{R})\|\\
&\le \|\boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\boldsymbol{\beta}_0,\boldsymbol{R})\|+\|\boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\hat{\boldsymbol{\beta}}_w,\tilde{\boldsymbol{R}})-\boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\boldsymbol{\beta}_0,\boldsymbol{R})\|\\
&= O_P(r^{-1/2}).
\end{aligned}
$$

(b) To show $\nabla_{\boldsymbol{\beta}}\boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\hat{\boldsymbol{\beta}}_w,\tilde{\boldsymbol{R}})-\boldsymbol{\Phi}=o_P(1)$, denote

$$
\nabla_{\boldsymbol{\beta}}\boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\boldsymbol{\beta}_0,\boldsymbol{R})=\frac{1}{n}\sum_{i=1}^{n}\frac{\eta_i}{\pi_i}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i.
$$

It is simple to obtain

$$
E[\nabla_{\boldsymbol{\beta}}\boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\boldsymbol{\beta}_0,\boldsymbol{R})]=E[\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i]=\boldsymbol{\Phi}.
$$

Denote $[\nabla_{\boldsymbol{\beta}}\boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\boldsymbol{\beta}_0,\boldsymbol{R})]_{l_1,l_2}$ and $[\boldsymbol{\Phi}]_{l_1,l_2}$ as the (l_1,l_2)-th component of $\nabla_{\boldsymbol{\beta}}\boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\boldsymbol{\beta}_0,\boldsymbol{R})$ and $\boldsymbol{\Phi}$ for any $1\le l_1\le l_2\le p$, respectively, then

$$
\begin{aligned}
&Var\{[\nabla_{\boldsymbol{\beta}}\boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\boldsymbol{\beta}_0,\boldsymbol{R})]_{l_1,l_2}\}\\
=&E(Var\{[\nabla_{\boldsymbol{\beta}}\boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\boldsymbol{\beta}_0,\boldsymbol{R})]_{l_1,l_2}|\mathcal{F}_n\})+Var(E\{[\nabla_{\boldsymbol{\beta}}\boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\boldsymbol{\beta}_0,\boldsymbol{R})]_{l_1,l_2}|\mathcal{F}_n\})\\
\le&E\Big[\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{\pi_i}[\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i]_{l_1,l_2}^2\Big]\\
=&O(r^{-1}).
\end{aligned}
$$

By Chebyshev's inequality, $[\nabla_{\boldsymbol{\beta}} \boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\boldsymbol{\beta}_0,\boldsymbol{R})]_{l_1,l_2} - [\boldsymbol{\Phi}]_{l_1,l_2} = O_P(r^{-1/2})$, and it holds for any pair of (l_1,l_2) for $1 \le l_1 \le l_2 \le p$. Thus, we conclude that $\nabla_{\boldsymbol{\beta}} \boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\boldsymbol{\beta}_0,\boldsymbol{R}) - \boldsymbol{\Phi} = O_P(r^{-1/2})$. Next, notice that

$$\begin{aligned}
&\nabla_{\boldsymbol{\beta}} \boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\hat{\boldsymbol{\beta}}_w,\tilde{\boldsymbol{R}}) - \nabla_{\boldsymbol{\beta}} \boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\boldsymbol{\beta}_0,\boldsymbol{R})\\
=&\frac{1}{n}\sum_{i=1}^{n}\frac{\eta_i}{\pi_i}\boldsymbol{X}_i^{\mathrm{T}}\{\boldsymbol{A}_i^{1/2}(\hat{\boldsymbol{\beta}}_w)\tilde{\boldsymbol{R}}^{-1}\boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}_w) - \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\}\boldsymbol{A}_i(\boldsymbol{\beta}_0)\boldsymbol{X}_i\\
=&o_P(1).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\|\nabla_{\boldsymbol{\beta}} \boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\hat{\boldsymbol{\beta}}_w,\tilde{\boldsymbol{R}}) - \boldsymbol{\Phi}\|\\
\le&\|\nabla_{\boldsymbol{\beta}} \boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\hat{\boldsymbol{\beta}}_w,\tilde{\boldsymbol{R}}) - \nabla_{\boldsymbol{\beta}} \boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\boldsymbol{\beta}_0,\boldsymbol{R})\| + \|\nabla_{\boldsymbol{\beta}} \boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\boldsymbol{\beta}_0,\boldsymbol{R}) - \boldsymbol{\Phi}\|\\
=&o_P(1).
\end{aligned}$$

Hence, $\|\boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\hat{\boldsymbol{\beta}}_w,\tilde{\boldsymbol{R}})\| = O_P(r^{-1/2})$, $\|\nabla_{\boldsymbol{\beta}} \boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\hat{\boldsymbol{\beta}}_w,\tilde{\boldsymbol{R}})(\hat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}_0)\| = O_P(\|\hat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}_0\|)$, and we conclude $\|\hat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}_0\| = O_P(r^{-1/2})$ from the Taylor's expansion.

(ii) Now we prove the asymptotic normality. Denote

$$\boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\boldsymbol{\beta}_0,\boldsymbol{R}) = \sum_{i=1}^{n}\frac{\eta_i}{n\pi_i}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\} =: \sum_{i=1}^{n}\boldsymbol{\xi}_i.$$

Below we check the conditions of Lindeberg-Feller central limit theorem. For $\forall\epsilon > 0$,

$$\begin{aligned}
&\sum_{i=1}^{n}E[\|\boldsymbol{\xi}_i\|^2 I(\|\boldsymbol{\xi}_i\| \ge \epsilon)]\\
\le&\frac{1}{\epsilon}\sum_{i=1}^{n}E(\|\boldsymbol{\xi}_i\|^3)
\end{aligned}$$

$$
\begin{aligned}
&=\frac{1}{\epsilon}\sum_{i=1}^{n} E[E(\|\boldsymbol{\xi}_i\|^3|\mathcal{F}_n)]\\
&=\frac{1}{\epsilon}\sum_{i=1}^{n} E\Big[\frac{1}{n^3\pi_i^2}\|\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}\|^3\Big]\\
&\leq\frac{1}{\epsilon}E\Big[\Big(\max_{1\leq i\leq n}\frac{1}{n\pi_i}\Big)^2\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}\|^3\Big]\\
&=o(1).
\end{aligned}
$$

Thus the condition of Lindeberg-Feller central limit theorem is satisfied. Note that

$$
\begin{aligned}
&\sum_{i=1}^{n} E(\boldsymbol{\xi}_i)=\sum_{i=1}^{n} E[E(\boldsymbol{\xi}_i|\mathcal{F}_n)]=E[\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}]=\boldsymbol{0},\\
&\sum_{i=1}^{n} Var(\boldsymbol{\xi}_i)\\
=&\sum_{i=1}^{n} Var[E(\boldsymbol{\xi}_i|\mathcal{F}_n)]+\sum_{i=1}^{n} E[Var(\boldsymbol{\xi}_i|\mathcal{F}_n)]\\
=&E\Big[\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{\pi_i}[\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}]^{\otimes 2}\Big]\\
=&E\Big(E\Big[\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{\pi_i}[\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}]^{\otimes 2}\Big|\{\boldsymbol{X}_i\}_{i=1}^{n}\Big]\Big)\\
=&E\Big(\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{\pi_i}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}E\{[\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}]^{\otimes 2}|\{\boldsymbol{X}_i\}_{i=1}^{n}\}\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i\Big)\\
=&E\Big[\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{\pi_i}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i\Big]\\
=&\boldsymbol{\Xi}_\pi.
\end{aligned}
$$

By the Lindeberg-Feller central limit theorem, we obtain

$$\boldsymbol{\Xi}_\pi^{-1/2}\boldsymbol{S}_w^*(\boldsymbol{\beta}_0;\boldsymbol{\beta}_0,\boldsymbol{R})\xrightarrow{d}N(\boldsymbol{0},\boldsymbol{I}_p).$$

According to the Taylor's expansion,

$$\hat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}_0 = -[\nabla_{\boldsymbol{\beta}} \boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \hat{\boldsymbol{\beta}}_w, \tilde{\boldsymbol{R}})]^{-1} \boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \hat{\boldsymbol{\beta}}_w, \tilde{\boldsymbol{R}}) + O_P(\|\hat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}_0\|^2).$$

Note that

$$-[\nabla_{\boldsymbol{\beta}} \boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \hat{\boldsymbol{\beta}}_w, \tilde{\boldsymbol{R}})]^{-1} = \boldsymbol{\Phi}^{-1} - [\nabla_{\boldsymbol{\beta}} \boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \hat{\boldsymbol{\beta}}_w, \tilde{\boldsymbol{R}})]^{-1} - \boldsymbol{\Phi}^{-1},$$

$$\boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \hat{\boldsymbol{\beta}}_w, \tilde{\boldsymbol{R}}) = \boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \hat{\boldsymbol{\beta}}_w, \tilde{\boldsymbol{R}}) - \boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \boldsymbol{\beta}_0, \boldsymbol{R}) + \boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \boldsymbol{\beta}_0, \boldsymbol{R}),$$

and we have shown that

$$\boldsymbol{\Phi}^{-1} - [\nabla_{\boldsymbol{\beta}} \boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \hat{\boldsymbol{\beta}}_w, \tilde{\boldsymbol{R}})]^{-1} = o_P(1), \ \boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \hat{\boldsymbol{\beta}}_w, \tilde{\boldsymbol{R}}) - \boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \boldsymbol{\beta}_0, \boldsymbol{R}) = o_P(r^{-1/2}).$$

Therefore, $\hat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}_0 = -\boldsymbol{\Phi}^{-1} \boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \boldsymbol{\beta}_0, \boldsymbol{R}) + o_P(1)$. Finally,

$$(\boldsymbol{\Phi}^{-1} \boldsymbol{\Xi}_\pi \boldsymbol{\Phi}^{-1})^{-1/2} (\hat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}_0) = -(\boldsymbol{\Phi}^{-1} \boldsymbol{\Xi}_\pi \boldsymbol{\Phi}^{-1})^{-1/2} \boldsymbol{\Phi}^{-1} \boldsymbol{\Xi}_\pi^{1/2} \boldsymbol{\Xi}_\pi^{-1/2} \boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \boldsymbol{\beta}_0, \boldsymbol{R}) + o_P(1),$$

and

$$(\boldsymbol{\Phi}^{-1} \boldsymbol{\Xi}_\pi \boldsymbol{\Phi}^{-1})^{-1/2} (\hat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}_0) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I}_p).$$

□

Proof of Theorem 2. The problem that minimizing $\mathrm{tr}(\boldsymbol{D} \boldsymbol{\Phi}^{-1} \boldsymbol{\Xi}_\pi \boldsymbol{\Phi}^{-1} \boldsymbol{D}^{\mathrm{T}})$, where $\boldsymbol{D} \in \mathbb{R}^{p \times p}$ is a fixed and known nonsingular matrix, is to solve the following optimization problem:

$$\begin{aligned} &\min\ \tilde{H} = \sum_{i=1}^{n} \mathrm{tr}\big\{\frac{1}{\pi_i}[\boldsymbol{D} \boldsymbol{\Phi}^{-1} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0) \boldsymbol{R}^{-1} \boldsymbol{\Pi}_0 \boldsymbol{R}^{-1} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0) \boldsymbol{X}_i \boldsymbol{\Phi}^{-1} \boldsymbol{D}^{\mathrm{T}}]\big\} \\ &\text{s.t} \ \sum_{i=1}^{n} \pi_i = r, \ 0 \le \pi_i \le 1 \text{ for } i \in [n], \end{aligned}$$

---

where $[n] = \{1, 2, \ldots, n\}$. Recall that

$$h_i^{\boldsymbol{D}} = [\mathrm{tr}\{\boldsymbol{D}\boldsymbol{\Phi}^{-1}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i\boldsymbol{\Phi}^{-1}\boldsymbol{D}^{\mathrm{T}}\}]^{1/2}.$$

Without loss of generality, assume that $0 < h_1^{\boldsymbol{D}} \leq h_2^{\boldsymbol{D}} \leq \cdots \leq h_n^{\boldsymbol{D}}$ and $h_{n+1}^{\boldsymbol{D}} = +\infty$. From the Cauchy-Schwarz inequality,

$$\tilde{H} = \frac{1}{r}\sum_{j=1}^{n}\pi_j\sum_{i=1}^{n}\frac{1}{\pi_i}(h_i^{\boldsymbol{D}})^2 \geq \frac{1}{r}(\sum_{i=1}^{n}h_i^{\boldsymbol{D}})^2,$$

where the equality holds if and only if $\pi_i \propto h_i^{\boldsymbol{D}}$. We consider the following two cases:

Case 1: If all $rh_i^{\boldsymbol{D}}/\sum_{j=1}^{n}h_j^{\boldsymbol{D}} \leq 1$, then $\pi_i = rh_i^{\boldsymbol{D}}/\sum_{j=1}^{n}h_j^{\boldsymbol{D}}$ for $i \in [n]$ gives the optimal solution.

Case 2: Assume that there exists some i such that $rh_i^{\boldsymbol{D}}/\sum_{j=1}^{n}h_j^{\boldsymbol{D}} > 1$. By the definition of ω, it satisfies $(r-\omega+1)h_{n-\omega+1}^{\boldsymbol{D}} \geq \sum_{j=1}^{n-\omega+1}h_j^{\boldsymbol{D}}$ and $(r-\omega)h_{n-\omega}^{\boldsymbol{D}} < \sum_{j=1}^{n-\omega}h_j^{\boldsymbol{D}}$, and the number of such i is ω. Therefore, the original optimization turns into the following optimization problem:

$$\begin{aligned}
\min \sum_{i=1}^{n-\omega}&\mathrm{tr}\{\frac{1}{\pi_i}[\boldsymbol{D}\boldsymbol{\Phi}^{-1}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i\boldsymbol{\Phi}^{-1}\boldsymbol{D}^{\mathrm{T}}]\}\\
\text{s.t} \sum_{i=1}^{n-\omega}&\pi_i = r-\omega,\ 0 \leq \pi_i \leq 1 \text{ for } i \in [n-\omega].\\
&\pi_{n-\omega+1} = \cdots = \pi_n = 1.
\end{aligned}$$

Similar to the calculation of the optimal π_i's in Case 1, from the Cauchy-Schwarz

