Supplemental Material for "Task-driven Optimal Subsampling for Massive Data"

S.1 Proofs of the Theorems

S.1.1 Proof of Theorem 1

Lemma 1 *(Theorem 1 of Wang et al. (2022)) Under Assumptions 1-5, as $r \to \infty$ and $n \to \infty$, the estimator $\tilde{\boldsymbol{\theta}}_{r,R}$ satisfies that*

$$\sqrt{r}\{V_R(\hat{\boldsymbol{\theta}}_n)\}^{-1/2}(\tilde{\boldsymbol{\theta}}_{r,R} - \hat{\boldsymbol{\theta}}_n) \overset{|\mathcal{D}_n}{\rightsquigarrow} \mathbb{N}(\mathbf{0}, I_p). \tag{1}$$

According to Lemma 1, we have

$$\sqrt{r}\{V_R(\hat{\boldsymbol{\theta}}_n)\}^{-1/2}(\tilde{\boldsymbol{\theta}}_{r,R} - \hat{\boldsymbol{\theta}}_n) \overset{|\mathcal{D}_n}{\rightsquigarrow} \mathbb{N}(\mathbf{0}, I_p).$$

Combining Delta method, as $r \to \infty$ and $n \to \infty$,

$$\sqrt{r}\{\dot{\mu}^\top(\hat{\boldsymbol{\theta}}_n)V_R(\hat{\boldsymbol{\theta}}_n)\dot{\mu}(\hat{\boldsymbol{\theta}}_n)\}^{-1/2}\{\mu(\tilde{\boldsymbol{\theta}}_{r,R}) - \mu(\hat{\boldsymbol{\theta}}_n)\} \overset{|\mathcal{D}_n}{\rightsquigarrow} \mathbb{N}(\mathbf{0}, I_q).$$

S.1.2 Proof of Theorem 2

Lemma 2 *(Theorem 2 of Wang et al. (2022)) Under Assumptions 1-5, as $r \to \infty$ and $n \to \infty$, the estimator $\tilde{\boldsymbol{\theta}}_{r,P}$ satisfies that*

$$\sqrt{r}\{V_P(\hat{\boldsymbol{\theta}}_n)\}^{-1/2}(\tilde{\boldsymbol{\theta}}_{r,P} - \hat{\boldsymbol{\theta}}_n) \overset{|\mathcal{D}_n}{\rightsquigarrow} \mathbb{N}(\mathbf{0}, I_p). \tag{2}$$

According to Lemma 2, we have

$$\sqrt{r}\{V_P(\hat{\boldsymbol{\theta}}_n)\}^{-1/2}(\tilde{\boldsymbol{\theta}}_{r,P} - \hat{\boldsymbol{\theta}}_n) \overset{|\mathcal{D}_n}{\rightsquigarrow} \mathbb{N}(\mathbf{0}, I_p).$$

Combining Delta method, as $r \to \infty$ and $n \to \infty$,

$$\sqrt{r}\{\dot{\mu}^\top(\hat{\boldsymbol{\theta}}_n)V_P(\hat{\boldsymbol{\theta}}_n)\dot{\mu}(\hat{\boldsymbol{\theta}}_n)\}^{-1/2}\{\mu(\tilde{\boldsymbol{\theta}}_{r,P}) - \mu(\hat{\boldsymbol{\theta}}_n)\} \overset{|\mathcal{D}_n}{\rightsquigarrow} \mathbb{N}(\mathbf{0}, I_q).$$

S.1.3 Proof of Theorem 3

Note that

$$\begin{aligned}
& tr\{\dot{\mu}^\top(\hat{\boldsymbol{\theta}}_n)V_R(\hat{\boldsymbol{\theta}}_n)\dot{\mu}(\hat{\boldsymbol{\theta}}_n)\} \\
=& tr\{\dot{\mu}^\top(\hat{\boldsymbol{\theta}}_n)\ddot{M}_n^{-1}(\hat{\boldsymbol{\theta}}_n)\Lambda_R(\hat{\boldsymbol{\theta}}_n)\ddot{M}_n^{-1}(\hat{\boldsymbol{\theta}}_n)\dot{\mu}(\hat{\boldsymbol{\theta}}_n)\} \\
=& tr\left\{\frac{1}{n^2}\sum_{i=1}^n \frac{1}{\pi_i} t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)t^\top(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\right\} \\
=& \frac{1}{n^2}\sum_{i=1}^n \frac{1}{\pi_i}\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2 \\
=& \frac{1}{n^2}\sum_{j=1}^n \pi_j \sum_{i=1}^n \frac{1}{\pi_i}\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2 \\
\geq & \left(\frac{1}{n}\sum_{i=1}^n \|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\right)^2,
\end{aligned}$$

where the final inequality follows the Cauchy-Schwarz inequality and the equality holds when π_i is proportional to $\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|$. This completes the proof of Theorem 3.

S.1.4 Proof of Theorem 4

Note that

$$\begin{aligned}
& tr\{\dot{\mu}(\hat{\boldsymbol{\theta}}_n)^\top V_P(\hat{\boldsymbol{\theta}}_n)\dot{\mu}(\hat{\boldsymbol{\theta}}_n)\} \\
=& tr\left\{\frac{1}{n^2}\sum_{i=1}^n \frac{1-\pi_i}{\pi_i} t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)t^\top(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\right\} \\
=& \frac{1}{n^2}\left\{\sum_{i=1}^n \frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2}{\pi_i} - \sum_{i=1}^n \|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\right\}.
\end{aligned}$$

Thus, minimizing $tr\{\dot{\mu}(\hat{\boldsymbol{\theta}}_n)^\top V_P(\hat{\boldsymbol{\theta}}_n)\dot{\mu}(\hat{\boldsymbol{\theta}}_n)\}$ is equivalent to minimizing

$$\sum_{i=1}^n \frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2}{\pi_i}.$$

Let $t_i = \|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|$, and denote its order statistics by $t_{(i)} = \|t(\boldsymbol{Z},\hat{\boldsymbol{\theta}}_n)\|_{(i)}$, where $t_{(1)} \leq ... \leq t_{(n)}$. Without loss of generality, assume $t_1 \leq ... \leq t_n$. The problem of determining the optimal sampling probabilities can then be formulated as the following constrained optimization problem:

$$\begin{aligned}
\min_{\boldsymbol{\pi}} \quad & \sum_{i=1}^n \frac{t_{(i)}^2}{\pi_i} \\
\text{s.t.} \quad & 0 \leq \pi_i \leq \frac{1}{r}, \quad i = 1, ..., n,
\end{aligned}$$

$$\sum_{i=1}^{n} \pi_i = 1.$$

To solve this problem, we apply the method of Lagrange multipliers. Introducing slack variables $\omega_1^2, ..., \omega_n^2$, we define the Lagrangian as

$$\begin{aligned} &A(\pi_1, ..., \pi_n, \tau, \delta_1, ..., \delta_n, \omega_1, ..., \omega_n) \\ =&\sum_{i=1}^{n} \frac{t_{(i)}^2}{\pi_i} + \tau\left(\sum_{i=1}^{n} \pi_i - 1\right) + \sum_{i=1}^{n} \delta_i\left(\pi_i + \omega_i^2 - \frac{1}{r}\right). \end{aligned}$$

The Karush-Kuhn-Tucker (KKT) conditions for this problem are given by:

$$\frac{\partial A}{\partial \pi_i} = -\frac{t_{(i)}^2}{\pi_i^2} + \tau + \delta_i = 0, \quad i = 1, ..., n. \tag{3}$$

$$\frac{\partial A}{\partial \tau} = \sum_{i=1}^{n} \pi_i - 1 = 0, \tag{4}$$

$$\frac{\partial A}{\partial \delta_i} = \pi_i + \omega_i^2 - \frac{1}{r} = 0, \quad i = 1, ..., n. \tag{5}$$

$$\frac{\partial A}{\partial \omega_i} = 2\delta_i\omega_i = 0, \quad i = 1, ..., n. \tag{6}$$

$$\delta_i \geq 0, \quad i = 1, ..., n. \tag{7}$$

From (3), we have

$$\pi_i = \frac{t_{(i)}}{\sqrt{\tau + \delta_i}}, \quad i = 1, ..., n. \tag{8}$$

Combining this with the condition (5), we have

$$\frac{t_{(i)}}{\sqrt{\tau + \delta_i}} + \omega_i^2 = \frac{1}{r}, \quad i = 1, ..., n. \tag{9}$$

Meanwhile, according to (6), we conclude that at least one of δ_i and ω_i is equal to zero. Thus, we have the following cases:

$$\text{If} \quad t_{(i)} < \frac{\sqrt{\tau}}{r}, \quad \text{then} \quad \delta_i = 0 \quad \text{and} \quad \pi_i = \frac{t_{(i)}}{\sqrt{\tau}} < \frac{1}{r}; \tag{10}$$

$$\text{If} \quad t_{(i)} \geq \frac{\sqrt{\tau}}{r}, \quad \text{then} \quad \omega_i = 0 \quad \text{and} \quad \pi_i = \frac{t_{(i)}}{\sqrt{\tau + \delta_i}} = \frac{1}{r}. \tag{11}$$

Let g be the number of indices such that $t_{(i)} \geq \sqrt{\tau}/r$. Using the constraint (4) and the fact that $t_{(i)}$ is non-decreasing, we obtain

$$1 = \sum_{i=1}^{n} \pi_i = \sum_{i=1}^{n-g} \frac{t_{(i)}}{\sqrt{\tau}} + \sum_{i=n-g+1}^{n} \frac{1}{r} = \frac{\sum_{i=1}^{n-g} t_{(i)}}{\sqrt{\tau}} + \frac{g}{r}, \tag{12}$$

which implies

$$\sqrt{\tau} = \frac{r}{r-g} \sum_{i=1}^{n-g} t_{(i)}. \tag{13}$$

Combining (10), (11) and (13), we have

$$\pi_i = \begin{cases} \dfrac{t_{(i)}(r-g)}{r\sum_{i=1}^{n-g} t_{(i)}}, & i = 1, ..., n-g; \\ \dfrac{1}{r}, & i = n-g+1, ..., n. \end{cases} \tag{14}$$

Finally, let

$$G = \frac{\sum_{i=1}^{n-g} t_{(i)}}{r-g} = \frac{\sqrt{\tau}}{r}. \tag{15}$$

Then $t_{(i)} < G$ for $i = 1, ..., n-g$, and $t_{(i)} \geq G$ for $i = n-g+1, ..., n$. This implies

$$\sum_{i=1}^{n} (t_{(i)} \wedge G) = \sum_{i=1}^{n-g} t_{(i)} + \sum_{i=n-g+1}^{n} G = rG. \tag{16}$$

Therefore, for $i = 1, ..., n-g$,

$$\pi_i = \frac{t_{(i)}}{rG} = \frac{t_{(i)} \wedge G}{\sum_{i=1}^{n}(t_{(i)} \wedge G)}; \tag{17}$$

and for $i = n-g+1, ..., n$,

$$\pi_{n,i} = \frac{G}{rG} = \frac{t_{(i)} \wedge G}{\sum_{i=1}^{n}(t_{(i)} \wedge G)}. \tag{18}$$

This completes the proof of Theorem 4.

S.1.5 Proof of Theorem 5

First, we present the required lemmas and their proofs.

Lemma 3 *Under Assumption 3, if* $\|\tilde{\boldsymbol{\theta}}^{\alpha}_{r,R} - \hat{\boldsymbol{\theta}}_n\| = o_p(1)$*, then conditional on* $\mathcal{D}_n$ *and* $\tilde{\boldsymbol{\theta}}_{0,R}$*,*

$$B_r^{\alpha} - \ddot{M}_n(\hat{\boldsymbol{\theta}}_n) = o_p(1), \tag{19}$$

where

$$B_r^{\alpha} = \int_0^1 \ddot{M}^*_{R\alpha}(\hat{\boldsymbol{\theta}}_n + \lambda(\tilde{\boldsymbol{\theta}}^{\alpha}_{r,R} - \hat{\boldsymbol{\theta}}_n))d\lambda = \int_0^1 \frac{1}{r}\sum_{i=1}^{r} \frac{\ddot{m}(\boldsymbol{Z}_i^*, \hat{\boldsymbol{\theta}}_n + \lambda(\tilde{\boldsymbol{\theta}}^{\alpha}_{r,R} - \hat{\boldsymbol{\theta}}_n))}{n\tilde{\pi}^{TOSR*}_{\alpha i}} d\lambda.$$

Proof Notice that

$$\|B_r^{\alpha} - \ddot{M}_n(\hat{\boldsymbol{\theta}}_n)\|$$

$$
\begin{aligned}
&\leq \int_0^1 \left\| \frac{1}{r}\sum_{i=1}^r \frac{\ddot{m}(\boldsymbol{Z}_i^*, \hat{\boldsymbol{\theta}}_n + \lambda(\tilde{\boldsymbol{\theta}}_{r,R}^{\alpha} - \hat{\boldsymbol{\theta}}_n))}{n\tilde{\pi}_{\alpha i}^{TOSR*}} - \frac{1}{n}\sum_{i=1}^n \ddot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n) \right\| d\lambda \\
&\leq \int_0^1 \left\{ \left\| \frac{1}{r}\sum_{i=1}^r \frac{\ddot{m}(\boldsymbol{Z}_i^*, \hat{\boldsymbol{\theta}}_n + \lambda(\tilde{\boldsymbol{\theta}}_{r,R}^{\alpha} - \hat{\boldsymbol{\theta}}_n))}{n\tilde{\pi}_{\alpha i}^{TOSR*}} - \frac{1}{r}\sum_{i=1}^r \frac{\ddot{m}(\boldsymbol{Z}_i^*, \hat{\boldsymbol{\theta}}_n)}{n\tilde{\pi}_{\alpha i}^{TOSR*}} \right\| \right. \\
&\quad \left. + \left\| \frac{1}{r}\sum_{i=1}^r \frac{\ddot{m}(\boldsymbol{Z}_i^*, \hat{\boldsymbol{\theta}}_n)}{n\tilde{\pi}_{\alpha i}^{TOSR*}} - \frac{1}{n}\sum_{i=1}^n \ddot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n) \right\| \right\} d\lambda. \qquad (20)
\end{aligned}
$$

First, for every $k, l = 1, 2, ..., p$, according to Assumption 3, we have

$$
\begin{aligned}
&\left| \frac{1}{r}\sum_{i=1}^r \frac{\ddot{m}_{k,l}(\boldsymbol{Z}_i^*, \hat{\boldsymbol{\theta}}_n + \lambda(\tilde{\boldsymbol{\theta}}_{r,R}^{\alpha} - \hat{\boldsymbol{\theta}}_n))}{n\tilde{\pi}_{\alpha i}^{TOSR*}} - \frac{1}{r}\sum_{i=1}^r \frac{\ddot{m}_{k,l}(\boldsymbol{Z}_i^*, \hat{\boldsymbol{\theta}}_n)}{n\tilde{\pi}_{\alpha i}^{TOSR*}} \right| \\
\leq & \frac{1}{r}\sum_{i=1}^r \frac{|\ddot{m}_{k,l}(\boldsymbol{Z}_i^*, \hat{\boldsymbol{\theta}}_n + \lambda(\tilde{\boldsymbol{\theta}}_{r,R}^{\alpha} - \hat{\boldsymbol{\theta}}_n)) - \ddot{m}_{k,l}(\boldsymbol{Z}_i^*, \hat{\boldsymbol{\theta}}_n)|}{n\tilde{\pi}_{\alpha i}^{TOSR*}} \\
\leq & \frac{1}{r}\sum_{i=1}^r \frac{\lambda\psi(\boldsymbol{Z}_i^*)\|\tilde{\boldsymbol{\theta}}_{r,R}^{\alpha} - \hat{\boldsymbol{\theta}}_n\|}{n\tilde{\pi}_{\alpha i}^{TOSR*}} \\
= & o_p(1).
\end{aligned}
$$

That is to say,

$$
\left\| \frac{1}{r}\sum_{i=1}^r \frac{\ddot{m}(\boldsymbol{Z}_i^*, \hat{\boldsymbol{\theta}}_n + \lambda(\tilde{\boldsymbol{\theta}}_{r,R}^{\alpha} - \hat{\boldsymbol{\theta}}_n))}{n\tilde{\pi}_{\alpha i}^{TOSR*}} - \frac{1}{r}\sum_{i=1}^r \frac{\ddot{m}(\boldsymbol{Z}_i^*, \hat{\boldsymbol{\theta}}_n)}{n\tilde{\pi}_{\alpha i}^{TOSR*}} \right\| = o_p(1). \qquad (21)
$$

Second, for every $k, l = 1, 2, ..., p$, we have

$$
\mathbb{E}\left\{ \frac{1}{r}\sum_{i=1}^r \frac{\ddot{m}_{k,l}(\boldsymbol{Z}_i^*, \hat{\boldsymbol{\theta}}_n)}{n\tilde{\pi}_{\alpha i}^{TOSR*}} \middle| \mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,R} \right\} = \frac{1}{n}\sum_{i=1}^n \ddot{m}_{k,l}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n),
$$

and

$$
\begin{aligned}
& Var\left\{ \frac{1}{r}\sum_{i=1}^r \frac{\ddot{m}_{k,l}(\boldsymbol{Z}_i^*, \hat{\boldsymbol{\theta}}_n)}{n\tilde{\pi}_{\alpha i}^{TOSR*}} \middle| \mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,R} \right\} \\
\leq & \frac{1}{n^2}\frac{1}{r}\sum_{i=1}^n \frac{\ddot{m}_{k,l}^2(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)}{\tilde{\pi}_{\alpha i}^{TOSR}} \\
\leq & \frac{1}{\alpha r}\frac{1}{n}\sum_{i=1}^n \ddot{m}_{k,l}^2(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n) \\
= & O_p(r^{-1}).
\end{aligned}
$$

Therefore, according to Chebyshev's inequality, conditional on $\mathcal{D}_n$ and $\tilde{\boldsymbol{\theta}}_{0,R}$, we have

$$
\left\| \frac{1}{r}\sum_{i=1}^r \frac{\ddot{m}(\boldsymbol{Z}_i^*, \hat{\boldsymbol{\theta}}_n)}{n\tilde{\pi}_{\alpha i}^{TOSR*}} - \frac{1}{n}\sum_{i=1}^n \ddot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n) \right\| = o_p(1). \qquad (22)
$$

Combining (20)-(22), we finish the proof of Lemma 3.

Lemma 4 *Let* $M^*_{R\alpha}(\boldsymbol{\theta}) = (1/n)\sum_{i=1}^r m(\boldsymbol{Z}_i^*, \boldsymbol{\theta})/(r\tilde{\pi}_{\alpha i}^{TOSR*})$. *Under Assumption 4, conditional on* $\mathcal{D}_n$ *and* $\tilde{\boldsymbol{\theta}}_{0,R}$,

$$\sqrt{r}\{\Lambda_R^\alpha(\tilde{\boldsymbol{\theta}}_{0,R})\}^{-1/2}\dot{M}^*_{R\alpha}(\hat{\boldsymbol{\theta}}_n) \overset{|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_{0,R}}{\rightsquigarrow} N(\mathbf{0}, I_p). \tag{23}$$

Proof Note that

$$\sqrt{r}\dot{M}^*_{R\alpha}(\hat{\boldsymbol{\theta}}_n) = \frac{1}{\sqrt{r}}\sum_{i=1}^r \frac{\dot{m}(\boldsymbol{Z}_i^*, \hat{\boldsymbol{\theta}}_n)}{n\tilde{\pi}_{\alpha i}^{TOSR*}}.$$

Let $\eta_i^* = \dot{m}(\boldsymbol{Z}_i^*, \hat{\boldsymbol{\theta}}_n)/(n\tilde{\pi}_{\alpha i}^{TOSR*})$. Thus, we have

$$\mathbb{E}(\eta_i^*|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,R}) = \frac{1}{n}\sum_{i=1}^n \dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n) = 0,$$

and

$$Var(\eta_i^*|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,R}) = \frac{1}{n^2}\sum_{i=1}^n \frac{\dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\dot{m}^\top(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)}{\tilde{\pi}_{\alpha i}^{TOSR}} = \Lambda_R^\alpha(\tilde{\boldsymbol{\theta}}_{0,R}).$$

Besides, for $\forall\epsilon > 0$, and $\delta \in (0, 2]$,

$$\begin{aligned}
&\sum_{i=1}^r \mathbb{E}\Big\{\Big\|\frac{\eta_i^*}{\sqrt{r}}\Big\|^2 I\Big(\Big\|\frac{\eta_i^*}{\sqrt{r}}\Big\| > \epsilon\Big)\Big|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,R}\Big\}\\
=&\frac{1}{r}\sum_{i=1}^r \mathbb{E}\{\|\eta_i^*\|^2 I(\|\eta_i^*\| > \sqrt{r}\epsilon)|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,R}\}\\
\le&\frac{1}{r}\frac{1}{r^{\delta/2}\epsilon^\delta}\sum_{i=1}^r \mathbb{E}\{\|\eta_i^*\|^{2+\delta} I(\|\eta_i^*\| > \sqrt{r}\epsilon)|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,R}\}\\
\le&\frac{1}{r^{1+\delta/2}\epsilon^\delta}\sum_{i=1}^r \mathbb{E}\{\|\eta_i^*\|^{2+\delta}|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,R}\}\\
=&\frac{1}{r^{1+\delta/2}\epsilon^\delta} r\sum_{i=1}^n \frac{\|\dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|^{2+\delta}}{n^{2+\delta}\tilde{\pi}_{\alpha i}^{TOSR(1+\delta)}}\\
\le&\frac{1}{r^{\delta/2}\epsilon^\delta\alpha^{1+\delta}}\frac{1}{n}\sum_{i=1}^n \|\dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|^{2+\delta}\\
=&O_p(r^{-\delta/2}),
\end{aligned} \tag{24}$$

where the fifth inequality follows the fact that $n\tilde{\pi}_{\alpha i}^{TOSR} = n\{(1-\alpha)\tilde{\pi}_i^{TOSR} + \alpha/n\} \ge \alpha$, and the final equality holds due to Assumption 4. By the Lindeberg-Feller central limit theorem, we obtain

$$\sqrt{r}\{\Lambda_R^\alpha(\tilde{\boldsymbol{\theta}}_{0,R})\}^{-1/2}\dot{M}^*_{R\alpha}(\hat{\boldsymbol{\theta}}_n) \overset{|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_{0,R}}{\rightsquigarrow} N(\mathbf{0}, I_p). \tag{25}$$

Lemma 5 *Under Assumptions 3, 4, and 6,*

$$\|\Lambda_R^\alpha(\tilde{\boldsymbol{\theta}}_{0,R}) - \Lambda_R^\alpha(\hat{\boldsymbol{\theta}}_n)\| = o_p(1).$$

Proof We determine the distance between $\|t(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,R})\|$ and $\|t(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|$ at first, where $i = 1, ..., n$.

$$
\begin{aligned}
&|\|t(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,R})\| - \|t(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\||\\
\leq&|\|t(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,R})\| - \|\dot{\mu}^\top(\tilde{\boldsymbol{\theta}}_{0,R})\ddot{M}_n^{-1}(\tilde{\boldsymbol{\theta}}_{0,R})\dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\||\\
&+|\|\dot{\mu}^\top(\tilde{\boldsymbol{\theta}}_{0,R})\ddot{M}_n^{-1}(\tilde{\boldsymbol{\theta}}_{0,R})\dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\| - \|t(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\||\\
\leq&\|\dot{\mu}^\top(\tilde{\boldsymbol{\theta}}_{0,R})\ddot{M}_n^{-1}(\tilde{\boldsymbol{\theta}}_{0,R})\{\dot{m}(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,R}) - \dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\}\|\\
+&\|\dot{\mu}^\top(\tilde{\boldsymbol{\theta}}_{0,R})\ddot{M}_n^{-1}(\tilde{\boldsymbol{\theta}}_{0,R}) - \dot{\mu}^\top(\hat{\boldsymbol{\theta}}_n)\ddot{M}_n^{-1}(\hat{\boldsymbol{\theta}}_n)\|\|\dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|\\
\leq&\|\dot{\mu}(\tilde{\boldsymbol{\theta}}_{0,R})\|\|\ddot{M}_n^{-1}(\tilde{\boldsymbol{\theta}}_{0,R})\|\|\dot{m}(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,R}) - \dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|\\
+&\|\dot{\mu}^\top(\tilde{\boldsymbol{\theta}}_{0,R})\ddot{M}_n^{-1}(\tilde{\boldsymbol{\theta}}_{0,R}) - \dot{\mu}^\top(\hat{\boldsymbol{\theta}}_n)\ddot{M}_n^{-1}(\tilde{\boldsymbol{\theta}}_{0,R})\|\|\dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|\\
+&\|\dot{\mu}^\top(\hat{\boldsymbol{\theta}}_n)\ddot{M}_n^{-1}(\tilde{\boldsymbol{\theta}}_{0,R}) - \dot{\mu}^\top(\hat{\boldsymbol{\theta}}_n)\ddot{M}_n^{-1}(\hat{\boldsymbol{\theta}}_n)\|\|\dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|\\
\leq&\|\dot{\mu}(\tilde{\boldsymbol{\theta}}_{0,R})\|\|\ddot{M}_n^{-1}(\tilde{\boldsymbol{\theta}}_{0,R})\|\|\dot{m}(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,R}) - \dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|\\
+&\|\dot{\mu}(\tilde{\boldsymbol{\theta}}_{0,R}) - \dot{\mu}(\hat{\boldsymbol{\theta}}_n)\|\|\ddot{M}_n^{-1}(\tilde{\boldsymbol{\theta}}_{0,R})\|\|\dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|\\
+&\|\dot{\mu}(\hat{\boldsymbol{\theta}}_n)\|\|\ddot{M}_n^{-1}(\tilde{\boldsymbol{\theta}}_{0,R}) - \ddot{M}_n^{-1}(\hat{\boldsymbol{\theta}}_n)\|\|\dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|\\
\leq&\|\dot{\mu}(\tilde{\boldsymbol{\theta}}_{0,R})\|\|\ddot{M}_n^{-1}(\tilde{\boldsymbol{\theta}}_{0,R})\|\|\dot{m}(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,R}) - \dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|\\
+&\|\dot{\mu}(\tilde{\boldsymbol{\theta}}_{0,R}) - \dot{\mu}(\hat{\boldsymbol{\theta}}_n)\|\|\ddot{M}_n^{-1}(\tilde{\boldsymbol{\theta}}_{0,R})\|\|\dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|\\
+&\|\dot{\mu}(\hat{\boldsymbol{\theta}}_n)\|\|\ddot{M}_n^{-1}(\tilde{\boldsymbol{\theta}}_{0,R})\|\|\ddot{M}_n(\tilde{\boldsymbol{\theta}}_{0,R}) - \ddot{M}_n(\hat{\boldsymbol{\theta}}_n)\|\|\ddot{M}_n^{-1}(\hat{\boldsymbol{\theta}}_n)\|\|\dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|. \qquad (26)
\end{aligned}
$$