---

inequality,

$$\begin{aligned}
&\sum_{i=1}^{n-\omega} \mathrm{tr}\{\frac{1}{\pi_i}[\boldsymbol{D}\boldsymbol{\Phi}^{-1}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i\boldsymbol{\Phi}^{-1}\boldsymbol{D}^{\mathrm{T}}]\} \\
=&\frac{1}{r-\omega}\sum_{j=1}^{n-\omega}\pi_j\sum_{i=1}^{n-\omega}\frac{1}{\pi_i}(h_i^{\boldsymbol{D}})^2 \geq \frac{1}{r-\omega}(\sum_{i=1}^{n-\omega} h_i^{\boldsymbol{D}})^2,
\end{aligned}$$

where the equality holds if and only if $\pi_i \propto h_i^{\boldsymbol{D}}$, i.e., $\pi_i = (r-\omega)h_i^{\boldsymbol{D}}/\sum_{j=1}^n h_j^{\boldsymbol{D}}, i \in [n-\omega]$. Then the optimal solution is min $\tilde{H} = (r-\omega)^{-1}(\sum_{i=1}^{n-\omega} h_i^{\boldsymbol{D}})^2 + \sum_{i=n-\omega+1}^{n}(h_i^{\boldsymbol{D}})^2$. Assume that $M^{\boldsymbol{D}}$ exists such that

$$\max_{1\leq i\leq n}\frac{r(h_i^{\boldsymbol{D}}\wedge M^{\boldsymbol{D}})}{\sum_{j=1}^n(h_j^{\boldsymbol{D}}\wedge M^{\boldsymbol{D}})} = 1,$$

and $h_{n-\omega}^{\boldsymbol{D}} < M^{\boldsymbol{D}} \leq h_{n-\omega+1}^{\boldsymbol{D}}$, then $\sum_{i=1}^{n-\omega} h_i^{\boldsymbol{D}} = (r-\omega)M^{\boldsymbol{D}}$ holds. Therefore it can be rewritten that min $\tilde{H} = (r-\omega)(M^{\boldsymbol{D}})^2 + \sum_{i=n-\omega+1}^{n}(h_i^{\boldsymbol{D}})^2$. Taking $\pi_i = r(h_i^{\boldsymbol{D}}\wedge M^{\boldsymbol{D}})/\sum_{j=1}^n(h_j^{\boldsymbol{D}}\wedge M^{\boldsymbol{D}})$ into $\tilde{H}$, the following equation holds:

$$\begin{aligned}
\tilde{H} =& \sum_{i=1}^{n-\omega}\frac{\sum_{j=1}^n(h_j^{\boldsymbol{D}}\wedge M^{\boldsymbol{D}})}{r(h_i^{\boldsymbol{D}}\wedge M^{\boldsymbol{D}})}(h_i^{\boldsymbol{D}})^2 + \sum_{i=n-\omega+1}^{n}\frac{\sum_{j=1}^n(h_j^{\boldsymbol{D}}\wedge M^{\boldsymbol{D}})}{r(h_i^{\boldsymbol{D}}\wedge M^{\boldsymbol{D}})}(h_i^{\boldsymbol{D}})^2 \\
=& \sum_{i=1}^{n-\omega}\frac{\sum_{j=1}^{n-\omega} h_j^{\boldsymbol{D}} + \omega M^{\boldsymbol{D}}}{r h_i^{\boldsymbol{D}}}(h_i^{\boldsymbol{D}})^2 + \sum_{i=n-\omega+1}^{n}\frac{\sum_{j=1}^{n-\omega} h_j^{\boldsymbol{D}} + \omega M^{\boldsymbol{D}}}{r M^{\boldsymbol{D}}}(h_i^{\boldsymbol{D}})^2 \\
=& \sum_{i=1}^{n-\omega} M^{\boldsymbol{D}} h_i^{\boldsymbol{D}} + \sum_{i=n-\omega+1}^{n}(h_i^{\boldsymbol{D}})^2 \\
=& (r-\omega)(M^{\boldsymbol{D}})^2 + \sum_{i=n-\omega+1}^{n}(h_i^{\boldsymbol{D}})^2 = \min\ \tilde{H}.
\end{aligned}$$

It can be seen that the above result echoes the minimum of the optimal problem and then $\pi_i = r(h_i^{\boldsymbol{D}}\wedge M^{\boldsymbol{D}})/\sum_{j=1}^n(h_j^{\boldsymbol{D}}\wedge M^{\boldsymbol{D}})$ satisfies the optimal solution.

Below we prove the existence of $M^{\boldsymbol{D}}$ and it satisfies $h_{n-\omega}^{\boldsymbol{D}} < M^{\boldsymbol{D}} \leq h_{n-\omega+1}^{\boldsymbol{D}}$. Note

---

that ω satisfies

$$\frac{(r-\omega+1)h^{\boldsymbol{D}}_{n-\omega+1}}{\sum_{i=1}^{n-\omega+1} h^{\boldsymbol{D}}_i} \geq 1 \text{ and } \frac{(r-\omega)h^{\boldsymbol{D}}_{n-\omega}}{\sum_{i=1}^{n-\omega} h^{\boldsymbol{D}}_i} < 1.$$

By taking $M^{\boldsymbol{D}}_1 = h^{\boldsymbol{D}}_{n-\omega+1}$ and $M^{\boldsymbol{D}}_2 = h^{\boldsymbol{D}}_{n-\omega}$,

$$\frac{(r-\omega+1)h^{\boldsymbol{D}}_{n-\omega+1} + (\omega-1)M^{\boldsymbol{D}}_1}{\sum_{i=1}^{n-\omega+1} h^{\boldsymbol{D}}_i + (\omega-1)M^{\boldsymbol{D}}_1} \geq 1 \text{ and } \frac{(r-\omega)h^{\boldsymbol{D}}_{n-\omega} + \omega M^{\boldsymbol{D}}_2}{\sum_{i=1}^{n-\omega} h^{\boldsymbol{D}}_i + \omega M^{\boldsymbol{D}}_2} < 1,$$

which implies the fact that

$$r\frac{(h^{\boldsymbol{D}}_i \wedge M^{\boldsymbol{D}}_1)}{\sum_{j=1}^{n}(h^{\boldsymbol{D}}_j \wedge M^{\boldsymbol{D}}_1)} \geq 1 \text{ and } r\frac{(h^{\boldsymbol{D}}_i \wedge M^{\boldsymbol{D}}_2)}{\sum_{j=1}^{n}(h^{\boldsymbol{D}}_j \wedge M^{\boldsymbol{D}}_2)} < 1.$$

Thus, the existence of $M^{\boldsymbol{D}}$ is proved, following the facts that $\max_{1\leq i\leq n}\{r(h^{\boldsymbol{D}}_i \wedge M^{\boldsymbol{D}})/\sum_{j=1}^{n}(h^{\boldsymbol{D}}_j \wedge M^{\boldsymbol{D}})\}$ is continuous on $M^{\boldsymbol{D}}$ conditional on $h^{\boldsymbol{D}}_1, \ldots, h^{\boldsymbol{D}}_n$. On the other hand, for any $h^{\boldsymbol{D}}_n \geq M'^{\boldsymbol{D}} > M^{\boldsymbol{D}}$, it can be verified that $(M'^{\boldsymbol{D}} \wedge h^{\boldsymbol{D}}_n) \geq (M^{\boldsymbol{D}} \wedge h^{\boldsymbol{D}}_n)$, and simple analysis leads to $M'^{\boldsymbol{D}}/\sum_{i=1}^{n}(h^{\boldsymbol{D}}_i \wedge M'^{\boldsymbol{D}}) \geq M^{\boldsymbol{D}}/\sum_{i=1}^{n}(h^{\boldsymbol{D}}_i \wedge M^{\boldsymbol{D}})$. Thus, $(h^{\boldsymbol{D}}_n \wedge M^{\boldsymbol{D}})/\sum_{i=1}^{n}(h^{\boldsymbol{D}}_i \wedge M^{\boldsymbol{D}})$ is nondecreasing on $M^{\boldsymbol{D}} \in (h^{\boldsymbol{D}}_1, h^{\boldsymbol{D}}_n)$. Therefore

$$\max_{1\leq i\leq n} \frac{r(h^{\boldsymbol{D}}_i \wedge M^{\boldsymbol{D}})}{\sum_{j=1}^{n}(h^{\boldsymbol{D}}_j \wedge M^{\boldsymbol{D}})} = 1$$

indicates that $h^{\boldsymbol{D}}_{n-\omega} < M^{\boldsymbol{D}} \leq h^{\boldsymbol{D}}_{n-\omega+1}$. The proof of Theorem 2 is complete. $\square$

Proof of Theorem 3. It is easy to verify that the approximated optimal probabilities $\{\tilde{\pi}^{\boldsymbol{D}}_i\}_{i=1}^{n}$ still satisfies Assumption (A.3). Therefore, by the consistency of the pilot estimators $\tilde{\boldsymbol{\beta}}_{\mathrm{p}}$, $\tilde{\boldsymbol{R}}_{\mathrm{p}}$, $\tilde{\boldsymbol{\Pi}}_{\mathrm{p}}$ and $\tilde{\boldsymbol{\Phi}}_{\mathrm{p}}$, one can apply the proof of Theorem 1 directly and thus the consistency and asymptotic normality still holds. We only prove the

---

derivation of $\Xi^{\boldsymbol{D}}$ from Ξ_π by inserting $\{\tilde{\pi}_i^{\boldsymbol{D}}\}_{i\in[n]}$. Since

$$\begin{aligned}
&E\Big[\frac{1}{n^2}\sum_{i=1}^{n}\frac{r}{\tilde{\pi}_i^{\boldsymbol{D}}}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i\Big]\\
=&E\Big[\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{1}{n}\sum_{i=1}^{n}\tilde{h}_i^{\boldsymbol{D}}}{\tilde{h}_i^{\boldsymbol{D}}}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i\Big]\\
\to& m^{\boldsymbol{D}}E\Big[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i^{\boldsymbol{D}}}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i\Big]\\
=&m^{\boldsymbol{D}}\Xi^{\boldsymbol{D}},
\end{aligned}$$

the desired results for $\hat{\boldsymbol{\beta}}_w^{\boldsymbol{D}}$ hold. □

Proof of Theorem 4. (i) For consistency, the first step is to show that $\|\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}}-\boldsymbol{\beta}_0\| = o_P(1)$. For any $\boldsymbol{\beta}\in\mathbb{R}^p$, denote

$$\begin{aligned}
\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta};\boldsymbol{R}) &= \frac{1}{n}\sum_{i=1}^{n}\eta_i^{\boldsymbol{D}}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta})\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta})\},\\
\boldsymbol{S}_{uw}(\boldsymbol{\beta};\boldsymbol{R}) &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta})\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta})\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta})\},
\end{aligned}$$

where $\eta_i^{\boldsymbol{D}}$ is the indicator function taking value 1 if the i-th cluster $(\boldsymbol{X}_i,\boldsymbol{Y}_i)$ is selected and 0 otherwise. Similar to the proof of Theorem 1, we obtain $\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta};\boldsymbol{R}) = E[\boldsymbol{S}_{uw}(\boldsymbol{\beta};\boldsymbol{R})]+o_P(1)$. Due to the consistency of $\tilde{\boldsymbol{R}}_{\mathrm{p}}$ to $\boldsymbol{R}$, $\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta};\tilde{\boldsymbol{R}}_{\mathrm{p}}) = \boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta};\boldsymbol{R})+o_P(1)$, and thus $\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta};\tilde{\boldsymbol{R}}_{\mathrm{p}}) = E[\boldsymbol{S}_{uw}(\boldsymbol{\beta};\boldsymbol{R})]+o_P(1)$. Applying Theorem 5.9 of van der Vaart (1998), it yields $\|\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}}-\boldsymbol{\beta}_0\| = o_P(1)$. Next, applying Taylor expansion,

$$\boldsymbol{0} = \boldsymbol{S}_{uw}^{*}(\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}};\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}},\tilde{\boldsymbol{R}}_{\mathrm{p}})$$

$$= \boldsymbol{S}^*_{uw}(\boldsymbol{\beta}_0; \hat{\boldsymbol{\beta}}^{\boldsymbol{D}}_{uw}, \tilde{\boldsymbol{R}}_{\mathrm{p}}) + \nabla_{\boldsymbol{\beta}} \boldsymbol{S}^*_{uw}(\boldsymbol{\beta}_0; \hat{\boldsymbol{\beta}}^{\boldsymbol{D}}_{uw}, \tilde{\boldsymbol{R}}_{\mathrm{p}})(\hat{\boldsymbol{\beta}}^{\boldsymbol{D}}_{uw} - \boldsymbol{\beta}_0) + O_P(\|\hat{\boldsymbol{\beta}}^{\boldsymbol{D}}_{uw} - \boldsymbol{\beta}_0\|^2),$$

where

$$\begin{aligned}
\boldsymbol{S}^*_{uw}(\boldsymbol{\beta}; \hat{\boldsymbol{\beta}}^{\boldsymbol{D}}_{uw}, \tilde{\boldsymbol{R}}_{\mathrm{p}}) &= \frac{1}{n}\sum_{i=1}^{n} \eta_i^{\boldsymbol{D}} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\hat{\boldsymbol{\beta}}^{\boldsymbol{D}}_{uw}) \tilde{\boldsymbol{R}}_{\mathrm{p}}^{-1} \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}^{\boldsymbol{D}}_{uw}) \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\}, \\
\nabla_{\boldsymbol{\beta}} \boldsymbol{S}^*_{uw}(\boldsymbol{\beta}; \hat{\boldsymbol{\beta}}^{\boldsymbol{D}}_{uw}, \tilde{\boldsymbol{R}}_{\mathrm{p}}) &= \frac{1}{n}\sum_{i=1}^{n} \eta_i^{\boldsymbol{D}} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\hat{\boldsymbol{\beta}}^{\boldsymbol{D}}_{uw}) \tilde{\boldsymbol{R}}_{\mathrm{p}}^{-1} \boldsymbol{A}_i^{-1/2}(\hat{\boldsymbol{\beta}}^{\boldsymbol{D}}_{uw}) \boldsymbol{A}_i(\boldsymbol{\beta}) \boldsymbol{X}_i.
\end{aligned}$$

Following the same proof of Theorem 1, we obtain $\|\boldsymbol{S}^*_{uw}(\boldsymbol{\beta}_0; \hat{\boldsymbol{\beta}}^{\boldsymbol{D}}_{uw}, \tilde{\boldsymbol{R}}_{\mathrm{p}})\| = O_P(r^{-1/2})$, $\|\nabla_{\boldsymbol{\beta}} \boldsymbol{S}^*_{uw}(\boldsymbol{\beta}_0; \hat{\boldsymbol{\beta}}^{\boldsymbol{D}}_{uw}, \tilde{\boldsymbol{R}}_{\mathrm{p}})\| = O_P(\|\hat{\boldsymbol{\beta}}^{\boldsymbol{D}}_{uw} - \boldsymbol{\beta}_0\|)$, and we conclude $\|\hat{\boldsymbol{\beta}}^{\boldsymbol{D}}_{uw} - \boldsymbol{\beta}_0\| = O_P(r^{-1/2})$ from the Taylor's expansion.