Note that for every $i = 1, ..., n$,

$$
\begin{aligned}
\|\dot{m}(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,R}) - \dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\| =&\sqrt{\sum_{k=1}^{p}\{\dot{m}_k(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,R}) - \dot{m}_k(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\}^2}\\
\leq&\sum_{k=1}^{p}|\dot{m}_k(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,R}) - \dot{m}_k(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)|\\
\leq&\sum_{k=1}^{p}|\ddot{m}_k^\top(\boldsymbol{Z}_i, \xi_k)(\tilde{\boldsymbol{\theta}}_{0,R} - \hat{\boldsymbol{\theta}}_n)|\\
\leq&\|\tilde{\boldsymbol{\theta}}_{0,R} - \hat{\boldsymbol{\theta}}_n\|\sum_{k=1}^{p}\|\ddot{m}_k(\boldsymbol{Z}_i, \xi_k)\|\\
\triangleq&\|\tilde{\boldsymbol{\theta}}_{0,R} - \hat{\boldsymbol{\theta}}_n\|h(\boldsymbol{Z}_i), \qquad (27)
\end{aligned}
$$

where $\dot{m}_k(\boldsymbol{Z}_i, \boldsymbol{\theta})$ is the k-th element of $\dot{m}(\boldsymbol{Z}_i, \boldsymbol{\theta})$ and ξ_k is between $\tilde{\boldsymbol{\theta}}_{0,R}$ and $\hat{\boldsymbol{\theta}}_n$. From Assumption 3, we derive that

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}h^2(\boldsymbol{Z}_i) =&\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{k=1}^{p}\|\ddot{m}_k(\boldsymbol{Z}_i, \xi_k)\|\right)^2\\
\leq&p\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{p}\|\ddot{m}_k(\boldsymbol{Z}_i, \xi_k)\|^2
\end{aligned}
$$

$$
\begin{aligned}
&= p\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{p}\sum_{l=1}^{p}\ddot{m}_{k,l}^2(\boldsymbol{Z}_i,\xi_k)\\
&\le p\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{p}\sum_{l=1}^{p}(2\ddot{m}_{k,l}^2(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)+2\psi^2(\boldsymbol{Z}_i)\|\tilde{\boldsymbol{\theta}}_{0,R}-\hat{\boldsymbol{\theta}}_n\|^2)\\
&= O_p(1).
\end{aligned} \tag{28}
$$

Furthermore, under the assumption that $\dot{\mu}(\boldsymbol{\theta})$ is Lipschitz continuous with respect to $\boldsymbol{\theta}$, that is to say, there exists a constant L such that for any $\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\Theta$, it holds that $\|\dot{\mu}(\boldsymbol{\theta}_1)-\dot{\mu}(\boldsymbol{\theta}_2)\|\le L\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|$. Then, by combining (26)-(27), we obtain that

$$
\begin{aligned}
|\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,R})\|-\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|| \le&\|\dot{\mu}(\tilde{\boldsymbol{\theta}}_{0,R})\|\|\ddot{M}_n^{-1}(\tilde{\boldsymbol{\theta}}_{0,R})\|\|\tilde{\boldsymbol{\theta}}_{0,R}-\hat{\boldsymbol{\theta}}_n\|h(\boldsymbol{Z}_i)\\
&+L\|\tilde{\boldsymbol{\theta}}_{0,R}-\hat{\boldsymbol{\theta}}_n\|\|\ddot{M}_n^{-1}(\tilde{\boldsymbol{\theta}}_{0,R})\|\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\\
&+\|\dot{\mu}(\hat{\boldsymbol{\theta}}_n)\|\|\ddot{M}_n^{-1}(\tilde{\boldsymbol{\theta}}_{0,R})\|\|\ddot{M}_n(\tilde{\boldsymbol{\theta}}_{0,R})-\ddot{M}_n(\hat{\boldsymbol{\theta}}_n)\|\|\ddot{M}_n^{-1}(\hat{\boldsymbol{\theta}}_n)\|\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|.
\end{aligned} \tag{29}
$$

Besides, similar to (22), we have

$$
\|\ddot{M}_n(\tilde{\boldsymbol{\theta}}_{0,R})-\ddot{M}_n(\hat{\boldsymbol{\theta}}_n)\|=O_p(r_0^{-1}). \tag{30}
$$

According to Assumption 6, we can derive that for every $\boldsymbol{\theta}\in\Theta$,

$$
\dot{\mu}(\boldsymbol{\theta})=O(1). \tag{31}
$$

Then,

$$
\begin{aligned}
&\|\Lambda_R^{\alpha}(\tilde{\boldsymbol{\theta}}_{0,R})-\Lambda_R^{\alpha}(\hat{\boldsymbol{\theta}}_n)\|\\
=&\left\|\frac{1}{n^2}\sum_{i=1}^{n}\frac{\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\dot{m}^\top(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)}{\tilde{\pi}_{\alpha i}^{TOSR}}-\frac{1}{n^2}\sum_{i=1}^{n}\frac{\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\dot{m}^\top(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)}{\pi_{\alpha i}^{TOSR}(\hat{\boldsymbol{\theta}}_n)}\right\|\\
\le&\frac{1}{\alpha^2}\sum_{i=1}^{n}\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2|\tilde{\pi}_i^{TOSR}-\pi_i^{TOSR}|\\
=&\frac{1}{\alpha^2}\sum_{i=1}^{n}\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\left|\frac{\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,R})\|}{\sum_{j=1}^n\|t(\boldsymbol{Z}_j,\tilde{\boldsymbol{\theta}}_{0,R})\|}-\frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|}{\sum_{j=1}^n\|t(\boldsymbol{Z}_j,\hat{\boldsymbol{\theta}}_n)\|}\right|\\
\le&\frac{1}{\alpha^2}\sum_{i=1}^{n}\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\left\{\left|\frac{\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,R})\|}{\sum_{j=1}^n\|t(\boldsymbol{Z}_j,\tilde{\boldsymbol{\theta}}_{0,R})\|}-\frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|}{\sum_{j=1}^n\|t(\boldsymbol{Z}_j,\tilde{\boldsymbol{\theta}}_{0,R})\|}\right|\right.\\
&\left.+\left|\frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|}{\sum_{j=1}^n\|t(\boldsymbol{Z}_j,\tilde{\boldsymbol{\theta}}_{0,R})\|}-\frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|}{\sum_{j=1}^n\|t(\boldsymbol{Z}_j,\hat{\boldsymbol{\theta}}_n)\|}\right|\right\}\\
\le&\frac{1}{\alpha^2}\sum_{i=1}^{n}\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\frac{|\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,R})\|-\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\||}{\sum_{j=1}^n\|t(\boldsymbol{Z}_j,\tilde{\boldsymbol{\theta}}_{0,R})\|}\\
&+\frac{1}{\alpha^2}\sum_{i=1}^{n}\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\frac{\sum_{j=1}^n|\|t(\boldsymbol{Z}_j,\tilde{\boldsymbol{\theta}}_{0,R})\|-\|t(\boldsymbol{Z}_j,\hat{\boldsymbol{\theta}}_n)\||}{\sum_{j=1}^n\|t(\boldsymbol{Z}_j,\tilde{\boldsymbol{\theta}}_{0,R})\|\sum_{j=1}^n\|t(\boldsymbol{Z}_j,\hat{\boldsymbol{\theta}}_n)\|}
\end{aligned}
$$

$$\triangleq \Delta_1 + \Delta_2. \tag{32}$$

Note that

$$\begin{aligned}
\Delta_1 =& \frac{1}{\alpha^2}\sum_{i=1}^n \|\dot m(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|^2 \frac{|\|t(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,R})\| - \|t(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\||}{\sum_{j=1}^n \|t(\boldsymbol{Z}_j, \tilde{\boldsymbol{\theta}}_{0,R})\|} \\
\leq & \frac{O_p(\|\tilde{\boldsymbol{\theta}}_{0,R} - \hat{\boldsymbol{\theta}}_n\|)}{n} \sum_{i=1}^n \|\dot m(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|^2 h(\boldsymbol{Z}_i) \\
& + \frac{O_p(\|\tilde{\boldsymbol{\theta}}_{0,R} - \hat{\boldsymbol{\theta}}_n\|)}{n} \sum_{i=1}^n \|\dot m(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|^3 + \frac{O_p(r_0^{-1})}{n} \sum_{i=1}^n \|\dot m(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|^3 \\
\leq & \frac{O_p(\|\tilde{\boldsymbol{\theta}}_{0,R} - \hat{\boldsymbol{\theta}}_n\|)}{n} \left\{ \sum_{i=1}^n \|\dot m(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|^4 \right\}^{1/2} \left\{ \sum_{i=1}^n h^2(\boldsymbol{Z}_i) \right\}^{1/2} \\
& + \frac{O_p(\|\tilde{\boldsymbol{\theta}}_{0,R} - \hat{\boldsymbol{\theta}}_n\|)}{n} \sum_{i=1}^n \|\dot m(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|^3 + \frac{O_p(r_0^{-1})}{n} \sum_{i=1}^n \|\dot m(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|^3 \\
=& o_p(1),
\end{aligned} \tag{33}$$

and

$$\begin{aligned}
\Delta_2 =& \frac{1}{\alpha^2}\sum_{i=1}^n \|\dot m(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|^2 \|t(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\| \frac{\sum_{j=1}^n |\|t(\boldsymbol{Z}_j, \tilde{\boldsymbol{\theta}}_{0,R})\| - \|t(\boldsymbol{Z}_j, \hat{\boldsymbol{\theta}}_n)\||}{\sum_{j=1}^n \|t(\boldsymbol{Z}_j, \tilde{\boldsymbol{\theta}}_{0,R})\| \sum_{j=1}^n \|t(\boldsymbol{Z}_j, \hat{\boldsymbol{\theta}}_n)\|} \\
\leq & \frac{1}{\alpha^2} \frac{\left\{ \sum_{i=1}^n \|\dot m(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|^4 \right\}^{1/2} \left\{ \sum_{i=1}^n \|t(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|^2 \right\}^{1/2}}{\sum_{j=1}^n \|t(\boldsymbol{Z}_j, \tilde{\boldsymbol{\theta}}_{0,R})\| \sum_{j=1}^n \|t(\boldsymbol{Z}_j, \hat{\boldsymbol{\theta}}_n)\|} \left\{ O_p(\|\tilde{\boldsymbol{\theta}}_{0,R} - \hat{\boldsymbol{\theta}}_n\|) \sum_{j=1}^n h(\boldsymbol{Z}_j) \right. \\
& \left. + O_p(\|\tilde{\boldsymbol{\theta}}_{0,R} - \hat{\boldsymbol{\theta}}_n\|) \sum_{j=1}^n \|\dot m(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\| + O_p(r_0^{-1}) \sum_{j=1}^n \|\dot m(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\| \right\} \\
=& o_p(1).
\end{aligned} \tag{34}$$

Combining (32)-(34), we finish the proof of Lemma 5.

Proof of Theorem 5:
To prove Theorem 5, we begin by proving

$$\sqrt{r}\{V_R^\alpha(\hat{\boldsymbol{\theta}}_n)\}^{-1/2}(\tilde{\boldsymbol{\theta}}_{r,R}^\alpha - \hat{\boldsymbol{\theta}}_n) \overset{|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,R}}{\rightsquigarrow} N(\mathbf{0}, I_p). \tag{35}$$

Note that

$$\tilde{\boldsymbol{\theta}}_{r,R}^\alpha = \arg\max_{\boldsymbol{\theta}} \frac{1}{r}\sum_{i=1}^r \frac{m(\boldsymbol{Z}_i^*, \boldsymbol{\theta})}{n\tilde{\pi}_{\alpha i}^{TOSR*}},$$

and

$$\hat{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta}} M_n(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\theta}} \frac{1}{n}\sum_{i=1}^n m(\boldsymbol{Z}_i, \boldsymbol{\theta}).$$

We can obtain that for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\begin{aligned}\mathbb{E}\{M^*_{R\alpha}(\boldsymbol{\theta})|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,R}\} =&\mathbb{E}\left\{\frac{1}{r}\sum_{i=1}^{r}\frac{m(\boldsymbol{Z}^*_i,\boldsymbol{\theta})}{n\tilde{\pi}^{TOSR*}_{\alpha i}}\middle|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,R}\right\}\\ =&\frac{1}{n}\sum_{i=1}^{n}\tilde{\pi}^{TOSR}_{\alpha i}\frac{m(\boldsymbol{Z}_i,\boldsymbol{\theta})}{\tilde{\pi}^{TOSR}_{\alpha i}}\\ =&M_n(\boldsymbol{\theta}),\end{aligned}$$

and

$$\begin{aligned}Var\{M^*_{R\alpha}(\boldsymbol{\theta})|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,R}\} =&\frac{1}{n^2}\sum_{i=1}^{r}\frac{1}{r^2}Var\left\{\frac{m(\boldsymbol{Z}^*_i,\boldsymbol{\theta})}{\tilde{\pi}^{TOSR*}_{\alpha i}}\middle|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,R}\right\}\\ \leq&\frac{1}{n^2 r}\mathbb{E}\left\{\left(\frac{m(\boldsymbol{Z}^*_i,\boldsymbol{\theta})}{\tilde{\pi}^{TOSR*}_{\alpha i}}\right)^2\middle|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,R}\right\}\\ =&\frac{1}{n^2 r}\sum_{i=1}^{n}\frac{m^2(\boldsymbol{Z}_i,\boldsymbol{\theta})}{\tilde{\pi}^{TOSR}_{\alpha i}}\\ \leq&\frac{1}{\alpha r n}\sum_{i=1}^{n}m^2(\boldsymbol{Z}_i,\boldsymbol{\theta})\\ =&O_p(r^{-1}),\end{aligned}$$

where the fourth inequality follows the fact that $n\tilde{\pi}^{TOSR}_{\alpha i} = n\{(1-\alpha)\tilde{\pi}^{TOSR}_{i} + \alpha/n\} \geq \alpha$, and the final equality holds due to Assumption 2.