(ii) Below we prove the asymptotic normality. Let

$$m^{\boldsymbol{D}} \boldsymbol{S}^*_{uw}(\boldsymbol{\beta}_0; \boldsymbol{R}) = \sum_{i=1}^{n} \frac{m^{\boldsymbol{D}} \eta_i^{\boldsymbol{D}}}{n} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0) \boldsymbol{R}^{-1} \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0) \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\} := \sum_{i=1}^{n} \boldsymbol{\xi}_i^{\boldsymbol{D}}.$$

Now, we check the conditions of Lindeberg-Feller central limit theorem. For $\forall \epsilon > 0$,

$$\begin{aligned}
\sum_{i=1}^{n} E[\|\boldsymbol{\xi}_i^{\boldsymbol{D}}\|^2 I(\|\boldsymbol{\xi}_i\| \geq \epsilon)] &\leq \frac{1}{\epsilon} \sum_{i=1}^{n} E(\|\boldsymbol{\xi}_i^{\boldsymbol{D}}\|^3) \\
&= \frac{1}{\epsilon} \sum_{i=1}^{n} E[E(\|\boldsymbol{\xi}_i^{\boldsymbol{D}}\|^3 | \mathcal{F}_n)] \\
&\leq \frac{(m^{\boldsymbol{D}})^3}{\epsilon} \sum_{i=1}^{n} E\Big[\frac{1}{n^3} \|\boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0) \boldsymbol{R}^{-1} \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0) \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}\|^3\Big] \\
&= o(1).
\end{aligned}$$

Thus the condition of Lindeberg-Feller central limit theorem is satisfied. Note that

$$E[m^{\boldsymbol{D}} \boldsymbol{S}^*_{uw}(\boldsymbol{\beta}_0; \boldsymbol{R})] = E\{E[m^{\boldsymbol{D}} \boldsymbol{S}^*_{uw}(\boldsymbol{\beta}_0; \boldsymbol{R}) | \mathcal{F}_n]\}$$

$$
\begin{aligned}
&=E\Big[\sum_{i=1}^{n}\frac{m^{\boldsymbol{D}}\tilde{\pi}_i^{\boldsymbol{D}}}{n}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}\Big]\\
&=E\Big\{E\Big[\sum_{i=1}^{n}\frac{m^{\boldsymbol{D}}\tilde{\pi}_i^{\boldsymbol{D}}}{n}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}\Big|\{\boldsymbol{X}_i\}_{i=1}^{n}\Big]\Big\}\\
&=\boldsymbol{0},
\end{aligned}
$$

and

$$
\begin{aligned}
&Var[m^{\boldsymbol{D}}\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta}_0;\boldsymbol{R})]\\
=&E\{Var[m^{\boldsymbol{D}}\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta}_0;\boldsymbol{R})|\mathcal{F}_n]\}+Var\{E[m^{\boldsymbol{D}}\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta}_0;\boldsymbol{R})|\mathcal{F}_n]\}\\
=&E\Big\{\frac{(m^{\boldsymbol{D}})^2}{n^2}\sum_{i=1}^{n}\tilde{\pi}_i^{\boldsymbol{D}}(1-\tilde{\pi}_i^{\boldsymbol{D}})[\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}]^{\otimes 2}\Big\}\\
&+Var\Big[\frac{m^{\boldsymbol{D}}}{n}\sum_{i=1}^{n}\tilde{\pi}_i^{\boldsymbol{D}}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}\Big].
\end{aligned}
$$

To proceed, by using the iterated expectation and iterated variance formula again,

$$
\begin{aligned}
&E\Big\{\frac{(m^{\boldsymbol{D}})^2}{n^2}\sum_{i=1}^{n}\tilde{\pi}_i^{\boldsymbol{D}}(1-\tilde{\pi}_i^{\boldsymbol{D}})[\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}]^{\otimes 2}\Big\}\\
=&E\Big(E\Big\{\frac{(m^{\boldsymbol{D}})^2}{n^2}\sum_{i=1}^{n}\tilde{\pi}_i^{\boldsymbol{D}}(1-\tilde{\pi}_i^{\boldsymbol{D}})[\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}]^{\otimes 2}\Big|\{\boldsymbol{X}_i\}_{i=1}^{n}\Big\}\Big)\\
=&E\Big[\frac{(m^{\boldsymbol{D}})^2}{n^2}\sum_{i=1}^{n}\tilde{\pi}_i^{\boldsymbol{D}}(1-\tilde{\pi}_i^{\boldsymbol{D}})\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i\Big],\\
&Var\Big[\frac{m^{\boldsymbol{D}}}{n}\sum_{i=1}^{n}\tilde{\pi}_i^{\boldsymbol{D}}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}\Big]\\
=&E\Big\{Var\Big[\frac{m^{\boldsymbol{D}}}{n}\sum_{i=1}^{n}\tilde{\pi}_i^{\boldsymbol{D}}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}\Big|\{\boldsymbol{X}_i\}_{i=1}^{n}\Big]\Big\}\\
&+Var\Big\{E\Big[\frac{m^{\boldsymbol{D}}}{n}\sum_{i=1}^{n}\tilde{\pi}_i^{\boldsymbol{D}}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}\Big|\{\boldsymbol{X}_i\}_{i=1}^{n}\Big]\Big\}
\end{aligned}
$$

$$=E\Big[\frac{(m^{\boldsymbol{D}})^2}{n^2}\sum_{i=1}^{n}(\tilde{\pi}_i^{\boldsymbol{D}})^2\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i\Big].$$

Consequently,

$$\begin{aligned}Var[m^{\boldsymbol{D}}\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta}_0;\boldsymbol{R})] &= E\Big[\frac{(m^{\boldsymbol{D}})^2}{n^2}\sum_{i=1}^{n}\tilde{\pi}_i^{\boldsymbol{D}}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i\Big]\\ &\to m^{\boldsymbol{D}}\boldsymbol{\Delta}^{\boldsymbol{D}}.\end{aligned}$$

By the Lindeberg-Feller central limit theorem,

$$(m^{\boldsymbol{D}}\boldsymbol{\Delta}^{\boldsymbol{D}})^{-1/2}[m^{\boldsymbol{D}}\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta}_0;\boldsymbol{R})]\xrightarrow{d}N(\boldsymbol{0},\boldsymbol{I}_p).$$

According to the Taylor's expansion,

$$\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}}-\boldsymbol{\beta}_0=-[\nabla_{\boldsymbol{\beta}}\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta}_0;\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}},\tilde{\boldsymbol{R}}_{\mathrm{p}})]^{-1}[\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta}_0;\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}},\tilde{\boldsymbol{R}}_{\mathrm{p}})+O_P(\|\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}}-\boldsymbol{\beta}_0\|^2)].$$

Note that

$$\begin{aligned}-[m^{\boldsymbol{D}}\nabla_{\boldsymbol{\beta}}\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta}_0;\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}},\tilde{\boldsymbol{R}}_{\mathrm{p}})]^{-1}&=(\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1}-[m^{\boldsymbol{D}}\nabla_{\boldsymbol{\beta}}\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta}_0;\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}},\tilde{\boldsymbol{R}}_{\mathrm{p}})]^{-1}-(\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1},\\ m^{\boldsymbol{D}}\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta}_0;\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}},\tilde{\boldsymbol{R}}_{\mathrm{p}})&=m^{\boldsymbol{D}}\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta}_0;\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}},\tilde{\boldsymbol{R}}_{\mathrm{p}})-m^{\boldsymbol{D}}\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta}_0;\boldsymbol{R})+m^{\boldsymbol{D}}\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta}_0;\boldsymbol{R}).\end{aligned}$$

Following the similar argument to the proof of Theorem 1,

$$(\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1}-[m^{\boldsymbol{D}}\nabla_{\boldsymbol{\beta}}\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta}_0;\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}},\tilde{\boldsymbol{R}}_{\mathrm{p}})]^{-1}=o_P(1),\ m^{\boldsymbol{D}}\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta}_0;\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}},\tilde{\boldsymbol{R}}_{\mathrm{p}})-m^{\boldsymbol{D}}\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta}_0;\boldsymbol{R})=o_P(r^{-1/2}).$$

Therefore, $\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}}-\boldsymbol{\beta}_0=-(\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1}m^{\boldsymbol{D}}\boldsymbol{S}_{uw}^{*}(\boldsymbol{\beta}_0;\boldsymbol{R})+o_P(1)$. Finally,

$$[m^{\boldsymbol{D}}(\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1}\boldsymbol{\Delta}^{\boldsymbol{D}}(\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1}]^{-1/2}(\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}}-\boldsymbol{\beta}_0)$$

$$= -\,[m^{\boldsymbol{D}}(\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1}\boldsymbol{\Delta}^{\boldsymbol{D}}(\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1}]^{-1/2}(\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1}(m^{\boldsymbol{D}}\boldsymbol{\Delta}^{\boldsymbol{D}})^{1/2}(m^{\boldsymbol{D}}\boldsymbol{\Delta}^{\boldsymbol{D}})^{-1/2}[m^{\boldsymbol{D}}\boldsymbol{S}^{*}_{uw}(\boldsymbol{\beta}_0;\boldsymbol{R})] + o_P(1),$$

and

$$[m^{\boldsymbol{D}}(\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1}\boldsymbol{\Delta}^{\boldsymbol{D}}(\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1}]^{-1/2}(\hat{\boldsymbol{\beta}}^{\boldsymbol{D}}_{uw} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I}_p).$$

(iii) Finally, we prove that when $\boldsymbol{R} = \boldsymbol{\Pi}_0$, the inequality $(\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1}\boldsymbol{\Delta}^{\boldsymbol{D}}(\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1} \leq \boldsymbol{\Phi}^{-1}\boldsymbol{\Xi}^{\boldsymbol{D}}\boldsymbol{\Phi}^{-1}$ holds under the Loewner ordering. A key observation is that $(\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1}\boldsymbol{\Delta}^{\boldsymbol{D}}(\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1} = (\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1}$ when $\boldsymbol{R} = \boldsymbol{\Pi}_0$. Since $\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i := \boldsymbol{V}_i$ is a real symmetric positive-definite matrix, Cholesky decomposition gives $\boldsymbol{V}_i = \mathbf{v}_i\mathbf{v}_i^{\mathrm{T}}$ for some $\mathbf{v}_i \in \mathbb{R}^{p\times p}$. We then only need to prove

$$E(h_i^{\boldsymbol{D}}\mathbf{v}_i\mathbf{v}_i^{\mathrm{T}})^{-1} \leq E(\mathbf{v}_i\mathbf{v}_i^{\mathrm{T}})^{-1}E(\mathbf{v}_i\mathbf{v}_i^{\mathrm{T}}/h_i^{\boldsymbol{D}})E(\mathbf{v}_i\mathbf{v}_i^{\mathrm{T}})^{-1}.$$

Denote

$$\mathbf{f}_i := \sqrt{h_i^{\boldsymbol{D}}}E(h_i^{\boldsymbol{D}}\mathbf{v}_i\mathbf{v}_i^{\mathrm{T}})^{-1}\mathbf{v}_i - \frac{1}{\sqrt{h_i^{\boldsymbol{D}}}}E(\mathbf{v}_i\mathbf{v}_i^{\mathrm{T}})^{-1}\mathbf{v}_i.$$

Since $\mathbf{f}_i\mathbf{f}_i^{\mathrm{T}} \geq 0$, we have

$$0 \leq E(\mathbf{f}_i\mathbf{f}_i^{\mathrm{T}}) = E(\mathbf{v}_i\mathbf{v}_i^{\mathrm{T}})^{-1}E(\mathbf{v}_i\mathbf{v}_i^{\mathrm{T}}/h_i^{\boldsymbol{D}})E(\mathbf{v}_i\mathbf{v}_i^{\mathrm{T}})^{-1} - E(h_i^{\boldsymbol{D}}\mathbf{v}_i\mathbf{v}_i^{\mathrm{T}})^{-1}.$$

Therefore the proof of Theorem 4 is completed. $\square$

Before we prove Theorems 5-7, some lemmas are provided as follows.

Lemma 1. *Under Assumptions (A.4)-(A.8),*

$$(i)\|\breve{\boldsymbol{\gamma}}_{\mathrm{p}} - \boldsymbol{\gamma}_0\|_1 = O_P(s_{\boldsymbol{\beta}}\sqrt{\log p/r^{\mathrm{p}}}),\ \ \|\breve{\boldsymbol{\gamma}}_{\mathrm{p}} - \boldsymbol{\gamma}_0\| = O_P(\sqrt{s_{\boldsymbol{\beta}}\log p/r^{\mathrm{p}}}),$$

$$(ii) \|\breve{\boldsymbol{R}}_{\mathrm{p}} - \boldsymbol{R}\| = O_P(\sqrt{s_{\boldsymbol{\beta}} \log p / r^{\mathrm{p}}}),$$

$$(iii) \max_{1 \leq j \leq d} \|(\breve{\boldsymbol{W}}_{\mathrm{p}} - \boldsymbol{W}_0)_j\|_1 = O_P((s_{\boldsymbol{W}} \vee s_{\boldsymbol{\beta}}) \sqrt{\log p / r^{\mathrm{p}}}),$$

$$(iv) \max_{1 \leq j \leq d} \|(\breve{\boldsymbol{W}}_{\mathrm{p}}^{\boldsymbol{D}} - \boldsymbol{W}_0^{\boldsymbol{D}})_j\|_1 = O_P((s_{\boldsymbol{W}^{\boldsymbol{D}}} \vee s_{\boldsymbol{\beta}}) \sqrt{\log p / r^{\mathrm{p}}}),$$

where $(\boldsymbol{A})_j$ *denotes the* j*th column of a matrix* $\boldsymbol{A}$*.*

Proof of Lemma 1. For (i), see Lemma 2 in Gao et al. (2025). (ii) follows from (i). For (iii), see Lemma S.1.6 in Fang et al. (2020). The proof of (iv) is similar to (iii). □

Proof of Theorem 5. (i) The first step is to show that $\|\breve{\boldsymbol{\theta}}_w - \boldsymbol{\theta}_0\| = o_P(1)$. Let

$$\begin{aligned}
&\boldsymbol{Q}_w^*(\boldsymbol{\vartheta}, \breve{\gamma}_{\mathrm{p}}, \breve{\boldsymbol{W}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}}) \\
=&\frac{1}{n} \sum_{i=1}^{n} \frac{\eta_i}{\varpi_i} (\boldsymbol{Z}_i - \boldsymbol{U}_i \breve{\boldsymbol{W}}_{\mathrm{p}})^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\vartheta}, \breve{\gamma}_{\mathrm{p}}) \breve{\boldsymbol{R}}_{\mathrm{p}}^{-1} \boldsymbol{A}_i^{-1/2}(\boldsymbol{\vartheta}, \breve{\gamma}_{\mathrm{p}}) \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\vartheta}, \breve{\gamma}_{\mathrm{p}})\}, \\
&\boldsymbol{Q}_w^*(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R}) \\
=&\frac{1}{n} \sum_{i=1}^{n} \frac{\eta_i}{\varpi_i} (\boldsymbol{Z}_i - \boldsymbol{U}_i \boldsymbol{W}_0)^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0) \boldsymbol{R}^{-1} \boldsymbol{A}_i^{-1/2}(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0) \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0)\}, \\
&\boldsymbol{Q}_w(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R}) \\
=&\frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{Z}_i - \boldsymbol{U}_i \boldsymbol{W}_0)^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0) \boldsymbol{R}^{-1} \boldsymbol{A}_i^{-1/2}(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0) \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0)\}.
\end{aligned}$$

It can be seen that

$$\begin{aligned}
&E[\boldsymbol{Q}_w^*(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R})] = E\{E[\boldsymbol{Q}_w^*(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R}) | \mathcal{F}_n]\} = E[\boldsymbol{Q}_w(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R})], \\
&Var[\boldsymbol{Q}_w^*(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R})]
\end{aligned}$$

$$
\begin{aligned}
=&E\{Var[\boldsymbol{Q}_w^*(\boldsymbol{\vartheta},\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R})|\mathcal{F}_n]\}+Var\{E[\boldsymbol{Q}_w^*(\boldsymbol{\vartheta},\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R})|\mathcal{F}_n]\}\\
\leq&E\Big[\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{\varpi_i}[(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\vartheta},\boldsymbol{\gamma}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\vartheta},\boldsymbol{\gamma}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\vartheta},\boldsymbol{\gamma}_0)\}]^{\otimes 2}\Big]\\
=&O(r^{-1}).
\end{aligned}
$$

Then we have

$$
\|\boldsymbol{Q}_w^*(\boldsymbol{\vartheta},\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R})-E[\boldsymbol{Q}_w(\boldsymbol{\vartheta},\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R})]\|=O_P\left(r^{-1/2}\right).
$$

By the consistency of $\breve{\boldsymbol{\gamma}}_{\mathrm{p}},\breve{\boldsymbol{W}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}}$ to $\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R}$ in Lemma 1, respectively, we have

$$
\|\boldsymbol{Q}_w^*(\boldsymbol{\vartheta},\breve{\boldsymbol{\gamma}}_{\mathrm{p}},\breve{\boldsymbol{W}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}})-\boldsymbol{Q}_w^*(\boldsymbol{\vartheta},\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R})\|=o_P(1).
$$

Combining the above results, we obtain

$$
\|\boldsymbol{Q}_w^*(\boldsymbol{\vartheta},\breve{\boldsymbol{\gamma}}_{\mathrm{p}},\breve{\boldsymbol{W}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}})-E[\boldsymbol{Q}_w(\boldsymbol{\vartheta},\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R})]\|=o_P(1),
$$

which indicates that $\|\breve{\boldsymbol{\theta}}_w-\boldsymbol{\theta}_0\|=o_P(1)$.