By Chebyshev's inequality, for any $\epsilon > 0$, we have

$$P(|M^*_{R\alpha}(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta})| > \epsilon|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,R}) \leq \frac{1}{\epsilon^2}Var\{M^*_{R\alpha}(\boldsymbol{\theta})|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,R}\} = O_p(r^{-1}),$$

which implies that for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$M^*_{R\alpha}(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}) = o_{p|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_{0,R}}(1). \tag{36}$$

Combining the above equation with Assumptions 1 and 2, as well as Theorem 5.9 and the corresponding remark on Van der Vaart (2000), we can derive

$$\|\tilde{\boldsymbol{\theta}}^{\alpha}_{r,R} - \hat{\boldsymbol{\theta}}_n\| = o_p(1). \tag{37}$$

Then, by Taylor expansion,

$$0 = \dot{M}^*_{R\alpha}(\tilde{\boldsymbol{\theta}}^{\alpha}_{r,R}) = \dot{M}^*_{R\alpha}(\hat{\boldsymbol{\theta}}_n) + B^{\alpha}_r(\tilde{\boldsymbol{\theta}}^{\alpha}_{r,R} - \hat{\boldsymbol{\theta}}_n), \tag{38}$$

where

$$B^{\alpha}_r = \int_0^1 \ddot{M}^*_{R\alpha}(\hat{\boldsymbol{\theta}}_n + \lambda(\tilde{\boldsymbol{\theta}}^{\alpha}_{r,R} - \hat{\boldsymbol{\theta}}_n))d\lambda = \int_0^1 \frac{1}{r}\sum_{i=1}^{r}\frac{\ddot{m}(\boldsymbol{Z}^*_i, \hat{\boldsymbol{\theta}}_n + \lambda(\tilde{\boldsymbol{\theta}}^{\alpha}_{r,R} - \hat{\boldsymbol{\theta}}_n))}{n\tilde{\pi}^{TOSR*}_{\alpha i}}d\lambda.$$

Thus,

$$\tilde{\boldsymbol{\theta}}^{\alpha}_{r,R} - \hat{\boldsymbol{\theta}}_n = -(B^{\alpha}_r)^{-1}\dot{M}^*_{R\alpha}(\hat{\boldsymbol{\theta}}_n). \tag{39}$$

Combining Lemmas 3-5, according to Slutsky's theorem, equation (35) holds. Then, by applying the Delta method, Theorem 5 is proved.

S.1.6 Proof of Theorem 6

First, we present the required lemmas and their proofs.

Lemma 6 *Under Assumption 3, if* $\|\tilde{\boldsymbol{\theta}}^{\alpha}_{r,P} - \hat{\boldsymbol{\theta}}_n\| = o_p(1)$*, then conditional on* $\mathcal{D}_n$ *and* $\tilde{\boldsymbol{\theta}}_{0,P}$*,*

$$D^{\alpha}_r - \ddot{M}_n(\hat{\boldsymbol{\theta}}_n) = o_p(1),$$

where

$$D^{\alpha}_r = \int_0^1 \ddot{M}^*_{P\alpha}(\hat{\boldsymbol{\theta}}_n + \lambda(\tilde{\boldsymbol{\theta}}^{\alpha}_{r,P} - \hat{\boldsymbol{\theta}}_n))d\lambda = \int_0^1 \frac{1}{n}\sum_{i=1}^{n} \frac{v_i \ddot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n + \lambda(\tilde{\boldsymbol{\theta}}^{\alpha}_{r,P} - \hat{\boldsymbol{\theta}}_n))}{(r\tilde{\pi}^{TOSP}_{\alpha i}) \wedge 1} d\lambda,$$

and $v_i = I(u_i \le r\tilde{\pi}^{TOSP}_{\alpha i})$ *is the indicator variable denoting whether the ith sample (*$i = 1, ..., n$*) is selected into the subsample.*

Proof Notice that

$$\begin{aligned}
&\|D^{\alpha}_r - \ddot{M}_n(\hat{\boldsymbol{\theta}}_n)\| \\
\le &\int_0^1 \left\| \frac{1}{n}\sum_{i=1}^{n} \frac{v_i \ddot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n + \lambda(\tilde{\boldsymbol{\theta}}^{\alpha}_{r,P} - \hat{\boldsymbol{\theta}}_n))}{(r\tilde{\pi}^{TOSP}_{\alpha i}) \wedge 1} - \frac{1}{n}\sum_{i=1}^{n} \ddot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n) \right\| d\lambda \\
\le &\int_0^1 \left\{ \left\| \frac{1}{n}\sum_{i=1}^{n} \frac{v_i \ddot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n + \lambda(\tilde{\boldsymbol{\theta}}^{\alpha}_{r,P} - \hat{\boldsymbol{\theta}}_n))}{(r\tilde{\pi}^{TOSP}_{\alpha i}) \wedge 1} - \frac{1}{n}\sum_{i=1}^{n} \frac{v_i \ddot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)}{(r\tilde{\pi}^{TOSP}_{\alpha i}) \wedge 1} \right\| \right. \\
&\left. + \left\| \frac{1}{n}\sum_{i=1}^{n} \frac{v_i \ddot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)}{(r\tilde{\pi}^{TOSP}_{\alpha i}) \wedge 1} - \frac{1}{n}\sum_{i=1}^{n} \ddot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n) \right\| \right\} d\lambda.
\end{aligned} \tag{40}$$

First, for every $k, l = 1, 2, ..., p$, according to Assumption 3, we have

$$\begin{aligned}
&\left| \frac{1}{n}\sum_{i=1}^{n} \frac{v_i \ddot{m}_{k,l}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n + \lambda(\tilde{\boldsymbol{\theta}}^{\alpha}_{r,P} - \hat{\boldsymbol{\theta}}_n))}{(r\tilde{\pi}^{TOSP}_{\alpha i}) \wedge 1} - \frac{1}{n}\sum_{i=1}^{n} \frac{v_i \ddot{m}_{k,l}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)}{(r\tilde{\pi}^{TOSP}_{\alpha i}) \wedge 1} \right| \\
\le &\frac{1}{n}\sum_{i=1}^{n} \frac{v_i |\ddot{m}_{k,l}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n + \lambda(\tilde{\boldsymbol{\theta}}^{\alpha}_{r,P} - \hat{\boldsymbol{\theta}}_n)) - \ddot{m}_{k,l}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)|}{(r\tilde{\pi}^{TOSP}_{\alpha i}) \wedge 1} \\
\le &\frac{1}{r}\sum_{i=1}^{r} \frac{\lambda \psi(\boldsymbol{Z}^*_i)\|\tilde{\boldsymbol{\theta}}^{\alpha}_{r,P} - \hat{\boldsymbol{\theta}}_n\|}{n(\tilde{\pi}^{TOSP*}_{\alpha i} \wedge r^{-1})} \\
= &o_p(1).
\end{aligned}$$

That is to say,

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \frac{v_i \ddot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n + \lambda(\tilde{\boldsymbol{\theta}}^{\alpha}_{r,P} - \hat{\boldsymbol{\theta}}_n))}{(r\tilde{\pi}^{TOSP}_{\alpha i}) \wedge 1} - \frac{1}{n}\sum_{i=1}^{n} \frac{v_i \ddot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)}{(r\tilde{\pi}^{TOSP}_{\alpha i}) \wedge 1} \right\| = o_p(1). \tag{41}$$

Second, for every $k, l = 1, 2, ..., p$, we have

$$\mathbb{E}\left\{ \frac{1}{n}\sum_{i=1}^{n} \frac{v_i \ddot{m}_{k,l}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)}{(r\tilde{\pi}^{TOSP}_{\alpha i}) \wedge 1} \middle| \mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,P} \right\} = \frac{1}{n}\sum_{i=1}^{n} \ddot{m}_{k,l}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n),$$

and

$$
\begin{aligned}
&Var\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{v_i\ddot{m}_{k,l}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)}{(r\tilde{\pi}_{\alpha i}^{TOSP})\wedge 1}\middle|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_{0,P}\right\}\\
\leq&\frac{1}{n^2}\sum_{i=1}^{n}\frac{\ddot{m}_{k,l}^2(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)}{(r\tilde{\pi}_{\alpha i}^{TOSP})\wedge 1}\\
\leq&\frac{1}{\alpha r}\frac{1}{n}\sum_{i=1}^{n}\ddot{m}_{k,l}^2(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\\
=&O_p(r^{-1}).
\end{aligned}
$$

Therefore, according to Chebyshev's inequality, conditional on $\mathcal{D}_n$ and $\tilde{\boldsymbol{\theta}}_{0,P}$, we have

$$
\left\|\frac{1}{n}\sum_{i=1}^{n}\frac{v_i\ddot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)}{(r\tilde{\pi}_{\alpha i}^{TOSP})\wedge 1}-\frac{1}{n}\sum_{i=1}^{n}\ddot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\right\|=o_p(1). \tag{42}
$$

Combining (40)-(42), we finish the proof of Lemma 6.

Lemma 7 *Let* $M_{P\alpha}^*(\boldsymbol{\theta})=(1/n)\sum_{i=1}^{n}v_i m(\boldsymbol{Z}_i,\boldsymbol{\theta})/\{(r\tilde{\pi}_{\alpha i}^{TOSP})\wedge 1\}$. *Under Assumption 4, conditional on* $\mathcal{D}_n$ *and* $\tilde{\boldsymbol{\theta}}_{0,P}$*,*

$$
\sqrt{r}\{\Lambda_P^{\alpha}(\tilde{\boldsymbol{\theta}}_{0,P})\}^{-1/2}\dot{M}_{P\alpha}^*(\hat{\boldsymbol{\theta}}_n)\overset{|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_{0,P}}{\rightsquigarrow}N(\boldsymbol{0},I_p).
$$

Proof Note that

$$
\sqrt{r}\dot{M}_{P\alpha}^*(\hat{\boldsymbol{\theta}}_n)=\sum_{i=1}^{n}\frac{v_i\sqrt{r}\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)}{n\{(r\tilde{\pi}_{\alpha i}^{TOSP})\wedge 1\}}.
$$

Let $\xi_i^*=v_i\sqrt{r}\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)/[n\{(r\tilde{\pi}_{\alpha i}^{TOSP})\wedge 1\}]$. Thus, we have

$$
\mathbb{E}(\sqrt{r}\dot{M}_{P\alpha}^*(\hat{\boldsymbol{\theta}}_n)|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_{0,P})=\frac{\sqrt{r}}{n}\sum_{i=1}^{n}\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)=0,
$$

and

$$
\begin{aligned}
&Var(\sqrt{r}\dot{M}_{P\alpha}^*(\hat{\boldsymbol{\theta}}_n)|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_{0,P})\\
=&\frac{r}{n^2}\sum_{i=1}^{n}\frac{Var(v_i|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_{0,P})\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\dot{m}^{\top}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)}{\{(r\tilde{\pi}_{\alpha i}^{TOSP})\wedge 1\}^2}\\
=&\frac{r}{n^2}\sum_{i=1}^{n}\frac{\{1-(r\tilde{\pi}_{\alpha i}^{TOSP})\wedge 1\}\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\dot{m}^{\top}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)}{(r\tilde{\pi}_{\alpha i}^{TOSP})\wedge 1}\\
\leq&\frac{1}{\alpha n}\sum_{i=1}^{n}\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\dot{m}^{\top}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\\
=&O_p(1).
\end{aligned}
$$

Besides, for $\forall\epsilon > 0$, and $\delta \in (0, 2]$,

$$
\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}\{\|\xi_i^*\|^2 I(\|\xi_i^*\| > \epsilon)|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,P}\} \\
\leq& \frac{1}{\epsilon^\delta} \sum_{i=1}^{n} \mathbb{E}\{\|\xi_i^*\|^{2+\delta} I(\|\xi_i^*\| > \epsilon)|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,P}\} \\
\leq& \frac{1}{\epsilon^\delta} \sum_{i=1}^{n} \mathbb{E}\{\|\eta_i^*\|^{2+\delta}|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,P}\} \\
=& \frac{1}{\epsilon^\delta} \sum_{i=1}^{n} \frac{r^{1+\delta/2}\|\dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|^{2+\delta}}{n^{2+\delta}\{(r\tilde{\pi}_{\alpha i}^{TOSP}) \wedge 1\}^{1+\delta}} \\
\leq& \frac{1}{r^{\delta/2}\epsilon^\delta \alpha^{1+\delta}} \frac{1}{n} \sum_{i=1}^{n} \|\dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|^{2+\delta} \\
=& O_p(r^{-\delta/2}),
\end{aligned} \tag{43}
$$

where the fifth inequality follows the fact that $n\tilde{\pi}_{\alpha i}^{TOSP} = n\{(1-\alpha)\tilde{\pi}_i^{TOSP} + \alpha/n\} \geq \alpha$, and the final equality holds due to Assumption 4. By the Lindeberg-Feller central limit theorem, we obtain

$$
\sqrt{r}\{\Lambda_P^\alpha(\tilde{\boldsymbol{\theta}}_{0,P})\}^{-1/2}\dot{M}_{P\alpha}^*(\hat{\boldsymbol{\theta}}_n) \overset{|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_{0,P}}{\rightsquigarrow} N(\mathbf{0}, I_p). \tag{44}
$$

Lemma 8 *Under Assumptions 3, 4, and 6,*

1. *Let $\rho_n = r/(bn)$ and $G_{\rho_n} = \|t(\boldsymbol{Z}, \hat{\boldsymbol{\theta}}_n)\|_{\rho_n}$. If $\rho_n \to \rho \in (0,1)$, then*
$$G_0 - G_{\rho_n} = o_p(1).$$
2. *Let $\Psi_{\rho_n} = (1/n)\sum_{i=1}^n\{\|t(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\| \wedge G_{\rho_n}\}$, then*
$$\Psi_0 - \Psi_{\rho_n} = o_p(1).$$
3. *If $\rho = 0$, then $\Psi_{\rho_n} - \Psi_\infty = o_p(1)$, where $\Psi_\infty = (1/n)\sum_{i=1}^n \|t(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|$.*

Proof The proof of this lemma is similar to the proof of Lemma 9 in Wang et al. (2022), except that $\dot{m}(\boldsymbol{Z}, \boldsymbol{\theta})$ is replaced with the corresponding $t(\boldsymbol{Z}, \boldsymbol{\theta})$.