In the second step, using Taylor's theorem,

$$
\begin{aligned}
\mathbf{0}&=\boldsymbol{Q}_w^*(\breve{\boldsymbol{\theta}}_w,\breve{\boldsymbol{\gamma}}_{\mathrm{p}},\breve{\boldsymbol{W}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w,\breve{\boldsymbol{\gamma}}_{\mathrm{p}})\\
&=\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\breve{\boldsymbol{W}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w,\breve{\boldsymbol{\gamma}}_{\mathrm{p}})+\nabla_{\boldsymbol{\vartheta}}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\breve{\boldsymbol{W}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w,\breve{\boldsymbol{\gamma}}_{\mathrm{p}})(\breve{\boldsymbol{\theta}}_w-\boldsymbol{\theta}_0)\\
&\quad+\nabla_{\boldsymbol{\varphi}}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\breve{\boldsymbol{W}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w,\breve{\boldsymbol{\gamma}}_{\mathrm{p}})(\breve{\boldsymbol{\gamma}}_{\mathrm{p}}-\boldsymbol{\gamma}_0)+\breve{\boldsymbol{\mathcal{R}}}
\end{aligned}
$$

---

where $\breve{\boldsymbol{\mathcal{R}}} = O_P(\|\breve{\boldsymbol{\theta}}_w - \boldsymbol{\theta}_0\|^2 + \|\breve{\gamma}_{\mathrm{p}} - \boldsymbol{\gamma}_0\|^2)$ and

$$\begin{aligned}
&\boldsymbol{Q}_w^*(\boldsymbol{\vartheta}, \boldsymbol{\varphi}, \breve{\boldsymbol{W}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w, \breve{\gamma}_{\mathrm{p}})\\
=&\frac{1}{n}\sum_{i=1}^{n}\frac{\eta_i}{\varpi_i}(\boldsymbol{Z}_i - \boldsymbol{U}_i\breve{\boldsymbol{W}}_{\mathrm{p}})^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\breve{\boldsymbol{\theta}}_w, \breve{\gamma}_{\mathrm{p}})\breve{\boldsymbol{R}}_{\mathrm{p}}^{-1}\boldsymbol{A}_i^{-1/2}(\breve{\boldsymbol{\theta}}_w, \breve{\gamma}_{\mathrm{p}})\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\vartheta}, \boldsymbol{\varphi})\},\\
&\nabla_{\boldsymbol{\vartheta}}\boldsymbol{Q}_w^*(\boldsymbol{\vartheta}, \boldsymbol{\varphi}, \breve{\boldsymbol{W}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w, \breve{\gamma}_{\mathrm{p}})\\
=&-\frac{1}{n}\sum_{i=1}^{n}\frac{\eta_i}{\varpi_i}(\boldsymbol{Z}_i - \boldsymbol{U}_i\breve{\boldsymbol{W}}_{\mathrm{p}})^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\breve{\boldsymbol{\theta}}_w, \breve{\gamma}_{\mathrm{p}})\breve{\boldsymbol{R}}_{\mathrm{p}}^{-1}\boldsymbol{A}_i^{-1/2}(\breve{\boldsymbol{\theta}}_w, \breve{\gamma}_{\mathrm{p}})\boldsymbol{A}_i(\boldsymbol{\vartheta}, \boldsymbol{\varphi})\boldsymbol{Z}_i,\\
&\nabla_{\boldsymbol{\varphi}}\boldsymbol{Q}_w^*(\boldsymbol{\vartheta}, \boldsymbol{\varphi}, \breve{\boldsymbol{W}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w, \breve{\gamma}_{\mathrm{p}})\\
=&-\frac{1}{n}\sum_{i=1}^{n}\frac{\eta_i}{\varpi_i}(\boldsymbol{Z}_i - \boldsymbol{U}_i\breve{\boldsymbol{W}}_{\mathrm{p}})^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\breve{\boldsymbol{\theta}}_w, \breve{\gamma}_{\mathrm{p}})\breve{\boldsymbol{R}}_{\mathrm{p}}^{-1}\boldsymbol{A}_i^{-1/2}(\breve{\boldsymbol{\theta}}_w, \breve{\gamma}_{\mathrm{p}})\boldsymbol{A}_i(\boldsymbol{\vartheta}, \boldsymbol{\varphi})\boldsymbol{U}_i
\end{aligned}$$

To show $\|\breve{\boldsymbol{\theta}}_w - \boldsymbol{\theta}_0\| = O_P(r^{-1/2})$, we need to discuss the order of $\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \breve{\boldsymbol{W}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w, \breve{\gamma}_{\mathrm{p}})$, $\nabla_{\boldsymbol{\vartheta}}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \breve{\boldsymbol{W}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w, \breve{\gamma}_{\mathrm{p}})$ and $\nabla_{\boldsymbol{\varphi}}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \breve{\boldsymbol{W}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w, \breve{\gamma}_{\mathrm{p}})(\breve{\gamma}_{\mathrm{p}} - \boldsymbol{\gamma}_0)$, respectively.

(a) $\|\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \breve{\boldsymbol{W}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w, \breve{\gamma}_{\mathrm{p}})\| = O_P(r^{-1/2})$.

To prove the above result, let

$$\begin{aligned}
&\boldsymbol{Q}_w^*(\boldsymbol{\vartheta}, \boldsymbol{\varphi}, \boldsymbol{W}_0, \boldsymbol{R}|\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)\\
=&\frac{1}{n}\sum_{i=1}^{n}\frac{\eta_i}{\varpi_i}(\boldsymbol{Z}_i - \boldsymbol{U}_i\boldsymbol{W})^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)\{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\vartheta}, \boldsymbol{\varphi})\}.
\end{aligned}$$

By the law of iterated expectation and variance,

$$\begin{aligned}
&E[\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R}|\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)]\\
=&E\{E[\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R}|\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)|\mathcal{F}_n]\} = \boldsymbol{0},\\
&Var[\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R}|\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)]
\end{aligned}$$

$$
\begin{aligned}
&=Var\{E[\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R}|\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0)|\mathcal{F}_n]\}+E\{Var[\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R}|\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0)|\mathcal{F}_n]\}\\
&=E\Big[\frac{1}{n^2}\sum_{i=1}^{n}\varpi_i^{-1}(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)\Big]\\
&\le E\Big[\Big(\max_{1\le i\le n}\frac{1}{n\varpi_i}\Big)\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)\Big]\\
&=O(r^{-1}).
\end{aligned}
$$

By Chebyshev's inequality, $\|\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R}|\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0)\|=O_P(r^{-1/2})$. Similar to the proof of Lemma S.1.10 of Fang et al. (2020),

$$
\|\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\breve{\boldsymbol{W}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w,\breve{\boldsymbol{\gamma}}_{\mathrm{p}})-\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R}|\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0)\|=O_P\Big\{\frac{(s_{\boldsymbol{W}}\vee s_{\boldsymbol{\beta}})\log p}{\sqrt{r^{\mathrm{p}}r}}\Big\}.
$$

Thus, we have

$$
\begin{aligned}
&\|\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\breve{\boldsymbol{W}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w,\breve{\boldsymbol{\gamma}}_{\mathrm{p}})\|\\
&\le\|\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R}|\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0)\|+\|\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\breve{\boldsymbol{W}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w,\breve{\boldsymbol{\gamma}}_{\mathrm{p}})-\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R}|\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0)\|\\
&=O_P\Big\{\frac{(s_{\boldsymbol{W}}\vee s_{\boldsymbol{\beta}})\log p}{\sqrt{r^{\mathrm{p}}r}}\Big\}+O_P(r^{-1/2})=O_P(r^{-1/2}),
\end{aligned}
$$

under the assumption in (A.7) that $(r^{\mathrm{p}})^{-1/2}(s_{\boldsymbol{W}}\vee s_{\boldsymbol{\beta}})\log p=o(1)$.

$(b)\|\nabla_{\boldsymbol{\vartheta}}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\breve{\boldsymbol{W}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w,\breve{\boldsymbol{\gamma}}_{\mathrm{p}})-\boldsymbol{\Lambda}\|=o_P(1)$. To show the above result, let

$$
\begin{aligned}
&\nabla_{\boldsymbol{\vartheta}}\boldsymbol{Q}_w^*(\boldsymbol{\vartheta},\boldsymbol{\varphi},\boldsymbol{W}_0,\boldsymbol{R}|\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0)\\
&=-\frac{1}{n}\sum_{i=1}^{n}\frac{\eta_i}{\varpi_i}(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0)\boldsymbol{A}_i(\boldsymbol{\vartheta},\boldsymbol{\varphi})\boldsymbol{Z}_i.
\end{aligned}
$$

It is easy to see that

$$
E[\nabla_{\boldsymbol{\vartheta}}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R}|\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0)]=E\{E[\nabla_{\boldsymbol{\vartheta}}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R}|\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0)|\mathcal{F}_n]\}=\boldsymbol{\Lambda}.
$$

For the (l_1, l_2)-th component of $\nabla_{\boldsymbol{\vartheta}} \boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R}|\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)$ with $1 \le l_1 \le l_2 \le d$,

$$
\begin{aligned}
&Var\{[\nabla_{\boldsymbol{\vartheta}} \boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R}|\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)]_{l_1,l_2}\} \\
=&E(Var\{[\nabla_{\boldsymbol{\vartheta}} \boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R}|\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)]_{l_1,l_2}|\mathcal{F}_n\}) + Var(E\{[\nabla_{\boldsymbol{\vartheta}} \boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R}|\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)]_{l_1,l_2}|\mathcal{F}_n\}) \\
\le&E\Big[\frac{1}{n^2}\sum_{i=1}^n \frac{1}{\varpi_i}\{(\boldsymbol{Z}_i - \boldsymbol{U}_i\boldsymbol{W}_0)_{l_1}^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)(\boldsymbol{Z}_i)_{l_2}\}^2\Big] \\
\le&E\Big[\Big(\max_{1\le i\le n}\frac{1}{n\varpi_i}\Big)\frac{1}{n}\sum_{i=1}^n \frac{1}{\varpi_i}\{(\boldsymbol{Z}_i - \boldsymbol{U}_i\boldsymbol{W}_0)_{l_1}^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)(\boldsymbol{Z}_i)_{l_2}\}^2\Big] \\
=&O(r^{-1}).
\end{aligned}
$$

Thus, we have

$$
\|\nabla_{\boldsymbol{\vartheta}} \boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R}|\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0) - \boldsymbol{\Lambda}\| = O_P(r^{-1/2}),
$$

Similar to the proof of Lemmas S.1.7 of Fang et al. (2020),

$$
\begin{aligned}
&\|\nabla_{\boldsymbol{\vartheta}} \boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \breve{\boldsymbol{W}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w, \breve{\gamma}_{\mathrm{p}}) - \nabla_{\boldsymbol{\vartheta}} \boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R}|\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)\| \\
=&O_P\Big\{\Big[\frac{(s_{\boldsymbol{W}} \vee s_{\boldsymbol{\beta}})\log p}{r^{\mathrm{p}}}\Big]^{1/2}\Big\}.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
&\|\nabla_{\boldsymbol{\vartheta}} \boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \breve{\boldsymbol{W}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w, \breve{\gamma}_{\mathrm{p}}) - \boldsymbol{\Lambda}\| \\
\le&\|\nabla_{\boldsymbol{\vartheta}} \boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \breve{\boldsymbol{W}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w, \breve{\gamma}_{\mathrm{p}}) - \nabla_{\boldsymbol{\vartheta}} \boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R}|\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)\| \\
&+ \|\nabla_{\boldsymbol{\vartheta}} \boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R}|\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0) - \boldsymbol{\Lambda}\| \\
=&O_P\Big\{\Big[\frac{(s_{\boldsymbol{W}} \vee s_{\boldsymbol{\beta}})\log p}{r^{\mathrm{p}}}\Big]^{1/2}\Big\} + O_P(r^{-1/2}) = o_P(1),
\end{aligned}
$$

under the assumption in (A.7) that $(r^{\mathrm{p}})^{-1/2}(s_{\boldsymbol{W}} \vee s_{\boldsymbol{\beta}})\log p = o(1)$.

(c)$\|\nabla_{\boldsymbol{\varphi}}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\breve{\boldsymbol{W}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w,\breve{\boldsymbol{\gamma}}_{\mathrm{p}})(\breve{\boldsymbol{\gamma}}_{\mathrm{p}}-\boldsymbol{\gamma}_0)\| = o_P(r^{-1/2})$.

To show the above result, let

$$\begin{aligned}&\nabla_{\boldsymbol{\varphi}}\boldsymbol{Q}_w^*(\boldsymbol{\vartheta},\boldsymbol{\varphi},\boldsymbol{W}_0,\boldsymbol{R}|\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0)\\ =&-\frac{1}{n}\sum_{i=1}^n\frac{\eta_i}{\varpi_i}(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0)\boldsymbol{A}_i(\boldsymbol{\vartheta},\boldsymbol{\varphi})\boldsymbol{U}_i\end{aligned}$$

Similar to the proof of Lemma S.1.5 of Fang et al. (2020),

$$\begin{aligned}&\|\nabla_{\boldsymbol{\varphi}}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R}|\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0)(\breve{\boldsymbol{\gamma}}_{\mathrm{p}}-\boldsymbol{\gamma}_0)\|\\ \le&\sqrt{d}\|\nabla_{\boldsymbol{\varphi}}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R}|\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0)\|_\infty\|\breve{\boldsymbol{\gamma}}_{\mathrm{p}}-\boldsymbol{\gamma}_0\|_1\\ =&O_P\Big\{\Big(\frac{\log p}{r}\Big)^{1/2}\times s_{\boldsymbol{\beta}}\Big(\frac{\log p}{r^{\mathrm{p}}}\Big)^{1/2}\Big\}=O_P\Big(\frac{s_{\boldsymbol{\beta}}\log p}{\sqrt{r^{\mathrm{p}}r}}\Big),\end{aligned}$$

and

$$\begin{aligned}&\Big\|\Big[\nabla_{\boldsymbol{\varphi}}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\breve{\boldsymbol{W}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w,\breve{\boldsymbol{\gamma}}_{\mathrm{p}})-\nabla_{\boldsymbol{\varphi}}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R}|\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0)\Big](\breve{\boldsymbol{\gamma}}_{\mathrm{p}}-\boldsymbol{\gamma}_0)\Big\|\\ =&O_P\Big\{\frac{(s_{\boldsymbol{W}}\vee s_{\boldsymbol{\beta}})\log p}{r^{\mathrm{p}}}\Big\}.\end{aligned}$$

Thus, we conclude that

$$\begin{aligned}&\|\nabla_{\boldsymbol{\varphi}}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\breve{\boldsymbol{W}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_w,\breve{\boldsymbol{\gamma}}_{\mathrm{p}})(\breve{\boldsymbol{\gamma}}_{\mathrm{p}}-\boldsymbol{\gamma}_0)\|\\ =&O_P\Big\{\frac{s_{\boldsymbol{\beta}}\log p}{\sqrt{r^{\mathrm{p}}r}}+\frac{(s_{\boldsymbol{W}}\vee s_{\boldsymbol{\beta}})\log p}{r^{\mathrm{p}}}\Big\}\\ =&o_P(r^{-1/2}),\end{aligned}$$

under the assumptions in (A.7) that $(r^{\mathrm{p}})^{-1/2}(s_{\boldsymbol{W}}\vee s_{\boldsymbol{\beta}})\log p=o(1)$ and $r^{1/2}(r^{\mathrm{p}})^{-1}(s_{\boldsymbol{W}}\vee s_{\boldsymbol{\beta}})\log p=o(1)$. Combining the results of (a), (b) and (c), we have $\|\breve{\boldsymbol{\theta}}_w-\boldsymbol{\theta}_0\|=O_P(r^{-1/2})$.

(ii) Now we prove the asymptotic normality. Denote

$$\begin{aligned}&\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\boldsymbol{W}_0,\boldsymbol{R}|\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0)\\=&\sum_{i=1}^n\frac{\eta_i}{n\varpi_i}(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}=:\sum_{i=1}^n\boldsymbol{\xi}_i.\end{aligned}$$

Below we check the conditions of Lindeberg-Feller central limit theorem. For $\forall\epsilon>0$,

$$\begin{aligned}&\sum_{i=1}^n E[\|\boldsymbol{\xi}_i\|^2 I(\|\boldsymbol{\xi}_i\|\geq\epsilon)]\\\leq&\frac{1}{\epsilon}\sum_{i=1}^n E(\|\boldsymbol{\xi}_i\|^3)\\=&\frac{1}{\epsilon}\sum_{i=1}^n E[E(\|\boldsymbol{\xi}_i\|^3|\mathcal{F}_n)]\\=&\frac{1}{\epsilon}\sum_{i=1}^n E\Big[\frac{1}{n^3\varpi_i^2}\|(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}\|^3\Big]\\\leq&\frac{1}{\epsilon}E\Big[\Big(\max_{1\leq i\leq n}\frac{1}{n\varpi_i}\Big)^2\frac{1}{n}\sum_{i=1}^n\|(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}\|^3\Big]\\=&o(1).\end{aligned}$$

Thus the condition of Lindeberg-Feller central limit theorem is satisfied. Note that

$$\begin{aligned}&\sum_{i=1}^n E(\boldsymbol{\xi}_i)\\=&\sum_{i=1}^n E[E(\boldsymbol{\xi}_i|\mathcal{F}_n)]=E[(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}]=\boldsymbol{0},\\&\sum_{i=1}^n Var(\boldsymbol{\xi}_i)\\=&\sum_{i=1}^n Var[E(\boldsymbol{\xi}_i|\mathcal{F}_n)]+\sum_{i=1}^n E[Var(\boldsymbol{\xi}_i|\mathcal{F}_n)]\end{aligned}$$

$$
\begin{aligned}
&=E\Big[\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{\pi_i}[(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}]^{\otimes 2}\Big]\\
&=E\Big(E\Big[\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{\pi_i}[(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}]^{\otimes 2}\Big|\{\boldsymbol{X}_i\}_{i=1}^{n}\Big]\Big)\\
&=E\Big[\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{\pi_i}(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)\Big]\\
&=\boldsymbol{\Upsilon}_{\varpi}.
\end{aligned}
$$

By the Lindeberg-Feller central limit theorem, we obtain

$$
\boldsymbol{\Upsilon}_{\varpi}^{-1/2}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\gamma_0,\boldsymbol{W}_0,\boldsymbol{R}|\boldsymbol{\theta}_0,\gamma_0)\xrightarrow{d}N(\boldsymbol{0},\boldsymbol{I}_d).
$$

According to the Taylor's expansion,

$$
\begin{aligned}
&\check{\boldsymbol{\theta}}_w-\boldsymbol{\theta}_0\\
=&-[\nabla_{\boldsymbol{\vartheta}}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\gamma_0,\check{\boldsymbol{W}}_{\mathrm{p}},\check{\boldsymbol{R}}_{\mathrm{p}}|\check{\boldsymbol{\theta}}_w,\check{\gamma}_{\mathrm{p}})]^{-1}[\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\gamma_0,\check{\boldsymbol{W}}_{\mathrm{p}},\check{\boldsymbol{R}}_{\mathrm{p}}|\check{\boldsymbol{\theta}}_w,\check{\gamma}_{\mathrm{p}})\\
&+\nabla_{\boldsymbol{\varphi}}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\gamma_0,\check{\boldsymbol{W}}_{\mathrm{p}},\check{\boldsymbol{R}}_{\mathrm{p}}|\check{\boldsymbol{\theta}}_w,\check{\gamma}_{\mathrm{p}})(\check{\gamma}_{\mathrm{p}}-\gamma_0)+\check{\boldsymbol{\mathcal{R}}}].
\end{aligned}
$$

Note that

$$
\begin{aligned}
-[\nabla_{\boldsymbol{\vartheta}}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\gamma_0,\check{\boldsymbol{W}}_{\mathrm{p}},\check{\boldsymbol{R}}_{\mathrm{p}}|\check{\boldsymbol{\theta}}_w,\check{\gamma}_{\mathrm{p}})]^{-1}&=\boldsymbol{\Lambda}^{-1}-[\nabla_{\boldsymbol{\vartheta}}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\gamma_0,\check{\boldsymbol{W}}_{\mathrm{p}},\check{\boldsymbol{R}}_{\mathrm{p}}|\check{\boldsymbol{\theta}}_w,\check{\gamma}_{\mathrm{p}})]^{-1}-\boldsymbol{\Lambda}^{-1},\\
\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\gamma_0,\check{\boldsymbol{W}}_{\mathrm{p}},\check{\boldsymbol{R}}_{\mathrm{p}}|\check{\boldsymbol{\theta}}_w,\check{\gamma}_{\mathrm{p}})&=\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\gamma_0,\check{\boldsymbol{W}}_{\mathrm{p}},\check{\boldsymbol{R}}_{\mathrm{p}}|\check{\boldsymbol{\theta}}_w,\check{\gamma}_{\mathrm{p}})-\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\gamma_0,\boldsymbol{W}_0,\boldsymbol{R}|\boldsymbol{\theta}_0,\gamma_0)\\
&\quad+\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\gamma_0,\boldsymbol{W}_0,\boldsymbol{R}|\boldsymbol{\theta}_0,\gamma_0),
\end{aligned}
$$

and we have shown that

$$
\boldsymbol{\Lambda}^{-1}-[\nabla_{\boldsymbol{\vartheta}}\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0,\gamma_0,\check{\boldsymbol{W}}_{\mathrm{p}},\check{\boldsymbol{R}}_{\mathrm{p}}|\check{\boldsymbol{\theta}}_w,\check{\gamma}_{\mathrm{p}})]^{-1}=o_P(1),
$$

$$\|\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \breve{\boldsymbol{W}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}} | \check{\boldsymbol{\theta}}_w, \breve{\boldsymbol{\gamma}}_{\mathrm{p}}) - \boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R} | \boldsymbol{\theta}_0, \boldsymbol{\gamma}_0)\| = o_P(r^{-1/2}),$$

$$\|\nabla_{\boldsymbol{\varphi}} \boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \breve{\boldsymbol{W}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}} | \check{\boldsymbol{\theta}}_w, \breve{\boldsymbol{\gamma}}_{\mathrm{p}})(\breve{\boldsymbol{\gamma}}_{\mathrm{p}} - \boldsymbol{\gamma}_0)\| = o_P(r^{-1/2}).$$

Therefore, $\check{\boldsymbol{\theta}}_w - \boldsymbol{\theta}_0 = -\boldsymbol{\Lambda}^{-1} \boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R} | \boldsymbol{\theta}_0, \boldsymbol{\gamma}_0) + o_P(1)$. Finally,

$$(\boldsymbol{\Lambda}^{-1} \boldsymbol{\Upsilon}_{\varpi} \boldsymbol{\Lambda}^{-1})^{-1/2} (\check{\boldsymbol{\theta}}_w - \boldsymbol{\theta}_0) = -(\boldsymbol{\Lambda}^{-1} \boldsymbol{\Upsilon}_{\varpi} \boldsymbol{\Lambda}^{-1})^{-1/2} \boldsymbol{\Lambda}^{-1} \boldsymbol{\Upsilon}_{\varpi}^{1/2} \boldsymbol{\Upsilon}_{\pi}^{-1/2} \boldsymbol{Q}_w^*(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R} | \boldsymbol{\theta}_0, \boldsymbol{\gamma}_0) + o_P(1),$$

and

$$(\boldsymbol{\Lambda}^{-1} \boldsymbol{\Upsilon}_{\varpi} \boldsymbol{\Lambda}^{-1})^{-1/2} (\check{\boldsymbol{\theta}}_w - \boldsymbol{\theta}_0) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I}_d).$$

□

Proof of Theorem 6. The problem that minimizing $\mathrm{tr}(\boldsymbol{D} \boldsymbol{\Lambda}^{-1} \boldsymbol{\Upsilon}_{\varpi} \boldsymbol{\Lambda}^{-1} \boldsymbol{D}^{\mathrm{T}})$, where $\boldsymbol{D} \in \mathbb{R}^{d \times d}$ is a fixed and known nonsingular matrix, is to solve the following optimization problem:

$$\min \tilde{W} = \sum_{i=1}^n \mathrm{tr}\{\frac{1}{\pi_i}[\boldsymbol{D} \boldsymbol{\Lambda}^{-1} (\boldsymbol{Z}_i - \boldsymbol{U}_i \boldsymbol{W}_0)^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0) \boldsymbol{R}^{-1} \boldsymbol{\Pi}_0 \boldsymbol{R}^{-1} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)(\boldsymbol{Z}_i - \boldsymbol{U}_i \boldsymbol{W}_0) \boldsymbol{\Lambda}^{-1} \boldsymbol{D}^{\mathrm{T}}]\}$$

$$\text{s.t } \sum_{i=1}^n \varpi_i = r, \; 0 \leq \pi_i \leq 1 \text{ for } i \in [n].$$

It is clear that the only difference between Theorem 6 and Theorem 2 is that $\boldsymbol{X}_i$ and π_i in Theorem 2 are replaced by $\boldsymbol{Z}_i - \boldsymbol{U}_i \boldsymbol{W}_0$ and ϖ_i in Theorem 6, respectively. Therefore, the proof of Theorem 6 is completed by directly applying the proof of Theorem 2. □

Proof of Theorem 7. It is easy to verify that the approximated optimal probabilities $\{\breve{\varpi}_i^{\boldsymbol{D}}\}_{i=1}^n$ still satisfies Assumption (A.8). Therefore, by the consistency of the pilot estimators $\breve{\boldsymbol{\beta}}_{\mathrm{p}}$, $\breve{\boldsymbol{R}}_{\mathrm{p}}$, $\breve{\boldsymbol{\Pi}}_{\mathrm{p}}$, $\breve{\boldsymbol{W}}_{\mathrm{p}}$ and $\breve{\boldsymbol{\Lambda}}_{\mathrm{p}}$, one can apply the proof of Theorem 5 directly and thus the consistency and asymptotic normality still holds. We only prove the derivation of $\boldsymbol{\Upsilon}^{\boldsymbol{D}}$ from $\boldsymbol{\Upsilon}_{\varpi}$ by inserting $\{\breve{\varpi}_i^{\boldsymbol{D}}\}_{i\in[n]}$. Since

$$
\begin{aligned}
&E\Big[\frac{1}{n^2}\sum_{i=1}^{n}\frac{r}{\breve{\varpi}_i^{\boldsymbol{D}}}(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)\Big]\\
=&E\Big[\frac{1}{n}\sum_{i=1}^{n}\frac{\frac{1}{n}\sum_{i=1}^{n}\breve{\hbar}_i^{\boldsymbol{D}}}{\breve{\hbar}_i^{\boldsymbol{D}}}(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)\Big]\\
\to&l^{\boldsymbol{D}}E\Big[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\hbar_i^{\boldsymbol{D}}}(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0)\Big]\\
=&l^{\boldsymbol{D}}\boldsymbol{\Upsilon}^{\boldsymbol{D}},
\end{aligned}
$$

the desired results for $\breve{\boldsymbol{\theta}}_w^{\boldsymbol{D}}$ hold. □

Before we prove Theorem 8, we present the following lemma.

Lemma 2. *Under Assumptions (A.2), (A.4)-(A.6) and (A.9)-(A.10),*

$$
\begin{aligned}
&(i)\|\boldsymbol{Q}_{uw}^{*}(\boldsymbol{\theta}_0,\gamma_0,\breve{\boldsymbol{W}}_{\mathrm{p}}^{\boldsymbol{D}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}},\breve{\gamma}_{\mathrm{p}})\|=O_P(r^{-1/2}),\\
&(ii)\|l^{\boldsymbol{D}}\nabla_{\boldsymbol{\vartheta}}\boldsymbol{Q}_{uw}^{*}(\boldsymbol{\theta}_0,\gamma_0,\breve{\boldsymbol{W}}_{\mathrm{p}}^{\boldsymbol{D}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}},\breve{\gamma}_{\mathrm{p}})-\boldsymbol{\Psi}^{\boldsymbol{D}}\|=o_P(1),\\
&(iii)\|l^{\boldsymbol{D}}\nabla_{\boldsymbol{\varphi}}\boldsymbol{Q}_{uw}^{*}(\boldsymbol{\theta}_0,\gamma_0,\breve{\boldsymbol{W}}_{\mathrm{p}}^{\boldsymbol{D}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}},\breve{\gamma}_{\mathrm{p}})(\breve{\gamma}_{\mathrm{p}}-\gamma_0)\|=o_P(r^{-1/2}).
\end{aligned}
$$

where

$$
\begin{aligned}
&\boldsymbol{Q}_{uw}^{*}(\boldsymbol{\theta}_0,\gamma_0,\breve{\boldsymbol{W}}_{\mathrm{p}}^{\boldsymbol{D}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}},\breve{\gamma}_{\mathrm{p}})\\
=&\frac{1}{n}\sum_{i=1}^{n}\eta_i^{\boldsymbol{D}}(\boldsymbol{Z}_i-\boldsymbol{U}_i\breve{\boldsymbol{W}}_{\mathrm{p}}^{\boldsymbol{D}})^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}},\breve{\gamma}_{\mathrm{p}})\breve{\boldsymbol{R}}_{\mathrm{p}}^{-1}\boldsymbol{A}_i^{-1/2}(\breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}},\breve{\gamma}_{\mathrm{p}})\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\vartheta},\boldsymbol{\varphi})\},
\end{aligned}
$$

$$\begin{aligned}
&\nabla_{\boldsymbol{\vartheta}} \boldsymbol{Q}^{*}_{uw}(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \breve{\boldsymbol{W}}^{\boldsymbol{D}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}} | \breve{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw}, \breve{\gamma}_{\mathrm{p}}) \\
=& -\frac{1}{n}\sum_{i=1}^{n} \eta_i^{\boldsymbol{D}} (\boldsymbol{Z}_i - \boldsymbol{U}_i \breve{\boldsymbol{W}}^{\boldsymbol{D}}_{\mathrm{p}})^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\breve{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw}, \breve{\gamma}_{\mathrm{p}}) \breve{\boldsymbol{R}}_{\mathrm{p}}^{-1} \boldsymbol{A}_i^{-1/2}(\breve{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw}, \breve{\gamma}_{\mathrm{p}}) \boldsymbol{A}_i(\boldsymbol{\beta}_0) \boldsymbol{Z}_i, \\
&\nabla_{\boldsymbol{\varphi}} \boldsymbol{Q}^{*}_{uw}(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0, \breve{\boldsymbol{W}}^{\boldsymbol{D}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}} | \breve{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw}, \breve{\gamma}_{\mathrm{p}}) \\
=& -\frac{1}{n}\sum_{i=1}^{n} \eta_i^{\boldsymbol{D}} (\boldsymbol{Z}_i - \boldsymbol{U}_i \breve{\boldsymbol{W}}^{\boldsymbol{D}}_{\mathrm{p}})^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\breve{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw}, \breve{\gamma}_{\mathrm{p}}) \breve{\boldsymbol{R}}_{\mathrm{p}}^{-1} \boldsymbol{A}_i^{-1/2}(\breve{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw}, \breve{\gamma}_{\mathrm{p}}) \boldsymbol{A}_i(\boldsymbol{\beta}_0) \boldsymbol{U}_i.
\end{aligned}$$