1. Recall that $G_0 = \|t(\boldsymbol{Z}^{0*}, \tilde{\boldsymbol{\theta}}_{0,P}))\|_{\frac{r}{bn}}$. G_0 is the $\lceil r_0^* - r_0^*\rho_n\rceil$th order statistics of $\|t(\boldsymbol{Z}_i^{0*}, \tilde{\boldsymbol{\theta}}_{0,P})\|$, where $i = 1, ..., r_0^*$. For any $\rho > 0$, let $\tilde{G}_\rho$ denote the $\lceil n(1-\rho)\rceil$-th order statistics of $\|t(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,P}))\|$ for $i = 1, \ldots, n$. Define the indicator function $v_{(i)}^0$ such that $v_{(i)}^0 = 1$ if $\|t(\boldsymbol{Z}, \tilde{\boldsymbol{\theta}}_{0,P})\|_{(i)}$ is included in the set $\{\|t(\boldsymbol{Z}_i^{0*}, \tilde{\boldsymbol{\theta}}_{0,P})\|\}_{i=1}^{r_0^*}$, and $v_{(i)}^0 = 0$ otherwise.

   For any $\rho_+ > \rho$,
$$P(G_0 \leq \tilde{G}_{\rho_+}) = P\left(\sum_{i=1}^{\lceil n(1-\rho_+)\rceil} v_{(i)}^0 \geq \lceil r_0^* - r_0^*\rho_n\rceil\right).$$

Note that

$$\frac{1}{r_0}\sum_{i=1}^{\lceil n(1-\rho_+)\rceil} v_{(i)}^0 = 1-\rho_+ + o_p(1),$$

and

$$\frac{\lceil r_0^* - r_0^*\rho_n\rceil}{r_0} = 1-\rho + o_p(1).$$

Thus,

$$P(G_0 \le \tilde{G}_{\rho_+}) \to 0. \tag{45}$$

Similarly, for any $\rho_- < \rho$,

$$P(G_0 \le \tilde{G}_{\rho_-}) \to 1. \tag{46}$$

Note that $\tilde{G}_{\rho_+}$ lies between the $\lceil n(1-\rho_+)\rceil - r_0^*$-th and the $\lceil n(1-\rho_+)\rceil$-th order statistics of $\|t(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,P})\|$'s that are not included in $\{\|t(\boldsymbol{Z}_i^{0*}, \tilde{\boldsymbol{\theta}}_{0,P})\|\}_{i=1}^{r_0^*}$. Since the joint distribution of these $\|t(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,P})\|$'s are exchangeable and $r_0^*/n \to 0$ in probability, it follows that both the $\lceil n(1-\rho_+)\rceil - r_0^*$-th and $\lceil n(1-\rho_+)\rceil$-th order statistics of these values converge in probability to the ρ_+-quantile of the distribution of $\|t(\boldsymbol{Z}, \boldsymbol{\theta}_0)\|$, denoted as ζ_{ρ_+}, where $\boldsymbol{\theta}_0 = \arg\max_{\boldsymbol{\theta}} \mathbb{E}\{m(\boldsymbol{Z}, \boldsymbol{\theta})\}$(Chanda, 1971). Consequently, $\tilde{G}_{\rho_+}$ converges in probability to ζ_{ρ_+}. Similarly, $\tilde{G}_{\rho_-}$ converges in probability to ζ_{ρ_-}, the ρ_--quantile of the distribution of $\|t(\boldsymbol{Z}, \boldsymbol{\theta}_0)\|$. Combining (45) and (46), we can derive that for any $\epsilon > 0$,

$$P(\zeta_{\rho+} - \epsilon < G_0 < \zeta_{\rho-} + \epsilon) \to 1.$$

Since the distribution of $\boldsymbol{Z}$ is continuous, the same holds for $\|t(\boldsymbol{Z}, \boldsymbol{\theta}_0)\|$. Thus, we can select ρ_+ and ρ_- sufficiently close to ρ such that $\zeta_{\rho_-} - \zeta_\rho < \epsilon$ and $\zeta_\rho - \zeta_{\rho_+} < \epsilon$, which implies that for any $\epsilon > 0$,

$$P(\zeta_\rho - 2\varepsilon < G_0 < \zeta_\rho + 2\epsilon) \to 1.$$

Consequently, we obtain $G_0 = \zeta_\rho + o_p(1)$. Similarly, since $\|t(Z_i, \hat{\boldsymbol{\theta}}_n)\|$'s are exchangeable, it follows that

$$G_{\rho_n} = \zeta_\rho + o_p(1).$$

Therefore, we conclude that $G_0 - G_{\rho_n} = o_p(1)$.

2. Recall that

$$\Psi_0 = \frac{1}{r_0^*}\sum_{i=1}^{r_0^*}\{\|t(\boldsymbol{Z}_i^{0*}, \tilde{\boldsymbol{\theta}}_{0,P})\| \wedge G_0\}.$$

When $\rho = 0$ and $\|t(\boldsymbol{Z}, \boldsymbol{\theta})\|$ is bounded,

$$\Psi_0 = \sum_{i=1}^{\lceil r_0^* - r_0^*\rho_n\rceil} \frac{\|t(\boldsymbol{Z}^{0*}, \tilde{\boldsymbol{\theta}}_{0,P})\|_{(i)}}{r_0^*} + \frac{r_0^* - \lceil r_0^* - r_0^*\rho_n\rceil}{r_0^*} G_0 = \frac{1}{r_0^*}\sum_{i=1}^{r_0^*}\|t(\boldsymbol{Z}_i^{0*}, \tilde{\boldsymbol{\theta}}_{0,P})\| + o_p(1).$$

Similarly,

$$\Psi_{\rho_n} = \frac{1}{n}\sum_{i=1}^{n}\|t(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\| + o_p(1).$$

By Taylor expansion and Markov's inequality, we can derive that

$$\frac{1}{r_0^*}\sum_{i=1}^{r_0^*}\|t(\boldsymbol{Z}_i^{0*},\tilde{\boldsymbol{\theta}}_{0,P})\| = \frac{1}{n}\sum_{i=1}^{n}\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\| + o_p(1).$$

That is to say,

$$\Psi_0 - \Psi_{\rho_n} = o_p(1).$$

To establish the proof for other cases, let $v_i^0 = 1$ if the i-th observation is included in the pilot subsample, and $v_i^0 = 0$ otherwise. Thus,

$$\Psi_0 = \frac{1}{r_0^*}\sum_{i=1}^{n} v_i^0\{\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\| \wedge G_0\}.$$

Let

$$\tilde{\Psi}_0 = \frac{1}{r_0^*}\sum_{i=1}^{n} v_i^0\{\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\| \wedge G_{\rho_n}\},$$

and

$$\tilde{\Psi}_{\rho_n} = \frac{1}{r_0^*}\sum_{i=1}^{n} v_i^0\{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\| \wedge G_{\rho_n}\}.$$

Note that

$$|\Psi_0 - \Psi_{\rho_n}| \le |\Psi_0 - \tilde{\Psi}_0| + |\tilde{\Psi}_0 - \tilde{\Psi}_{\rho_n}| + |\tilde{\Psi}_{\rho_n} - \Psi_{\rho_n}|.$$

We will prove

$$|\Psi_0 - \tilde{\Psi}_0| = o_p(1), \tag{47}$$

$$|\tilde{\Psi}_0 - \tilde{\Psi}_{\rho_n}| = o_p(1), \tag{48}$$

and

$$|\tilde{\Psi}_{\rho_n} - \Psi_{\rho_n}| = o_p(1) \tag{49}$$

separately.

Firstly, when $\rho = 0$ and $\|t(\boldsymbol{Z},\boldsymbol{\theta})\|$ is unbounded, then $G_0 \wedge G_{\rho_n} \to \infty$ in probability. Therefore,

$$\begin{aligned}
|\Psi_0 - \tilde{\Psi}_0| \le & \frac{1}{r_0^*}\sum_{i=1}^{n} v_i^0\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\| I\{\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\| \ge G_0 \wedge G_{\rho_n}\} \\
& + \frac{G_0}{r_0^*}\sum_{i=1}^{n} v_i^0 I\{\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\| \ge G_0\} + \frac{G_{\rho_n}}{r_0^*}\sum_{i=1}^{n} v_i^0 I\{\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\| \ge G_{\rho_n}\} \\
\le & \left\{\frac{1}{G_0 \wedge G_{\rho_n}} + \frac{1}{G_0} + \frac{1}{G_{\rho_n}}\right\}\frac{1}{r_0^*}\sum_{i=1}^{n} v_i^0\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\|^2 \\
= & o_p(1).
\end{aligned} \tag{50}$$

When $\rho \in (0,1)$,

$$
\begin{aligned}
|\Psi_0 - \tilde{\Psi}_0| \leq & \frac{1}{r_0^*} \sum_{i=1}^n v_i^0 |\|t(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,P})\| \wedge G_0 - \|t(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,P})\| \wedge G_{\rho_n}| \\
\leq & \frac{|G_0 - G_{\rho_n}|}{r_0^*} \sum_{i=1}^n v_i^0 I\{\|t(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,P})\| \geq G_0 \wedge G_{\rho_n}\} \\
\leq & |G_0 - G_{\rho_n}| \\
= & o_p(1).
\end{aligned} \tag{51}
$$

Combining (50) and (51), (47) holds.

Furthermore,

$$
\begin{aligned}
|\tilde{\Psi}_0 - \tilde{\Psi}_{\rho_n}| \leq & \frac{1}{r_0^*} \sum_{i=1}^n v_i^0 |\|t(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,P})\| \wedge G_{\rho_n} - \|t(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\| \wedge G_{\rho_n}| \\
\leq & \frac{1}{r_0^*} \sum_{i=1}^n v_i^0 |\|t(\boldsymbol{Z}_i, \tilde{\boldsymbol{\theta}}_{0,P})\| - \|t(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|| \\
\leq & \frac{O_p(\|\tilde{\boldsymbol{\theta}}_{0,P} - \hat{\boldsymbol{\theta}}_n\|)}{r_0^*} \sum_{i=1}^n v_i^0 h(\boldsymbol{Z}_i) + \frac{O_p(\|\tilde{\boldsymbol{\theta}}_{0,P} - \hat{\boldsymbol{\theta}}_n\|)}{r_0^*} \sum_{i=1}^n v_i^0 \|\dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\| \\
& + \frac{O_p(r_0^{-1})}{r_0^*} \sum_{i=1}^n v_i^0 \|\dot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\| \\
= & o_p(1),
\end{aligned}
$$

where the third inequality follows (29). (48) holds.

Then, by applying the mean and variance calculations, along with Chebyshev's inequality, we establish the validity of (49). This completes the proof.

3. If $\|t(\boldsymbol{Z}, \boldsymbol{\theta})\|$ is bounded,

$$
\begin{aligned}
|\Psi_{\rho_n} - \Psi_\infty| \leq & \frac{1}{n} \sum_{i=1}^n |\|t(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\| - \|t(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\| \wedge G_{\rho_n}| \\
\leq & \frac{n - \lceil n(1-\rho_n) \rceil}{n} \|t(\boldsymbol{Z}, \hat{\boldsymbol{\theta}}_n)\|_{(n)} = o_p(1);
\end{aligned}
$$

otherwise,

$$
\begin{aligned}
|\Psi_{\rho_n} - \Psi_\infty| \leq & \frac{1}{n} \sum_{i=1}^n |\|t(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\| - \|t(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\| \wedge G_{\rho_n}| \\
\leq & \frac{1}{n} \sum_{i=1}^n \|t(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\| I\{\|t(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\| \geq G_{\rho_n}\} \\
\leq & \frac{1}{n G_{\rho_n}} \sum_{i=1}^n \|t(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n)\|^2 = o_p(1).
\end{aligned}
$$

This completes the proof.

Lemma 9 *Under Assumptions 3, 4, and 6, conditional on $\mathcal{D}_n$ and $\tilde{\boldsymbol{\theta}}_{0,P}$,*

$$\|\Lambda_P^\alpha(\tilde{\boldsymbol{\theta}}_{0,P}) - \Lambda_P^\alpha(\hat{\boldsymbol{\theta}}_n)\| = o_p(1).$$

Proof Note that

$$\begin{aligned}
&\|\Lambda_P^\alpha(\tilde{\boldsymbol{\theta}}_{0,P}) - \Lambda_P^\alpha(\hat{\boldsymbol{\theta}}_n)\| \\
=&\left\| \frac{r}{n^2}\sum_{i=1}^n \frac{\{1-(r\tilde{\pi}_{\alpha i}^{TOSP})\wedge 1\}\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\dot{m}^\top(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)}{(r\tilde{\pi}_{\alpha i}^{TOSP})\wedge 1} \right. \\
&\left. - \frac{r}{n^2}\sum_{i=1}^n \frac{\{1-(r\pi_{\alpha i}^{TOSP}(\hat{\boldsymbol{\theta}}_n)\wedge 1\}\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\dot{m}^\top(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)}{(r\pi_{\alpha i}^{TOSP}(\hat{\boldsymbol{\theta}}_n))\wedge 1} \right\| \\
\leq&\frac{r}{n^2}\sum_{i=1}^n \|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2 \left| \frac{1}{(r\tilde{\pi}_{\alpha i}^{TOSP})\wedge 1} - \frac{1}{(r\pi_{\alpha i}^{TOSP}(\hat{\boldsymbol{\theta}}_n))\wedge 1} \right| \\
\leq&\frac{1}{\alpha^2}\sum_{i=1}^n \|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2 |\tilde{\pi}_i^{TOSP} - \pi_i^{TOSP}(\hat{\boldsymbol{\theta}}_n)|, \qquad (52)
\end{aligned}$$

where

$$\pi_i^{TOSP}(\hat{\boldsymbol{\theta}}_n) = \begin{cases} \frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|}{n\Psi_\infty} & \text{If} \quad \rho = 0; \\ \frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\wedge G_{\rho_n}}{n\Psi_{\rho_n}} & \text{Otherwise.} \end{cases}$$

If $\rho = 0$,

$$\begin{aligned}
&n|\tilde{\pi}_i^{TOSP} - \pi_i^{TOSP}(\hat{\boldsymbol{\theta}}_n)| \\
=&\left| \frac{\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\|\wedge G_0}{\Psi_0} - \frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|}{\Psi_\infty} \right| \\
\leq&\left| \frac{\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\|\wedge G_0}{\Psi_0} - \frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\wedge G_{\rho_n}}{\Psi_0} \right| + \left| \frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\wedge G_{\rho_n}}{\Psi_0} - \frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\wedge G_{\rho_n}}{\Psi_{\rho_n}} \right| \\
&+ \left| \frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\wedge G_{\rho_n}}{\Psi_{\rho_n}} - \frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|}{\Psi_{\rho_n}} \right| + \left| \frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|}{\Psi_{\rho_n}} - \frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|}{\Psi_\infty} \right| \\
\leq&\frac{|\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\|\wedge G_0 - \|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\wedge G_\rho|}{\Psi_0} + \|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\frac{|\Psi_0 - \Psi_{\rho_n}|}{\Psi_0\Psi_{\rho_n}} \\
&+ \frac{|\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\wedge G_{\rho_n} - \|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\||}{\Psi_{\rho_n}} + \|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\frac{|\Psi_{\rho_n} - \Psi_\infty|}{\Psi_{\rho_n}\Psi_\infty} \\
\leq&\frac{|\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\| - \|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\||}{\Psi_0} + \|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\frac{|\Psi_0 - \Psi_{\rho_n}|}{\Psi_0\Psi_{\rho_n}} \\
&+ \frac{\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\|}{\Psi_0} I\{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\| \geq G_{\rho_n}\} + \frac{G_{\rho_n}}{\Psi_0} I\{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\| \geq G_{\rho_n}\} \\
&+ \frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|}{\Psi_0} I\{\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\| \geq G_0\} + \frac{G_0}{\Psi_0} I\{\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\| \geq G_0\}
\end{aligned}$$

$$
\begin{aligned}
&+\frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2}{\Psi_{\rho_n}G_{\rho_n}}+\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\frac{|\Psi_{\rho_n}-\Psi_\infty|}{\Psi_{\rho_n}\Psi_\infty}\\
\triangleq&\Delta_{3i}+\Delta_{4i}+\Delta_{5i}+\Delta_{6i}+\Delta_{7i}+\Delta_{8i}+\Delta_{9i}+\Delta_{10i}. \qquad (53)
\end{aligned}
$$

From (29), Assumptions 4 and 6, and Lemma 8, we can derive that

$$\frac{1}{n}\sum_{i=1}^n\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\Delta_{3i}=o_p(1), \tag{54}$$

$$\frac{1}{n}\sum_{i=1}^n\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\Delta_{4i}=o_p(1), \tag{55}$$

$$\frac{1}{n}\sum_{i=1}^n\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\Delta_{9i}=o_p(1), \tag{56}$$

and

$$\frac{1}{n}\sum_{i=1}^n\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\Delta_{10i}=o_p(1), \tag{57}$$

Then,

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\Delta_{5i}\\
=&\frac{1}{n}\sum_{i=1}^n\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\frac{\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\|}{\Psi_0}I\{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\ge G_{\rho_n}\}\\
\le&\frac{\|\dot{\mu}(\tilde{\boldsymbol{\theta}}_{0,P})\|\|\ddot{M}^{-1}(\tilde{\boldsymbol{\theta}}_{0,P})\|}{\Psi_0}\frac{1}{n}\sum_{i=1}^n\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\|\dot{m}(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\|I\{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\ge G_{\rho_n}\}\\
\le&\frac{\|\dot{\mu}(\tilde{\boldsymbol{\theta}}_{0,P})\|\|\ddot{M}^{-1}(\tilde{\boldsymbol{\theta}}_{0,P})\|}{\Psi_0}\left\{\frac{1}{n}\sum_{i=1}^n\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^4\right\}^{1/2}\left\{\frac{1}{n}\sum_{i=1}^n\|\dot{m}(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\|^4\right\}^{1/4}\\
&\left[\frac{1}{n}\sum_{i=1}^n I\{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\ge G_{\rho_n}\}\right]^{1/4}\\
=&o_p(1), \qquad (58)
\end{aligned}
$$

where the last equation is obtained by $\rho=0$ and $(1/n)\sum_{i=1}^n I\{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\ge G_{\rho_n}\}=o_p(1)$. Similarly,

$$\frac{1}{n}\sum_{i=1}^n\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\Delta_{7i}=o_p(1). \tag{59}$$

If $\|t(\boldsymbol{Z},\boldsymbol{\theta})\|$ is bounded,

$$\frac{1}{n}\sum_{i=1}^n\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\Delta_{6i}$$

$$
\begin{aligned}
=&\frac{G_{\rho_n}}{\Psi_0}\frac{1}{n}\sum_{i=1}^{n}\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2 I\{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\geq G_{\rho_n}\}\\
\leq&\frac{G_{\rho_n}}{\Psi_0}\frac{1}{n}\sum_{i=\lceil n(1-\rho_n)\rceil}^{n}\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\\
\leq&\frac{G_{\rho_n}}{\Psi_0}\frac{n-\lceil n(1-\rho_n)\rceil}{n}\|\dot{m}(\boldsymbol{Z},\hat{\boldsymbol{\theta}}_n)\|^2_{(n)}\\
=&o_p(1).
\end{aligned}
$$

Otherwise,

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\Delta_{6i}\\
=&\frac{G_{\rho_n}}{\Psi_0}\frac{1}{n}\sum_{i=1}^{n}\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2 I\{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\geq G_{\rho_n}\}\\
\leq&\frac{1}{\Psi_0 G_{\rho_n}}\frac{1}{n}\sum_{i=1}^{n}\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\\
\leq&\frac{\|\dot{\mu}(\hat{\boldsymbol{\theta}}_n)\|^2\|\ddot{M}^{-1}(\hat{\boldsymbol{\theta}}_n)\|^2}{\Psi_0 G_{\rho_n}}\frac{1}{n}\sum_{i=1}^{n}\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^4\\
=&o_p(1).
\end{aligned}
$$

That is to say,

$$
\frac{1}{n}\sum_{i=1}^{n}\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\Delta_{6i}=o_p(1). \tag{60}
$$

Similarly,

$$
\frac{1}{n}\sum_{i=1}^{n}\|\dot{m}(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|^2\Delta_{8i}=o_p(1). \tag{61}
$$

Combining (52)-(61), if $\rho=0$, $\|\Lambda_P^{\alpha}(\tilde{\boldsymbol{\theta}}_{0,P})-\Lambda_P^{\alpha}(\hat{\boldsymbol{\theta}}_n)\|=o_p(1)$ holds.

If $\rho\in(0,1)$,

$$
\begin{aligned}
&n|\tilde{\pi}_i^{TOSP}-\pi_i^{TOSP}(\hat{\boldsymbol{\theta}}_n)|\\
=&\left|\frac{\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\|\wedge G_0}{\Psi_0}-\frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\wedge G_{\rho_n}}{\Psi_{\rho_n}}\right|\\
\leq&\left|\frac{\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\|\wedge G_0}{\Psi_0}-\frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\wedge G_{\rho_n}}{\Psi_0}\right|+\left|\frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\wedge G_{\rho_n}}{\Psi_0}-\frac{\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\wedge G_{\rho_n}}{\Psi_{\rho_n}}\right|\\
\leq&\frac{|\|t(\boldsymbol{Z}_i,\tilde{\boldsymbol{\theta}}_{0,P})\|-\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\||}{\Psi_0}+\frac{|G_0-G_{\rho_n}|}{\Psi_0}+\|t(\boldsymbol{Z}_i,\hat{\boldsymbol{\theta}}_n)\|\frac{|\Psi_0-\Psi_{\rho_n}|}{\Psi_0\Psi_{\rho_n}}.
\end{aligned}
$$

Combining (29), Assumptions 4 and 6, and Lemma 8, we can show that $\|\Lambda_P^{\alpha}(\tilde{\boldsymbol{\theta}}_{0,P})-\Lambda_P^{\alpha}(\hat{\boldsymbol{\theta}}_n)\|=o_p(1)$ holds. This completes the proof.

Proof of Theorem 6:

Similar to the proof of Theorem 5, we begin by proving

$$\sqrt{r}\{V_P^{\alpha}(\hat{\boldsymbol{\theta}}_n)\}^{-1/2}(\tilde{\boldsymbol{\theta}}_{r,P}^{\alpha}-\hat{\boldsymbol{\theta}}_n) \overset{|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_{0,P}}{\rightsquigarrow} N(\mathbf{0},I_p). \tag{62}$$

Note that

$$\tilde{\boldsymbol{\theta}}_{r,P}^{\alpha}=\arg\max_{\boldsymbol{\theta}}\frac{1}{n}\sum_{i=1}^{n}\frac{v_i m(\boldsymbol{Z}_i,\boldsymbol{\theta})}{(r\tilde{\pi}_{\alpha i}^{TOSP})\wedge 1}=\arg\max_{\boldsymbol{\theta}} M_{P\alpha}^*(\boldsymbol{\theta}).$$

We can obtain that for every $\boldsymbol{\theta}\in\boldsymbol{\Theta}$,

$$\begin{aligned}\mathbb{E}\{M_{P\alpha}^*(\boldsymbol{\theta})|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_{0,P}\} =&\mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{v_i m(\boldsymbol{Z}_i,\boldsymbol{\theta})}{(r\tilde{\pi}_{\alpha i}^{TOSP})\wedge 1}\middle|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_{0,R}\right\}\\ =&\frac{1}{n}\sum_{i=1}^{n}m(\boldsymbol{Z}_i,\boldsymbol{\theta})\\ =&M_n(\boldsymbol{\theta}),\end{aligned}$$

and

$$\begin{aligned}Var\{M_{P\alpha}^*(\boldsymbol{\theta})|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_{0,P}\} =&\frac{1}{n^2}\sum_{i=1}^{n}\frac{Var(v_i|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_{0,P})m(\boldsymbol{Z}_i,\boldsymbol{\theta})m^{\top}(\boldsymbol{Z}_i,\boldsymbol{\theta})}{\{(r\tilde{\pi}_{\alpha i}^{TOSP})\wedge 1\}^2}\\ \le&\frac{1}{n^2}\sum_{i=1}^{n}\frac{\mathbb{E}(v_i^2|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_{0,P})m(\boldsymbol{Z}_i,\boldsymbol{\theta})m^{\top}(\boldsymbol{Z}_i,\boldsymbol{\theta})}{\{(r\tilde{\pi}_{\alpha i}^{TOSP})\wedge 1\}^2}\\ =&\frac{1}{n^2}\sum_{i=1}^{n}\frac{m(\boldsymbol{Z}_i,\boldsymbol{\theta})m^{\top}(\boldsymbol{Z}_i,\boldsymbol{\theta})}{(r\tilde{\pi}_{\alpha i}^{TOSP})\wedge 1}\\ \le&\frac{1}{\alpha rn}\sum_{i=1}^{n}m(\boldsymbol{Z}_i,\boldsymbol{\theta})m^{\top}(\boldsymbol{Z}_i,\boldsymbol{\theta})\\ =&O_p(r^{-1}),\end{aligned}$$

where the fourth inequality follows the fact that $n\tilde{\pi}_{\alpha i}^{TOSP}=n\{(1-\alpha)\tilde{\pi}_i^{TOSP}+\alpha/n\}\ge\alpha$, and the final equality holds due to Assumption 2.