Proof of Lemma 2. The proof of (i), (ii) and (iii) are similar to the proof of (a), (b) and (c) in Theorem 5, respectively. □

Proof of Theorem 8. (i) For consistency, the first step is to show that $\|\breve{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw} - \boldsymbol{\theta}_0\| = o_P(1)$. Let

$$\begin{aligned}
&\boldsymbol{Q}^{*}_{uw}(\boldsymbol{\vartheta}, \breve{\gamma}_{\mathrm{p}}, \breve{\boldsymbol{W}}^{\boldsymbol{D}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}}) \\
=& \frac{1}{n}\sum_{i=1}^{n} \eta_i^{\boldsymbol{D}} (\boldsymbol{Z}_i - \boldsymbol{U}_i \breve{\boldsymbol{W}}^{\boldsymbol{D}}_{\mathrm{p}})^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\vartheta}, \breve{\gamma}_{\mathrm{p}}) \breve{\boldsymbol{R}}_{\mathrm{p}}^{-1} \{\boldsymbol{A}_i^{1/2}(\boldsymbol{\vartheta}, \breve{\gamma}_{\mathrm{p}})\}^{-1} \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\vartheta}, \breve{\gamma}_{\mathrm{p}})\}, \\
&\boldsymbol{Q}^{*}_{uw}(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0, \boldsymbol{W}^{\boldsymbol{D}}_{0}, \boldsymbol{R}) \\
=& \frac{1}{n}\sum_{i=1}^{n} \eta_i^{\boldsymbol{D}} (\boldsymbol{Z}_i - \boldsymbol{U}_i \boldsymbol{W}^{\boldsymbol{D}}_{0})^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0) \boldsymbol{R}^{-1} \{\boldsymbol{A}_i^{1/2}(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0)\}^{-1} \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0)\}, \\
&\boldsymbol{Q}_{uw}(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0, \boldsymbol{W}^{\boldsymbol{D}}_{0}, \boldsymbol{R}) \\
=& \frac{1}{n}\sum_{i=1}^{n} (\boldsymbol{Z}_i - \boldsymbol{U}_i \boldsymbol{W}^{\boldsymbol{D}}_{0})^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0) \boldsymbol{R}^{-1} \{\boldsymbol{A}_i^{1/2}(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0)\}^{-1} \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0)\}.
\end{aligned}$$

Similar to the proof of Theorem 5, we obtain $\boldsymbol{Q}^{*}_{uw}(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0, \boldsymbol{W}^{\boldsymbol{D}}_{0}, \boldsymbol{R}) = E[\boldsymbol{Q}_{uw}(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0, \boldsymbol{W}^{\boldsymbol{D}}_{0}, \boldsymbol{R})] + o_P(1)$. Due to the consistency of $\breve{\gamma}_{\mathrm{p}}, \breve{\boldsymbol{W}}^{\boldsymbol{D}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}}$ to $\boldsymbol{\gamma}_0, \boldsymbol{W}^{\boldsymbol{D}}_{0}, \boldsymbol{R}$ in Lemma 1, respectively, we have $\boldsymbol{Q}^{*}_{uw}(\boldsymbol{\vartheta}, \breve{\gamma}_0, \breve{\boldsymbol{W}}^{\boldsymbol{D}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}}) = \boldsymbol{Q}^{*}_{uw}(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0, \boldsymbol{W}^{\boldsymbol{D}}_{0}, \boldsymbol{R}) + o_P(1)$, and thus $\boldsymbol{Q}^{*}_{uw}(\boldsymbol{\vartheta}, \breve{\gamma}_0, \breve{\boldsymbol{W}}^{\boldsymbol{D}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}}) =$

$E[\boldsymbol{Q}_{uw}(\boldsymbol{\vartheta}, \boldsymbol{\gamma}_0, \boldsymbol{W}_0^{\boldsymbol{D}}, \boldsymbol{R})] + o_P(1)$. Applying Theorem 5.9 of van der Vaart (1998), it yields $\|\breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}} - \boldsymbol{\theta}_0\| = o_P(1)$. Next, using Taylor's theorem,

$$\begin{aligned}
\boldsymbol{0} &= \boldsymbol{Q}_{uw}^*(\breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}}, \breve{\gamma}_{\mathrm{p}}, \breve{\boldsymbol{W}}_{\mathrm{p}}^{\boldsymbol{D}}, \breve{\boldsymbol{R}}_{\mathrm{p}} | \breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}}, \breve{\gamma}_{\mathrm{p}}) \\
&= \boldsymbol{Q}_{uw}^*(\boldsymbol{\theta}_0, \gamma_0, \breve{\boldsymbol{W}}_{\mathrm{p}}^{\boldsymbol{D}}, \breve{\boldsymbol{R}}_{\mathrm{p}} | \breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}}, \breve{\gamma}_{\mathrm{p}}) + \nabla_{\boldsymbol{\vartheta}} \boldsymbol{Q}_{uw}^*(\boldsymbol{\theta}_0, \gamma_0, \breve{\boldsymbol{W}}_{\mathrm{p}}^{\boldsymbol{D}}, \breve{\boldsymbol{R}}_{\mathrm{p}} | \breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}}, \breve{\gamma}_{\mathrm{p}})(\breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}} - \boldsymbol{\theta}_0) \\
&\quad + \nabla_{\boldsymbol{\varphi}} \boldsymbol{Q}_{uw}^*(\boldsymbol{\theta}_0, \gamma_0, \breve{\boldsymbol{W}}_{\mathrm{p}}^{\boldsymbol{D}}, \breve{\boldsymbol{R}}_{\mathrm{p}} | \breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}}, \breve{\gamma}_{\mathrm{p}})(\breve{\gamma}_{\mathrm{p}} - \gamma_0) + \breve{\boldsymbol{\mathcal{R}}}
\end{aligned}$$

where $\breve{\boldsymbol{\mathcal{R}}} = O_P(\|\breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}} - \boldsymbol{\theta}_0\|^2 + \|\breve{\gamma}_{\mathrm{p}} - \boldsymbol{\gamma}_0\|^2)$. According to Lemma 2, we conclude $\|\breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}} - \boldsymbol{\theta}_0\| = O_P(r^{-1/2})$.

(ii) Below we prove the asymptotic normality. Let

$$\begin{aligned}
& l^{\boldsymbol{D}} \boldsymbol{Q}_{uw}^*(\boldsymbol{\theta}_0; \gamma_0, \boldsymbol{W}_0^{\boldsymbol{D}}, \boldsymbol{R}) \\
=& \sum_{i=1}^n \frac{l^{\boldsymbol{D}} \eta_i^{\boldsymbol{D}}}{n} (\boldsymbol{Z}_i - \boldsymbol{U}_i \boldsymbol{W}_0^{\boldsymbol{D}})^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0) \boldsymbol{R}^{-1} \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0) \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\} := \sum_{i=1}^n \boldsymbol{\xi}_i^{\boldsymbol{D}}.
\end{aligned}$$

Now, we check the conditions of Lindeberg-Feller central limit theorem. For $\forall \epsilon > 0$,

$$\begin{aligned}
& \sum_{i=1}^n E[\|\boldsymbol{\xi}_i^{\boldsymbol{D}}\|^2 I(\|\boldsymbol{\xi}_i\| \geq \epsilon)] \\
\leq & \frac{1}{\epsilon} \sum_{i=1}^n E(\|\boldsymbol{\xi}_i^{\boldsymbol{D}}\|^3) \\
= & \frac{1}{\epsilon} \sum_{i=1}^n E[E(\|\boldsymbol{\xi}_i^{\boldsymbol{D}}\|^3 | \mathcal{F}_n)] \\
\leq & \frac{(l^{\boldsymbol{D}})^3}{\epsilon} \sum_{i=1}^n E\Big[\frac{1}{n^3} \|(\boldsymbol{Z}_i - \boldsymbol{U}_i \boldsymbol{W}_0^{\boldsymbol{D}})^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0) \boldsymbol{R}^{-1} \boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0) \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}\|^3\Big] \\
= & o(1).
\end{aligned}$$

Thus the condition of Lindeberg-Feller central limit theorem is satisfied. Note that

$$
\begin{aligned}
& E[l^{\boldsymbol{D}}\boldsymbol{Q}_{uw}^{*}(\boldsymbol{\theta}_0;\gamma_0,\boldsymbol{W}_0^{\boldsymbol{D}},\boldsymbol{R})] \\
=& E\{E[l^{\boldsymbol{D}}\boldsymbol{Q}_{uw}^{*}(\boldsymbol{\theta}_0;\gamma_0,\boldsymbol{W}_0^{\boldsymbol{D}},\boldsymbol{R})|\mathcal{F}_n]\} \\
=& E\Big[\sum_{i=1}^{n}\frac{l^{\boldsymbol{D}}\breve{\varpi}_i^{\boldsymbol{D}}}{n}(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0^{\boldsymbol{D}})^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}\Big] \\
=& E\Big\{E\Big[\sum_{i=1}^{n}\frac{l^{\boldsymbol{D}}\breve{\varpi}_i^{\boldsymbol{D}}}{n}(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0^{\boldsymbol{D}})^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}\Big|\{\boldsymbol{X}_i\}_{i=1}^{n}\Big]\Big\} \\
=& \boldsymbol{0},
\end{aligned}
$$

and

$$
\begin{aligned}
& Var[l^{\boldsymbol{D}}\boldsymbol{Q}_{uw}^{*}(\boldsymbol{\theta}_0;\boldsymbol{\gamma}_0,\boldsymbol{W}_0^{\boldsymbol{D}},\boldsymbol{R})] \\
=& E\{Var[l^{\boldsymbol{D}}\boldsymbol{Q}_{uw}^{*}(\boldsymbol{\theta}_0;\boldsymbol{\gamma}_0,\boldsymbol{W}_0^{\boldsymbol{D}},\boldsymbol{R})|\mathcal{F}_n]\}+Var\{E[l^{\boldsymbol{D}}\boldsymbol{Q}_{uw}^{*}(\boldsymbol{\theta}_0;\boldsymbol{\gamma}_0,\boldsymbol{W}_0^{\boldsymbol{D}},\boldsymbol{R})|\mathcal{F}_n]\} \\
=& E\Big\{\frac{(l^{\boldsymbol{D}})^2}{n^2}\sum_{i=1}^{n}\breve{\varpi}_i^{\boldsymbol{D}}(1-\breve{\varpi}_i^{\boldsymbol{D}})[(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0^{\boldsymbol{D}})^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}]^{\otimes 2}\Big\} \\
& +Var\Big[\frac{l^{\boldsymbol{D}}}{n}\sum_{i=1}^{n}\breve{\varpi}_i^{\boldsymbol{D}}(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0^{\boldsymbol{D}})^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}\Big].
\end{aligned}
$$

To proceed, by using the iterated expectation and iterated variance formula again,

$$
\begin{aligned}
& E\Big\{\frac{(l^{\boldsymbol{D}})^2}{n^2}\sum_{i=1}^{n}\breve{\varpi}_i^{\boldsymbol{D}}(1-\breve{\varpi}_i^{\boldsymbol{D}})[(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0^{\boldsymbol{D}})^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}]^{\otimes 2}\Big\} \\
=& E\Big(E\Big\{\frac{(l^{\boldsymbol{D}})^2}{n^2}\sum_{i=1}^{n}\breve{\varpi}_i^{\boldsymbol{D}}(1-\breve{\varpi}_i^{\boldsymbol{D}}) \\
& \qquad [(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0^{\boldsymbol{D}})^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{-1/2}(\boldsymbol{\beta}_0)\{\boldsymbol{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_0)\}]^{\otimes 2}\Big|\{\boldsymbol{X}_i\}_{i=1}^{n}\Big\}\Big) \\
=& E\Big[\frac{(l^{\boldsymbol{D}})^2}{n^2}\sum_{i=1}^{n}\breve{\varpi}_i^{\boldsymbol{D}}(1-\breve{\varpi}_i^{\boldsymbol{D}})(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0^{\boldsymbol{D}})^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0^{\boldsymbol{D}})\Big],
\end{aligned}
$$

$$
\begin{aligned}
&Var\Big[\frac{l^{\boldsymbol D}}{n}\sum_{i=1}^{n}\breve{\varpi}_i^{\boldsymbol D}(\boldsymbol Z_i-\boldsymbol U_i\boldsymbol W_0^{\boldsymbol D})^{\rm T}\boldsymbol A_i^{1/2}(\boldsymbol\beta_0)\boldsymbol R^{-1}\boldsymbol A_i^{-1/2}(\boldsymbol\beta_0)\{\boldsymbol Y_i-\boldsymbol\mu_i(\boldsymbol\beta_0)\}\Big]\\
=&E\Big\{Var\Big[\frac{l^{\boldsymbol D}}{n}\sum_{i=1}^{n}\breve{\varpi}_i^{\boldsymbol D}(\boldsymbol Z_i-\boldsymbol U_i\boldsymbol W_0^{\boldsymbol D})^{\rm T}\boldsymbol A_i^{1/2}(\boldsymbol\beta_0)\boldsymbol R^{-1}\boldsymbol A_i^{-1/2}(\boldsymbol\beta_0)\{\boldsymbol Y_i-\boldsymbol\mu_i(\boldsymbol\beta_0)\}\Big|\{\boldsymbol X_i\}_{i=1}^{n}\Big]\Big\}\\
&+Var\Big\{E\Big[\frac{l^{\boldsymbol D}}{n}\sum_{i=1}^{n}\breve{\varpi}_i^{\boldsymbol D}(\boldsymbol Z_i-\boldsymbol U_i\boldsymbol W_0^{\boldsymbol D})^{\rm T}\boldsymbol A_i^{1/2}(\boldsymbol\beta_0)\boldsymbol R^{-1}\boldsymbol A_i^{-1/2}(\boldsymbol\beta_0)\{\boldsymbol Y_i-\boldsymbol\mu_i(\boldsymbol\beta_0)\}\Big|\{\boldsymbol X_i\}_{i=1}^{n}\Big]\Big\}\\
=&E\Big[\frac{(l^{\boldsymbol D})^2}{n^2}\sum_{i=1}^{n}(\breve{\varpi}_i^{\boldsymbol D})^2(\boldsymbol Z_i-\boldsymbol U_i\boldsymbol W_0^{\boldsymbol D})^{\rm T}\boldsymbol A_i^{1/2}(\boldsymbol\beta_0)\boldsymbol R^{-1}\boldsymbol\Pi_0\boldsymbol R^{-1}\boldsymbol A_i^{1/2}(\boldsymbol\beta_0)(\boldsymbol Z_i-\boldsymbol U_i\boldsymbol W_0^{\boldsymbol D})\Big].
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
&Var[l^{\boldsymbol D}\boldsymbol Q_{uw}^{*}(\boldsymbol\theta_0;\boldsymbol\gamma_0,\boldsymbol W_0^{\boldsymbol D},\boldsymbol R)]\\
=&E\Big[\frac{(l^{\boldsymbol D})^2}{n^2}\sum_{i=1}^{n}\breve{\varpi}_i^{\boldsymbol D}(\boldsymbol Z_i-\boldsymbol U_i\boldsymbol W_0^{\boldsymbol D})^{\rm T}\boldsymbol A_i^{1/2}(\boldsymbol\beta_0)\boldsymbol R^{-1}\boldsymbol\Pi_0\boldsymbol R^{-1}\boldsymbol A_i^{1/2}(\boldsymbol\beta_0)(\boldsymbol Z_i-\boldsymbol U_i\boldsymbol W_0^{\boldsymbol D})\Big]\\
\to& l^{\boldsymbol D}\boldsymbol\Omega^{\boldsymbol D}.
\end{aligned}
$$