By Chebyshev's inequality, for any $\epsilon>0$, we have

$$P(|M_{P\alpha}^*(\boldsymbol{\theta})-M_n(\boldsymbol{\theta})|>\epsilon|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_{0,P})\le\frac{1}{\epsilon^2}Var\{M_{P\alpha}^*(\boldsymbol{\theta})|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_{0,P}\}=O_p(r^{-1}),$$

which implies that for every $\boldsymbol{\theta}\in\boldsymbol{\Theta}$,

$$M_{P\alpha}^*(\boldsymbol{\theta})-M_n(\boldsymbol{\theta})=o_{p|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_{0,P}}(1). \tag{63}$$

Combining the above equation with Assumptions 1 and 2, as well as Theorem 5.9 and the corresponding remark on Van der Vaart (2000), we can derive that

$$\|\tilde{\boldsymbol{\theta}}_{r,P}^{\alpha}-\hat{\boldsymbol{\theta}}_n\|=o_p(1). \tag{64}$$

Then, by Taylor expansion,

$$0 = \dot{M}^*_{P\alpha}(\tilde{\boldsymbol{\theta}}^{\alpha}_{r,P}) = \dot{M}^*_{P\alpha}(\hat{\boldsymbol{\theta}}_n) + D^{\alpha}_r(\tilde{\boldsymbol{\theta}}^{\alpha}_{r,P} - \hat{\boldsymbol{\theta}}_n), \tag{65}$$

where

$$D^{\alpha}_r = \int_0^1 \ddot{M}^*_{P\alpha}(\hat{\boldsymbol{\theta}}_n + \lambda(\tilde{\boldsymbol{\theta}}^{\alpha}_{r,P} - \hat{\boldsymbol{\theta}}_n))d\lambda = \int_0^1 \frac{1}{n}\sum_{i=1}^{n} \frac{v_i \ddot{m}(\boldsymbol{Z}_i, \hat{\boldsymbol{\theta}}_n + \lambda(\tilde{\boldsymbol{\theta}}^{\alpha}_{r,P} - \hat{\boldsymbol{\theta}}_n))}{(r\tilde{\pi}^{TOSP}_{\alpha i}) \wedge 1} d\lambda.$$

Thus,

$$\tilde{\boldsymbol{\theta}}^{\alpha}_{r,P} - \hat{\boldsymbol{\theta}}_n = -(D^{\alpha}_r)^{-1}\dot{M}^*_{P\alpha}(\hat{\boldsymbol{\theta}}_n). \tag{66}$$

Combining Lemmas 6-9, according to Slutsky's theorem, equation (62) holds. Then, by applying the Delta method, Theorem 6 is proved.

S.2 Additional Experimental Results

S.2.1 The eMSE Results Under Other Scenarios

The simulation setup in this section is the same as that in Section 5. In the experiments of this section, we additionally consider other scenarios of μ. The design of μ is divided into the following scenarios:

Scenario 4: $\mu(\boldsymbol{\theta}) = \boldsymbol{a}_2^\top \boldsymbol{\theta}$, where $\boldsymbol{a}_2 = (1, 0, 0, 0, 0, 0, 0)^\top$;

Scenario 5: $\mu(\boldsymbol{\theta}) = \exp(\boldsymbol{a}_2^\top \boldsymbol{\theta})$.

The experimental results are presented in the following figures.

S.2.2 The Time Results of Other Experiments

In this section, we present the time results of all the simulation experiments, excluding those already shown in Section 5.

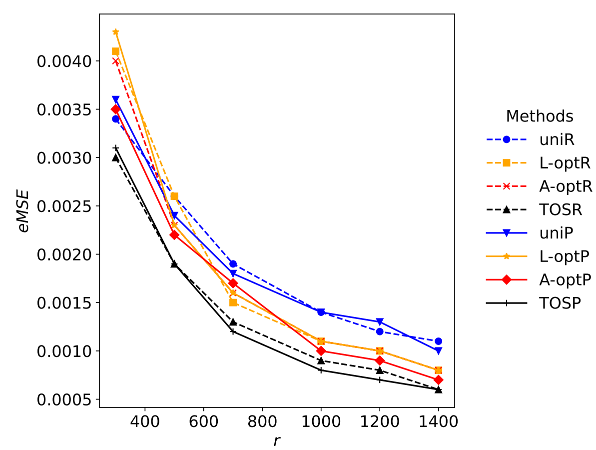
References

Chanda, K. (1971): "Asymptotic distribution of sample quantiles for exchangeable random variables," *Calcutta Statistical Association Bulletin*, 20, 135–142.

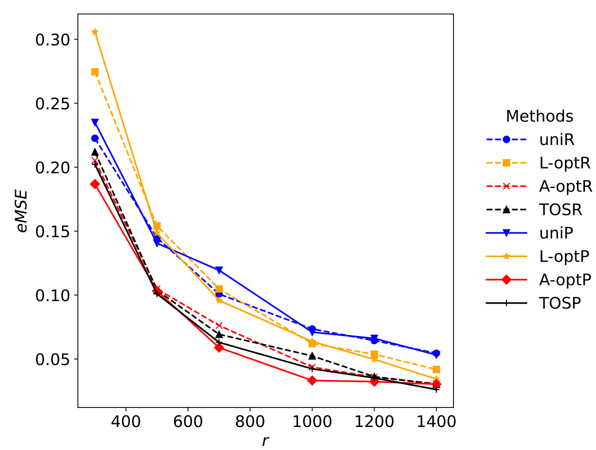
Van der Vaart, A. W. (2000): Asymptotic statistics, vol. 3, Cambridge university press.

Wang, J., J. Zou, and H. Wang (2022): "Sampling with replacement vs Poisson sampling: A comparative study in optimal subsampling," *IEEE Transactions on Information Theory*, 68, 6605–6630.

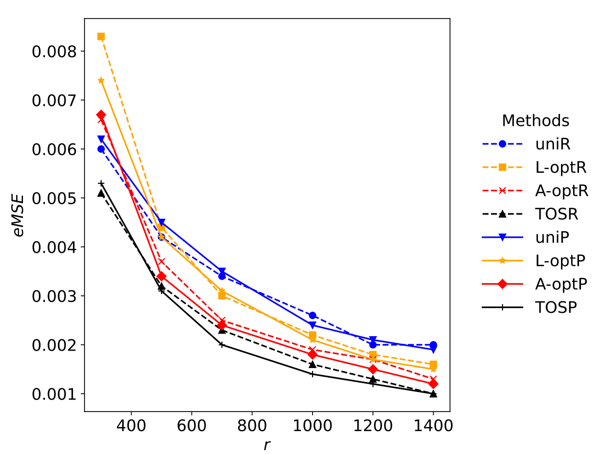
Figure 1: The eMSEs of Poisson regression for different subsampling methods in Scenario 4 when r varies.



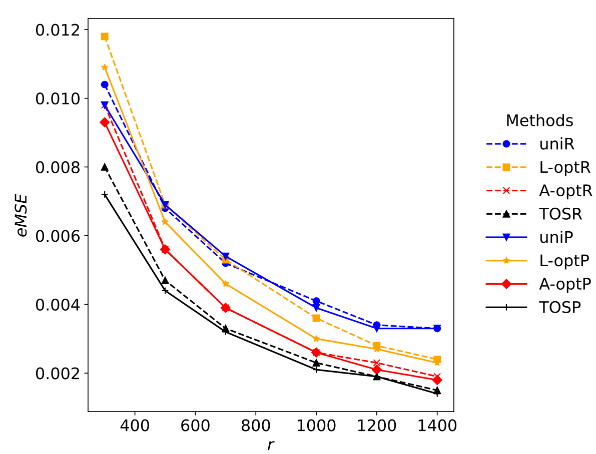
(a) Case 1: independent covariates



(b) Case 2: highly correlated covariates

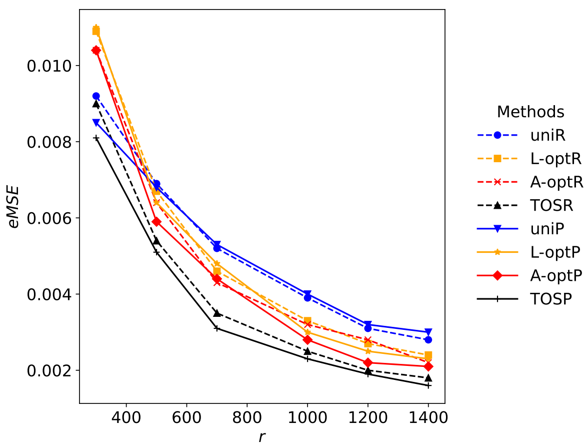


(c) Case 3: weakly correlated covariates

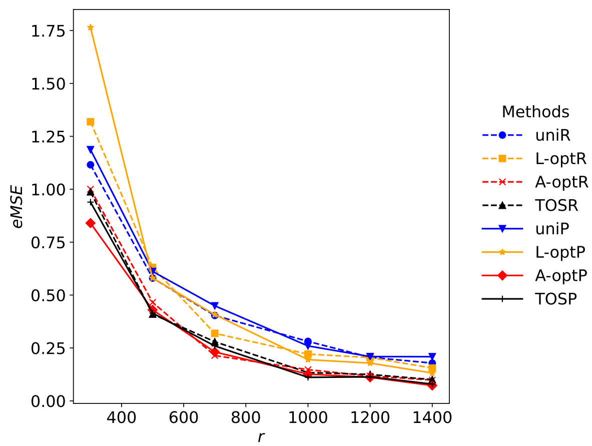


(d) Case 4: unequal bounds of covariates

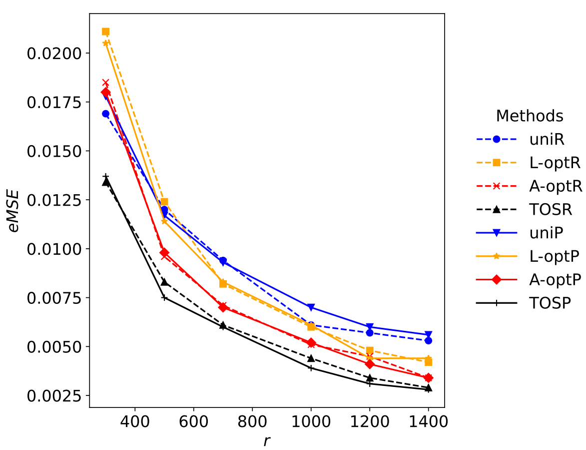
Figure 2: The eMSEs of Poisson regression for different subsampling methods in Scenario 5 when r varies.



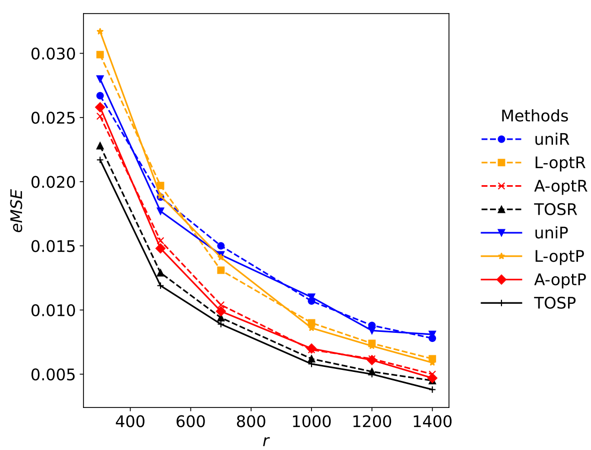
(a) Case 1: independent covariates



(b) Case 2: highly correlated covariates



(c) Case 3: weakly correlated covariates



(d) Case 4: unequal bounds of covariates

Table 1: The average computation time per repetition (measured in seconds) and their standard deviations of Poisson regression for different subsampling methods under Case 1 settings when r varies.

r Values	Full	uniR	L-optR	A-optR	TOSR	uniP	L-optP	A-optP	TOSP
Scenario 4: $\mu(\boldsymbol{\theta}) = \boldsymbol{a}_2^\top \boldsymbol{\theta}$, where $\boldsymbol{a}_2 = (1,0,0,0,0,0,0)^\top$.									
300	1.6107(0.0000)	0.0239(0.0136)	0.1696(0.0161)	0.1731(0.0162)	0.1379(0.0122)	0.0242(0.0054)	0.1688(0.0164)	0.1748(0.0226)	0.1391(0.0126)
500	1.6107(0.0000)	0.0234(0.0104)	0.1687(0.0154)	0.1711(0.0149)	0.1368(0.0125)	0.0236(0.0048)	0.1661(0.0157)	0.1729(0.0213)	0.1372(0.0126)
700	1.6107(0.0000)	0.0263(0.0090)	0.1678(0.0158)	0.1732(0.0157)	0.1370(0.0117)	0.0258(0.0049)	0.1663(0.0153)	0.1731(0.0217)	0.1378(0.0128)
1000	1.6107(0.0000)	0.0312(0.0108)	0.1725(0.0139)	0.1784(0.0133)	0.1421(0.0110)	0.0309(0.0058)	0.1742(0.0136)	0.1816(0.0212)	0.1458(0.0128)
1200	1.6107(0.0000)	0.0327(0.0116)	0.1822(0.0145)	0.1901(0.0132)	0.1517(0.0117)	0.0325(0.0059)	0.1822(0.0132)	0.1886(0.0219)	0.1518(0.0137)
1400	1.6107(0.0000)	0.0334(0.0108)	0.1844(0.0141)	0.1922(0.0137)	0.1537(0.0118)	0.0323(0.0060)	0.1847(0.0141)	0.1908(0.0213)	0.1526(0.0128)
Scenario 5: $\mu(\boldsymbol{\theta}) = \exp(\boldsymbol{a}_2^\top \boldsymbol{\theta})$.									
300	1.6107(0.0000)	0.0233(0.0102)	0.1695(0.0157)	0.1716(0.0154)	0.1369(0.0126)	0.0243(0.0053)	0.1679(0.0161)	0.1731(0.0216)	0.1389(0.0133)
500	1.6107(0.0000)	0.0236(0.0101)	0.1696(0.0156)	0.1716(0.0147)	0.1369(0.0123)	0.0236(0.0049)	0.1667(0.0157)	0.1729(0.0226)	0.1385(0.0129)
700	1.6107(0.0000)	0.0268(0.0113)	0.1688(0.0156)	0.1722(0.0149)	0.1372(0.0117)	0.0261(0.0051)	0.1664(0.0157)	0.1735(0.0210)	0.1375(0.0129)
1000	1.6107(0.0000)	0.0320(0.0145)	0.1730(0.0142)	0.1798(0.0139)	0.1430(0.0114)	0.0313(0.0056)	0.1755(0.0151)	0.1834(0.0226)	0.1473(0.0137)
1200	1.6107(0.0000)	0.0327(0.0115)	0.1803(0.0140)	0.1878(0.0136)	0.1509(0.0115)	0.0323(0.0059)	0.1814(0.0135)	0.1871(0.0203)	0.1513(0.0129)
1400	1.6107(0.0000)	0.0340(0.0114)	0.1879(0.0147)	0.1956(0.0145)	0.1564(0.0122)	0.0331(0.0062)	0.1868(0.0136)	0.1935(0.0218)	0.1559(0.0132)