By the Lindeberg-Feller central limit theorem,

$$
(l^{\boldsymbol D}\boldsymbol\Omega^{\boldsymbol D})^{-1/2}[l^{\boldsymbol D}\boldsymbol Q_{uw}^{*}(\boldsymbol\theta_0;\boldsymbol\gamma_0,\boldsymbol W_0^{\boldsymbol D},\boldsymbol R)]\xrightarrow{d}N(\boldsymbol 0,\boldsymbol I_d).
$$

According to the Taylor's expansion,

$$
\begin{aligned}
&\breve{\boldsymbol\theta}_{uw}^{\boldsymbol D}-\boldsymbol\theta_0\\
=&-[\nabla_{\boldsymbol\vartheta}\boldsymbol Q_{uw}^{*}(\boldsymbol\theta_0,\boldsymbol\gamma_0,\breve{\boldsymbol W}_{\rm p}^{\boldsymbol D},\breve{\boldsymbol R}_{\rm p}|\breve{\boldsymbol\theta}_{uw}^{\boldsymbol D},\breve{\gamma}_{\rm p})]^{-1}[\boldsymbol Q_{uw}^{*}(\boldsymbol\theta_0,\boldsymbol\gamma_0,\breve{\boldsymbol W}_{\rm p}^{\boldsymbol D},\breve{\boldsymbol R}_{\rm p}|\breve{\boldsymbol\theta}_{uw}^{\boldsymbol D},\breve{\gamma}_{\rm p})\\
&+\nabla_{\boldsymbol\varphi}\boldsymbol Q_{uw}^{*}(\boldsymbol\theta_0,\boldsymbol\gamma_0,\breve{\boldsymbol W}_{\rm p}^{\boldsymbol D},\breve{\boldsymbol R}_{\rm p}|\breve{\boldsymbol\theta}_{uw}^{\boldsymbol D},\breve{\gamma}_{\rm p})(\breve{\gamma}_{\rm p}-\boldsymbol\gamma_0)+\boldsymbol{\mathcal R}],
\end{aligned}
$$

Note that

$$
-[l^{\boldsymbol D}\nabla_{\boldsymbol\vartheta}\boldsymbol Q_{uw}^{*}(\boldsymbol\theta_0,\boldsymbol\gamma_0,\breve{\boldsymbol W}_{\rm p}^{\boldsymbol D},\breve{\boldsymbol R}_{\rm p}|\breve{\boldsymbol\theta}_{uw}^{\boldsymbol D},\breve{\gamma}_{\rm p})]^{-1}
$$

$$
\begin{aligned}
&=(\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1} - [l^{\boldsymbol{D}}\nabla_{\boldsymbol{\vartheta}}\boldsymbol{Q}^*_{uw}(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\breve{\boldsymbol{W}}^{\boldsymbol{D}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw},\breve{\boldsymbol{\gamma}}_{\mathrm{p}})]^{-1} - (\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1},\\
&l^{\boldsymbol{D}}\boldsymbol{Q}^*_{uw}(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\breve{\boldsymbol{W}}^{\boldsymbol{D}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw},\breve{\boldsymbol{\gamma}}_{\mathrm{p}})\\
&=l^{\boldsymbol{D}}\boldsymbol{Q}^*_{uw}(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\breve{\boldsymbol{W}}^{\boldsymbol{D}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw},\breve{\boldsymbol{\gamma}}_{\mathrm{p}}) - l^{\boldsymbol{D}}\boldsymbol{Q}^*_{uw}(\boldsymbol{\theta}_0;\boldsymbol{\gamma}_0,\boldsymbol{W}^{\boldsymbol{D}}_0,\boldsymbol{R}) + l^{\boldsymbol{D}}\boldsymbol{Q}^*_{uw}(\boldsymbol{\theta}_0;\boldsymbol{\gamma}_0,\boldsymbol{W}^{\boldsymbol{D}}_0,\boldsymbol{R}),
\end{aligned}
$$

and we have shown that

$$
\begin{aligned}
(\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1} - [l^{\boldsymbol{D}}\nabla_{\boldsymbol{\vartheta}}\boldsymbol{Q}^*_{uw}(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\breve{\boldsymbol{W}}^{\boldsymbol{D}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw},\breve{\boldsymbol{\gamma}}_{\mathrm{p}})]^{-1} &= o_P(1),\\
l^{\boldsymbol{D}}\boldsymbol{Q}^*_{uw}(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\breve{\boldsymbol{W}}^{\boldsymbol{D}}_{\mathrm{p}},\breve{\boldsymbol{R}}_{\mathrm{p}}|\breve{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw},\breve{\boldsymbol{\gamma}}_{\mathrm{p}}) - l^{\boldsymbol{D}}\boldsymbol{Q}^*_{uw}(\boldsymbol{\theta}_0;\boldsymbol{\gamma}_0,\boldsymbol{W}^{\boldsymbol{D}}_0,\boldsymbol{R}) &= o_P(r^{-1/2}).
\end{aligned}
$$

Therefore, $\breve{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw} - \boldsymbol{\theta}_0 = -(\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1}l^{\boldsymbol{D}}\boldsymbol{Q}^*_{uw}(\boldsymbol{\theta}_0;\boldsymbol{\gamma}_0,\boldsymbol{W}^{\boldsymbol{D}}_0,\boldsymbol{R}) + o_P(1)$. Finally,

$$
\begin{aligned}
&[l^{\boldsymbol{D}}(\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1}\boldsymbol{\Omega}^{\boldsymbol{D}}(\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1}]^{-1/2}(\breve{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw} - \boldsymbol{\theta}_0)\\
&=-[l^{\boldsymbol{D}}(\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1}\boldsymbol{\Omega}^{\boldsymbol{D}}(\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1}]^{-1/2}(\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1}(l^{\boldsymbol{D}}\boldsymbol{\Omega}^{\boldsymbol{D}})^{1/2}(l^{\boldsymbol{D}}\boldsymbol{\Omega}^{\boldsymbol{D}})^{-1/2}[l^{\boldsymbol{D}}\boldsymbol{Q}^*_{uw}(\boldsymbol{\theta}_0;\boldsymbol{\gamma}_0,\boldsymbol{W}^{\boldsymbol{D}}_0,\boldsymbol{R})] + o_P(1),
\end{aligned}
$$

and

$$[l^{\boldsymbol{D}}(\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1}\boldsymbol{\Omega}^{\boldsymbol{D}}(\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1}]^{-1/2}(\breve{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\boldsymbol{0},\boldsymbol{I}_d).$$

(iii) Finally, we prove that when $\boldsymbol{R} = \boldsymbol{\Pi}_0$, the inequality $(\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1}\boldsymbol{\Omega}^{\boldsymbol{D}}(\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1} \leq \boldsymbol{\Lambda}^{-1}\boldsymbol{\Upsilon}^{\boldsymbol{D}}\boldsymbol{\Lambda}^{-1}$ holds under the Loewner ordering. A key observation is that $(\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1}\boldsymbol{\Omega}^{\boldsymbol{D}}(\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1} = (\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1}$ when $\boldsymbol{R} = \boldsymbol{\Pi}_0$. Since $(\boldsymbol{Z}_i - \boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)(\boldsymbol{Z}_i - \boldsymbol{U}_i\boldsymbol{W}_0) := \boldsymbol{V}_{1i}$ and $(\boldsymbol{Z}_i - \boldsymbol{U}_i\boldsymbol{W}^{\boldsymbol{D}}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)(\boldsymbol{Z}_i - \boldsymbol{U}_i\boldsymbol{W}^{\boldsymbol{D}}_0) := \boldsymbol{V}_{2i}$ are both real symmetric positive-definite matrices, Cholesky decomposition gives $\boldsymbol{V}_{1i} = \mathbf{v}_{1i}\mathbf{v}_{1i}^{\mathrm{T}}$ and $\boldsymbol{V}_{2i} = \mathbf{v}_{2i}\mathbf{v}_{2i}^{\mathrm{T}}$ for some $\mathbf{v}_{1i}$, $\mathbf{v}_{2i} \in \mathbb{R}^{d\times d}$. We then only need to prove

$$E(\hbar^{\boldsymbol{D}}_i\mathbf{v}_{2i}\mathbf{v}_{2i}^{\mathrm{T}})^{-1} \leq E(\mathbf{v}_{1i}\mathbf{v}_{1i}^{\mathrm{T}})^{-1}E(\mathbf{v}_{1i}\mathbf{v}_{1i}^{\mathrm{T}}/\hbar^{\boldsymbol{D}}_i)E(\mathbf{v}_{1i}\mathbf{v}_{1i}^{\mathrm{T}})^{-1}.$$

Denote

$$\mathbf{f}_i := \sqrt{\hbar_i^{\boldsymbol{D}}}E(\hbar_i^{\boldsymbol{D}}\mathbf{v}_{2i}\mathbf{v}_{2i}^{\mathrm{T}})^{-1}\mathbf{v}_{2i} - \frac{1}{\sqrt{\hbar_i^{\boldsymbol{D}}}}E(\mathbf{v}_{1i}\mathbf{v}_{1i}^{\mathrm{T}})^{-1}\mathbf{v}_{1i}.$$

Since $\mathbf{f}_i\mathbf{f}_i^{\mathrm{T}} \geq 0$, we have

$$\begin{aligned} 0 \leq E(\mathbf{f}_i\mathbf{f}_i^{\mathrm{T}}) =& E(\mathbf{v}_{1i}\mathbf{v}_{1i}^{\mathrm{T}})^{-1}E(\mathbf{v}_{1i}\mathbf{v}_{1i}^{\mathrm{T}}/\hbar_i^{\boldsymbol{D}})E(\mathbf{v}_{1i}\mathbf{v}_{1i}^{\mathrm{T}})^{-1} + E(\hbar_i^{\boldsymbol{D}}\mathbf{v}_{2i}\mathbf{v}_{2i}^{\mathrm{T}})^{-1} \\ & - 2E(\mathbf{v}_{1i}\mathbf{v}_{1i}^{\mathrm{T}})^{-1}E(\mathbf{v}_{1i}\mathbf{v}_{2i}^{\mathrm{T}})E(\hbar_i^{\boldsymbol{D}}\mathbf{v}_{2i}\mathbf{v}_{2i}^{\mathrm{T}})^{-1}. \end{aligned}$$

Note that

$$\begin{aligned} & E(\mathbf{v}_{1i}\mathbf{v}_{1i}^{\mathrm{T}})^{-1}E(\mathbf{v}_{1i}\mathbf{v}_{2i}^{\mathrm{T}})E(\hbar_i^{\boldsymbol{D}}\mathbf{v}_{2i}\mathbf{v}_{2i}^{\mathrm{T}})^{-1} \\ =& E(\mathbf{v}_{1i}\mathbf{v}_{1i}^{\mathrm{T}})^{-1}E(\mathbf{v}_{1i}\mathbf{v}_{1i}^{\mathrm{T}})E(\hbar_i^{\boldsymbol{D}}\mathbf{v}_{2i}\mathbf{v}_{2i}^{\mathrm{T}})^{-1} + E(\mathbf{v}_{1i}\mathbf{v}_{1i}^{\mathrm{T}})^{-1}E(\mathbf{v}_{1i}(\mathbf{v}_{2i}-\mathbf{v}_{1i})^{\mathrm{T}})E(\hbar_i^{\boldsymbol{D}}\mathbf{v}_{2i}\mathbf{v}_{2i}^{\mathrm{T}})^{-1} \\ =& E(\hbar_i^{\boldsymbol{D}}\mathbf{v}_{2i}\mathbf{v}_{2i}^{\mathrm{T}})^{-1} + E(\mathbf{v}_{1i}\mathbf{v}_{1i}^{\mathrm{T}})^{-1}E[(\boldsymbol{Z}_i - \boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}\boldsymbol{U}_i(\boldsymbol{W}_0 - \boldsymbol{W}_0^{\boldsymbol{D}})]E(\hbar_i^{\boldsymbol{D}}\mathbf{v}_{2i}\mathbf{v}_{2i}^{\mathrm{T}})^{-1} \\ =& E(\hbar_i^{\boldsymbol{D}}\mathbf{v}_{2i}\mathbf{v}_{2i}^{\mathrm{T}})^{-1} \end{aligned}$$

Therefore the proof of Theorem 8 is completed. $\square$

S3 Additional simulation results

S3.1 Simulation results under HD-scenario

Tables S1 and S2 summarizes the bias, standard deviation (SD), estimated standard error (SE) and average coverage probability (ACP) of 95% confidence intervals with respect to the true parameter for both linear regression and logistic regression under HD-scenario based on 500 replications.

Table S1: Bias (×100), SE, SD, and ACP for linear regression under the HD-scenario.

Model	r		$\hat{\boldsymbol{\theta}}^{\mathrm{U}}$	$\hat{\boldsymbol{\theta}}_{\mathrm{d}}^{\mathrm{A}}$	$\hat{\boldsymbol{\theta}}_{\mathrm{d}}^{\mathrm{L}}$	$\hat{\boldsymbol{\theta}}_{\mathrm{w}}^{\mathrm{A}}$	$\hat{\boldsymbol{\theta}}_{\mathrm{w}}^{\mathrm{L}}$	$\hat{\boldsymbol{\theta}}_{\mathrm{uw}}^{\mathrm{A}}$	$\hat{\boldsymbol{\theta}}_{\mathrm{uw}}^{\mathrm{L}}$
IND	400	Bias	2.078	3.652	4.454	2.468	2.324	2.177	2.352
		SE	0.244	0.176	0.187	0.202	0.212	0.198	0.205
		SD	0.242	0.168	0.172	0.201	0.211	0.211	0.225
		ACP	0.960	0.955	0.970	0.935	0.960	0.935	0.920
	600	Bias	2.463	2.464	2.706	2.200	2.308	2.424	2.308
		SE	0.199	0.143	0.152	0.164	0.173	0.161	0.167
		SD	0.200	0.139	0.151	0.162	0.166	0.167	0.167
		ACP	0.940	0.955	0.940	0.960	0.960	0.945	0.960
	800	Bias	2.581	3.214	2.477	2.714	2.373	2.981	2.114
		SE	0.173	0.123	0.131	0.142	0.149	0.139	0.144
		SD	0.163	0.117	0.128	0.142	0.143	0.148	0.144
		ACP	0.965	0.965	0.935	0.955	0.940	0.945	0.950
	1000	Bias	2.125	3.573	1.998	2.071	2.310	1.934	2.122
		SE	0.154	0.110	0.117	0.127	0.133	0.125	0.129
		SD	0.145	0.101	0.113	0.128	0.134	0.134	0.140
		ACP	0.970	0.950	0.950	0.960	0.950	0.935	0.910
AR	400	Bias	1.131	1.030	1.284	0.815	0.900	0.622	0.597
		SE	0.145	0.105	0.112	0.120	0.126	0.108	0.115
		SD	0.141	0.101	0.106	0.118	0.128	0.104	0.116
		ACP	0.950	0.960	0.965	0.965	0.940	0.965	0.945
	600	Bias	1.478	1.364	1.037	0.905	1.143	1.074	0.736
		SE	0.118	0.086	0.091	0.098	0.103	0.088	0.094
		SD	0.117	0.083	0.090	0.099	0.101	0.089	0.096
		ACP	0.960	0.960	0.940	0.960	0.950	0.940	0.940
	800	Bias	1.510	0.909	0.860	1.164	1.165	1.274	0.642
		SE	0.103	0.074	0.079	0.085	0.089	0.076	0.081
		SD	0.097	0.074	0.076	0.088	0.088	0.077	0.084
		ACP	0.955	0.945	0.975	0.940	0.945	0.945	0.935
	1000	Bias	0.942	0.940	0.945	1.226	1.226	1.232	0.691
		SE	0.091	0.066	0.070	0.076	0.080	0.068	0.073
		SD	0.089	0.065	0.070	0.078	0.080	0.071	0.075
		ACP	0.945	0.940	0.955	0.950	0.965	0.940	0.935
CS	400	Bias	0.875	0.885	1.427	0.929	0.693	0.646	0.177
		SE	0.132	0.100	0.106	0.113	0.118	0.100	0.104
		SD	0.132	0.094	0.100	0.112	0.118	0.100	0.108
		ACP	0.945	0.965	0.960	0.960	0.945	0.940	0.935
	600	Bias	1.135	0.927	1.010	0.827	0.774	0.376	0.414
		SE	0.108	0.081	0.087	0.092	0.096	0.081	0.085
		SD	0.108	0.080	0.080	0.094	0.096	0.086	0.089
		ACP	0.945	0.950	0.975	0.940	0.940	0.940	0.935
	800	Bias	1.130	0.556	0.718	0.732	0.833	0.754	0.359
		SE	0.094	0.070	0.075	0.079	0.083	0.070	0.073
		SD	0.087	0.067	0.070	0.085	0.084	0.077	0.080
		ACP	0.960	0.955	0.980	0.945	0.940	0.915	0.930
	1000	Bias	0.559	0.555	0.488	0.855	0.793	0.802	0.198
		SE	0.084	0.063	0.067	0.071	0.074	0.063	0.065
		SD	0.081	0.057	0.061	0.070	0.073	0.065	0.069
		ACP	0.950	0.970	0.960	0.955	0.945	0.945	0.935

Table S2: Bias (×100), SE, SD, and ACP for logistic regression under the HD-scenario.

Model	r		$\hat{\boldsymbol{\theta}}^{\mathrm{U}}$	$\hat{\boldsymbol{\theta}}_{\mathrm{d}}^{\mathrm{A}}$	$\hat{\boldsymbol{\theta}}_{\mathrm{d}}^{\mathrm{L}}$	$\hat{\boldsymbol{\theta}}_{\mathrm{w}}^{\mathrm{A}}$	$\hat{\boldsymbol{\theta}}_{\mathrm{w}}^{\mathrm{L}}$	$\hat{\boldsymbol{\theta}}_{\mathrm{uw}}^{\mathrm{A}}$	$\hat{\boldsymbol{\theta}}_{\mathrm{uw}}^{\mathrm{L}}$
IND	400	Bias	0.922	0.736	1.494	1.585	1.480	1.920	1.473
		SE	0.195	0.172	0.182	0.182	0.187	0.176	0.179
		SD	0.188	0.162	0.168	0.177	0.191	0.171	0.174
		ACP	0.960	0.955	0.960	0.955	0.955	0.965	0.965
	600	Bias	1.766	1.043	1.521	1.658	1.900	1.813	2.270
		SE	0.160	0.140	0.148	0.148	0.153	0.143	0.146
		SD	0.156	0.137	0.141	0.146	0.157	0.140	0.148
		ACP	0.950	0.965	0.965	0.955	0.950	0.945	0.940
	800	Bias	2.107	1.357	1.956	1.516	1.843	1.699	2.130
		SE	0.138	0.121	0.127	0.128	0.132	0.124	0.127
		SD	0.129	0.123	0.124	0.126	0.126	0.122	0.121
		ACP	0.955	0.950	0.960	0.965	0.955	0.960	0.960
	1000	Bias	1.613	1.814	2.035	1.247	2.045	1.358	2.232
		SE	0.124	0.108	0.114	0.114	0.118	0.111	0.113
		SD	0.120	0.105	0.117	0.111	0.109	0.111	0.106
		ACP	0.975	0.960	0.950	0.955	0.965	0.960	0.985
AR	400	Bias	1.099	1.299	1.527	1.415	1.688	1.526	2.608
		SE	0.177	0.150	0.158	0.170	0.176	0.161	0.165
		SD	0.163	0.143	0.152	0.152	0.172	0.143	0.155
		ACP	0.970	0.975	0.955	0.965	0.955	0.970	0.965
	600	Bias	2.205	1.831	2.009	1.823	1.845	1.867	2.564
		SE	0.144	0.122	0.128	0.138	0.143	0.131	0.134
		SD	0.140	0.123	0.123	0.131	0.134	0.122	0.126
		ACP	0.955	0.965	0.955	0.960	0.960	0.955	0.955
	800	Bias	2.463	1.848	1.960	1.783	2.013	1.629	2.634
		SE	0.125	0.105	0.111	0.120	0.124	0.114	0.116
		SD	0.118	0.105	0.114	0.110	0.112	0.108	0.107
		ACP	0.955	0.955	0.925	0.970	0.980	0.965	0.970
	1000	Bias	1.851	1.837	1.898	1.752	1.845	1.652	2.571
		SE	0.112	0.094	0.099	0.107	0.111	0.102	0.104
		SD	0.110	0.096	0.097	0.102	0.105	0.097	0.101
		ACP	0.965	0.965	0.955	0.965	0.945	0.960	0.950
CS	400	Bias	1.214	0.779	1.816	1.598	1.991	1.943	2.332
		SE	0.172	0.145	0.153	0.167	0.172	0.158	0.162
		SD	0.161	0.140	0.147	0.151	0.169	0.142	0.153
		ACP	0.960	0.970	0.975	0.960	0.955	0.965	0.960
	600	Bias	2.306	1.774	1.943	1.778	1.506	2.104	1.763
		SE	0.141	0.118	0.124	0.136	0.141	0.128	0.132
		SD	0.137	0.116	0.124	0.127	0.133	0.117	0.124
		ACP	0.945	0.950	0.960	0.955	0.965	0.965	0.955
	800	Bias	2.541	1.911	1.915	1.741	1.976	1.809	2.112
		SE	0.122	0.102	0.107	0.118	0.122	0.111	0.115
		SD	0.114	0.099	0.108	0.109	0.111	0.106	0.103
		ACP	0.965	0.975	0.950	0.970	0.960	0.970	0.965
	1000	Bias	1.983	1.936	1.634	1.728	1.972	1.742	2.089
		SE	0.109	0.091	0.096	0.105	0.109	0.099	0.102
		SD	0.106	0.093	0.092	0.099	0.104	0.095	0.100
		ACP	0.965	0.935	0.965	0.975	0.960	0.960	0.955

S3.2 Effects of the allocation between r^{p} and r

We also provide an empirical guideline to choose r^{p} by minimizing the MSEs with respect to $\rho = r^{\mathrm{p}}/(r^{\mathrm{p}} + r)$ when $r^{\mathrm{p}} + r$ is fixed. Figure S1 shows the MSEs of our proposed L-optimal subsample estimators versus ρ with a fixed $r^{\mathrm{p}} + r = 100$ under LD-scenario and $r^{\mathrm{p}} + r = 500$ under HD-scenario for logistic regression, while the other settings remain the same as in Section 4.2. It can be seen that the two-step algorithm has the smallest MES when ρ is around 0.2. As ρ keeps increasing, the performance of all methods deteriorate severely. This behavior can be explained by two key factors: when r^{p} is too small, the initial estimate is inaccurate; conversely, when r^{p} becomes too large, the second-step subsample r becomes too small, reducing overall accuracy. Intuitively, the performance of our proposed estimator mainly depends on the more informative subsample drawn in the second step, while r^{p} contributes to an estimate of the population parameter. Based on our simulation results, in practice we may set $r^{\mathrm{p}} \approx 0.2(r^{\mathrm{p}} + r)$, and the rest is used to guarantee a satisfactory final estimator.

S3.3 Computation time

Table S3 records the average computational time (in seconds) based on 100 samplings under the LD-scenario of linear regression and CS structure described in Section 4.1, where T_1, T_2 and T_3 denote the computational time of calculating pilot estimators

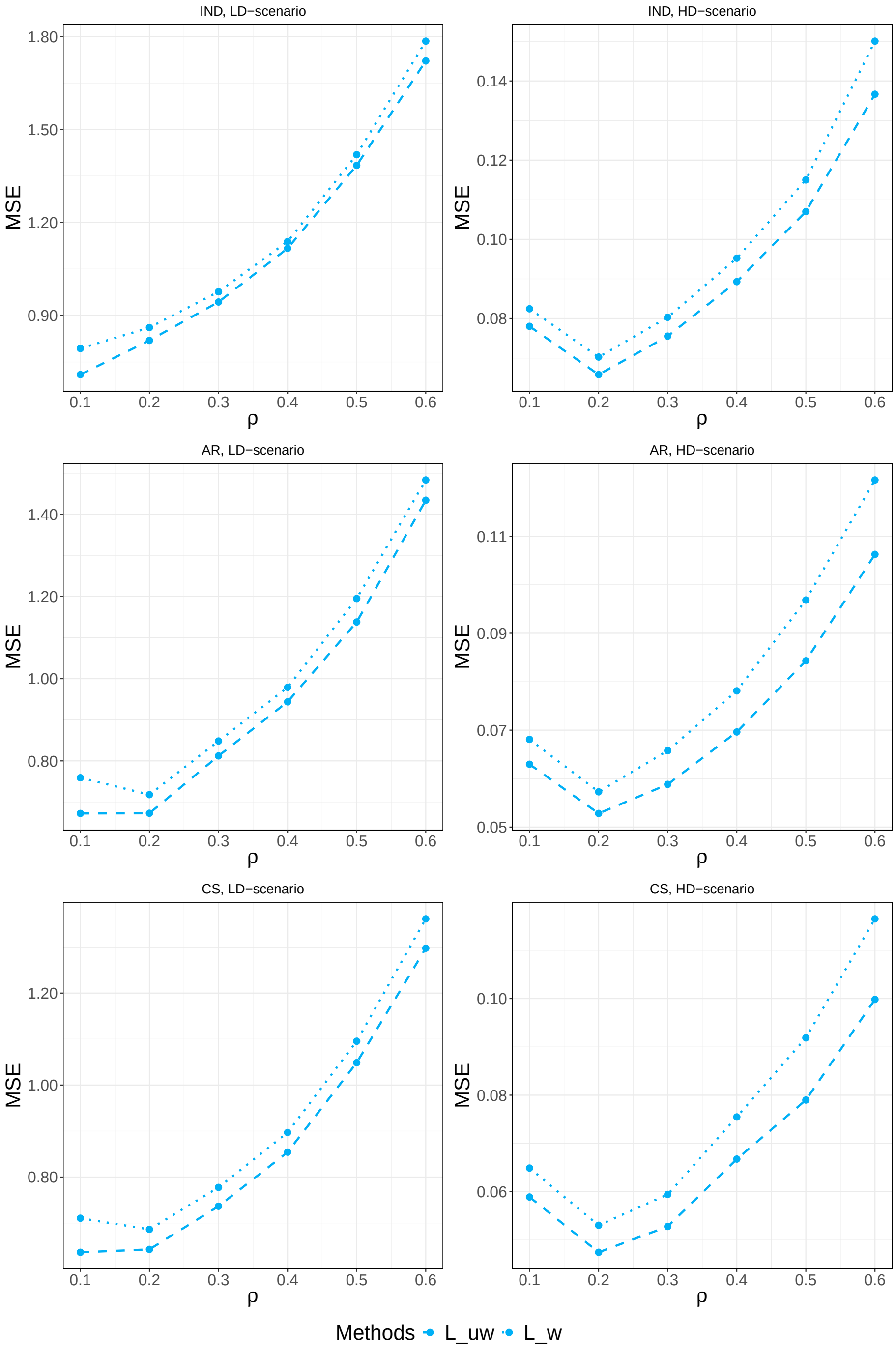

Figure S1: MSEs versus ρ under LD- and HD-scenarios for logistic regression.

based on $\mathcal{F}^{*\mathrm{p}}$, estimated optimal subsampling probability, and final estimators based on $\mathcal{F}^{*D}$, respectively. The programming language and platform is R (version 4.1.0) based on Intel(R) Xeon(R) Gold 5118 CPU @ 2.30GHz with memory 125 GB. Note that the computational time for the full dataset is 99.211s and 864.255s under $n = 10^6$ and $n = 10^7$, respectively. Table S3 reveals that the computational cost is dominated by T_2, while T_1 and T_3 contribute marginally. As expected, as T_2 involves computing subsampling probabilities for all n observations, T_2 approximately scales linearly with n. In contrast, T_1 and T_3 only involve calculations based on subsamples of expected size r^{p} and r, which are typically small in practice. Additionally, because L-optimal estimates do not require estimation of $\boldsymbol{\Phi}^{-1}$, their computational cost during T_2 is substantially lower than that of A-optimal estimates. Compared to the full-data estimates, our proposed subsampling methods achieve dramatic computational savings: for $n = 10^7$, the runtime is approximately 10 seconds, versus nearly 15 minutes for the full-data analysis.

Table S3: The average computational time (in seconds) based on 100 samplings.

n		$\hat{\boldsymbol{\beta}}^{\mathrm{A}}_{\mathrm{w}}$		$\hat{\boldsymbol{\beta}}^{\mathrm{A}}_{\mathrm{uw}}$		$\hat{\boldsymbol{\beta}}^{\mathrm{L}}_{\mathrm{w}}$		$\hat{\boldsymbol{\beta}}^{\mathrm{L}}_{\mathrm{uw}}$	
	r	400	1000	400	1000	400	1000	400	1000
10^6	T_1	0.080	0.080	0.080	0.080	0.076	0.076	0.076	0.076
	T_2	1.384	1.424	1.384	1.424	1.048	1.075	1.048	1.075
	T_3	0.112	0.135	0.112	0.136	0.092	0.136	0.092	0.137
10^7	T_1	0.575	0.595	0.575	0.595	0.570	0.591	0.570	0.591
	T_2	12.088	12.693	12.088	12.693	9.732	9.210	9.732	9.210
	T_3	0.844	0.909	0.845	0.907	0.676	0.727	0.679	0.725

References

Fang, E. X., Y. Ning, and R. Li (2020). Test of significance for high-dimensional longitudinal data. *The Annals of Statistics 48*(5), 2622–2645.

Gao, J., L. Wang, and J. Shao (2025). Distributed subsampling and quasi decorrelated score for cluster data: An application to beijing multisite air quality. *The Annals of Applied Statistics 19*(3), 1967–1987.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge university press.

Zhou, J. and A. Qu (2012). Informative estimation and selection of correlation structure for longitudinal data. *Journal of the American statistical Association 107*(498), 701–710.