Table 2: The average computation time per repetition (measured in seconds) and their standard deviations of Poisson regression for different subsampling methods under Case 2 settings when r varies.

r Values	Full	uniR	L-optR	A-optR	TOSR	uniP	L-optP	A-optP	TOSP
Scenario 1: $\mu(\boldsymbol{\theta}) = \boldsymbol{a}_0^\top \boldsymbol{\theta}$, where $\boldsymbol{a}_0 = I_7$ and I_p is a p-dimensional identical matrix.									
300	2.9575(0.0000)	0.2396(0.0710)	0.5651(0.0880)	0.5867(0.0907)	0.6171(0.0918)	0.2352(0.0659)	0.5717(0.0845)	0.5875(0.0968)	0.6174(0.1001)
500	2.9575(0.0000)	0.2556(0.0787)	0.5715(0.0966)	0.5903(0.1005)	0.6297(0.0977)	0.2413(0.0795)	0.5812(0.0928)	0.6008(0.1028)	0.6289(0.1085)
700	2.9575(0.0000)	0.4370(0.1356)	0.5740(0.0947)	0.5633(0.0974)	0.5966(0.0997)	0.4002(0.1457)	0.5818(0.0977)	0.5807(0.1106)	0.6020(0.1061)
1000	2.9575(0.0000)	0.8595(0.3037)	1.1622(0.2865)	1.1414(0.2859)	1.2126(0.2746)	0.8515(0.2965)	1.3999(0.3911)	1.5048(0.4193)	1.5100(0.4400)
1200	2.9575(0.0000)	1.3473(0.5034)	2.5513(0.5435)	2.5066(0.5726)	2.5727(0.5888)	1.3664(0.4770)	2.6176(0.5399)	2.5342(0.6172)	2.5809(0.6393)
1400	2.9575(0.0000)	1.4887(0.5400)	2.8511(0.6529)	2.7641(0.7068)	2.8673(0.7020)	1.4414(0.5420)	2.8404(0.6643)	2.8265(0.7110)	2.8556(0.7246)
Scenario 2: $\mu(\boldsymbol{\theta}) = \boldsymbol{a}_1^\top \boldsymbol{\theta}$, where $\boldsymbol{a}_1 = (1, 0, -1, 0.5, 1.5, -0.5, -1.5)^\top$.									
300	2.9575(0.0000)	0.2379(0.0686)	0.5471(0.0881)	0.5664(0.0871)	0.5392(0.0925)	0.2359(0.0656)	0.5541(0.0829)	0.5754(0.0876)	0.5464(0.0963)
500	2.9575(0.0000)	0.2578(0.0734)	0.5623(0.0952)	0.5828(0.0982)	0.5528(0.0977)	0.2528(0.0730)	0.5763(0.0906)	0.5916(0.0967)	0.5614(0.0969)
700	2.9575(0.0000)	0.4504(0.1339)	0.5682(0.0956)	0.5622(0.0943)	0.5377(0.0936)	0.4060(0.1507)	0.5742(0.0991)	0.5714(0.1054)	0.5461(0.1007)
1000	2.9575(0.0000)	0.8421(0.3052)	1.1934(0.2615)	1.1889(0.2614)	1.1403(0.2556)	0.8815(0.2896)	1.4217(0.3891)	1.4485(0.4036)	1.3930(0.3948)
1200	2.9575(0.0000)	1.4178(0.5029)	2.6387(0.5371)	2.5254(0.5752)	2.4773(0.5543)	1.4322(0.5057)	2.6629(0.5723)	2.6087(0.5946)	2.5399(0.6120)
1400	2.9575(0.0000)	1.5340(0.4866)	2.8832(0.6338)	2.8324(0.6619)	2.7577(0.6822)	1.4689(0.5031)	2.8961(0.6505)	2.9024(0.6457)	2.7624(0.6863)
Scenario 3: $\mu(\boldsymbol{\theta}) = \exp(\boldsymbol{a}_1^\top \boldsymbol{\theta})$.									
300	2.9575(0.0000)	0.2428(0.0704)	0.5515(0.0871)	0.5657(0.0900)	0.5479(0.0864)	0.2364(0.0628)	0.5582(0.0875)	0.5794(0.0902)	0.5522(0.0879)
500	2.9575(0.0000)	0.2524(0.0813)	0.5652(0.0922)	0.5798(0.0961)	0.5520(0.1006)	0.2452(0.0779)	0.5762(0.0962)	0.5902(0.1010)	0.5608(0.1033)
700	2.9575(0.0000)	0.4571(0.1278)	0.5680(0.0951)	0.5606(0.0922)	0.5305(0.0938)	0.4035(0.1501)	0.5758(0.0980)	0.5713(0.1053)	0.5418(0.1047)
1000	2.9575(0.0000)	0.8418(0.2997)	1.1955(0.2649)	1.1619(0.2816)	1.1276(0.2639)	0.8421(0.2990)	1.4091(0.3692)	1.4605(0.3986)	1.4031(0.4074)
1200	2.9575(0.0000)	1.4333(0.4999)	2.5970(0.5567)	2.5486(0.5875)	2.4933(0.5503)	1.4646(0.4752)	2.6174(0.5832)	2.6535(0.5843)	2.5394(0.5848)
1400	2.9575(0.0000)	1.4115(0.5268)	2.9161(0.5871)	2.8448(0.6508)	2.7171(0.6760)	1.4306(0.5233)	2.8926(0.6528)	2.8699(0.6424)	2.7203(0.6935)
Scenario 4: $\mu(\boldsymbol{\theta}) = \boldsymbol{a}_2^\top \boldsymbol{\theta}$, where $\boldsymbol{a}_2 = (1, 0, 0, 0, 0, 0, 0)^\top$.									
300	2.9575(0.0000)	0.2406(0.0653)	0.5495(0.0879)	0.5699(0.0899)	0.5381(0.0880)	0.2364(0.0604)	0.5519(0.0891)	0.5787(0.0902)	0.5501(0.0904)
500	2.9575(0.0000)	0.2537(0.0803)	0.5658(0.0958)	0.5810(0.0956)	0.5510(0.0957)	0.2478(0.0724)	0.5681(0.0948)	0.5944(0.0991)	0.5591(0.1013)
700	2.9575(0.0000)	0.4413(0.1358)	0.5625(0.0951)	0.5566(0.0890)	0.5276(0.0963)	0.4033(0.1449)	0.5690(0.0946)	0.5661(0.1091)	0.5467(0.1019)
1000	2.9575(0.0000)	0.8580(0.3091)	1.2014(0.2772)	1.1619(0.2836)	1.1165(0.2705)	0.8321(0.3049)	1.4113(0.3903)	1.4902(0.4222)	1.4257(0.4157)
1200	2.9575(0.0000)	1.3994(0.4702)	2.5795(0.5572)	2.5592(0.5506)	2.4553(0.5887)	1.3999(0.4942)	2.6495(0.5758)	2.6137(0.6132)	2.5106(0.6070)
1400	2.9575(0.0000)	1.4621(0.5355)	2.8568(0.6526)	2.8548(0.6800)	2.6763(0.6721)	1.4211(0.5204)	2.8715(0.6670)	2.8450(0.7178)	2.7421(0.6898)
Scenario 5: $\mu(\boldsymbol{\theta}) = \exp(\boldsymbol{a}_2^\top \boldsymbol{\theta})$.									
300	2.9575(0.0000)	0.2411(0.0746)	0.5542(0.0878)	0.5797(0.0943)	0.5495(0.0955)	0.2350(0.0659)	0.5592(0.0877)	0.5866(0.0915)	0.5546(0.0973)
500	2.9575(0.0000)	0.2499(0.0823)	0.5614(0.0924)	0.5741(0.1020)	0.5469(0.0950)	0.2443(0.0742)	0.5674(0.0887)	0.5876(0.0989)	0.5567(0.0961)
700	2.9575(0.0000)	0.4475(0.1356)	0.5576(0.0950)	0.5490(0.0949)	0.5285(0.0929)	0.4079(0.1459)	0.5648(0.0962)	0.5629(0.1007)	0.5379(0.0979)
1000	2.9575(0.0000)	0.8680(0.3122)	1.1965(0.2762)	1.1791(0.2817)	1.1264(0.2743)	0.8533(0.3130)	1.4196(0.3873)	1.5234(0.4058)	1.4366(0.4321)
1200	2.9575(0.0000)	1.4675(0.4769)	2.6110(0.5440)	2.5452(0.5530)	2.4781(0.5624)	1.4437(0.5233)	2.6322(0.5781)	2.6028(0.6396)	2.5527(0.6152)
1400	2.9575(0.0000)	1.5056(0.4894)	2.9180(0.6081)	2.8301(0.6859)	2.7185(0.6769)	1.4324(0.5268)	2.9173(0.6611)	2.9071(0.6738)	2.7440(0.7082)

Table 3: The average computation time per repetition (measured in seconds) and their standard deviations of Poisson regression for different subsampling methods under Case 3 settings when r varies.

r Values	Full	uniR	L-optR	A-optR	TOSR	uniP	L-optP	A-optP	TOSP
Scenario 1: $\mu(\boldsymbol{\theta}) = \boldsymbol{a}_0^\top \boldsymbol{\theta}$, where $\boldsymbol{a}_0 = I_7$ and I_p is a p-dimensional identical matrix.									
300	3.5824(0.0000)	0.0323(0.0140)	0.1993(0.0217)	0.2080(0.0218)	0.2287(0.0221)	0.0337(0.0091)	0.1990(0.0203)	0.2104(0.0251)	0.2315(0.0228)
500	3.5824(0.0000)	0.0322(0.0117)	0.1956(0.0191)	0.2062(0.0194)	0.2258(0.0206)	0.0334(0.0074)	0.1952(0.0182)	0.2066(0.0244)	0.2278(0.0223)
700	3.5824(0.0000)	0.0399(0.0144)	0.1958(0.0197)	0.2057(0.0206)	0.2260(0.0211)	0.0393(0.0107)	0.1954(0.0180)	0.2070(0.0245)	0.2274(0.0220)
1000	3.5824(0.0000)	0.0578(0.0200)	0.2126(0.0201)	0.2224(0.0220)	0.2439(0.0229)	0.0596(0.0188)	0.2221(0.0236)	0.2367(0.0340)	0.2582(0.0292)
1200	3.5824(0.0000)	0.0628(0.0180)	0.2459(0.0250)	0.2550(0.0289)	0.2794(0.0293)	0.0634(0.0158)	0.2455(0.0253)	0.2600(0.0362)	0.2817(0.0302)
1400	3.5824(0.0000)	0.0639(0.0168)	0.2560(0.0253)	0.2651(0.0309)	0.2895(0.0311)	0.0646(0.0146)	0.2538(0.0253)	0.2679(0.0339)	0.2905(0.0316)
Scenario 2: $\mu(\boldsymbol{\theta}) = \boldsymbol{a}_1^\top \boldsymbol{\theta}$, where $\boldsymbol{a}_1 = (1, 0, -1, 0.5, 1.5, -0.5, -1.5)^\top$.									
300	3.5824(0.0000)	0.0305(0.0131)	0.1851(0.0203)	0.1955(0.0210)	0.1597(0.0189)	0.0321(0.0089)	0.1858(0.0190)	0.1970(0.0250)	0.1637(0.0200)
500	3.5824(0.0000)	0.0298(0.0099)	0.1838(0.0186)	0.1930(0.0192)	0.1599(0.0182)	0.0317(0.0080)	0.1847(0.0183)	0.1961(0.0243)	0.1622(0.0188)
700	3.5824(0.0000)	0.0380(0.0129)	0.1834(0.0185)	0.1921(0.0182)	0.1593(0.0172)	0.0375(0.0110)	0.1841(0.0167)	0.1936(0.0227)	0.1611(0.0194)
1000	3.5824(0.0000)	0.0564(0.0197)	0.1989(0.0188)	0.2074(0.0202)	0.1731(0.0217)	0.0558(0.0161)	0.2073(0.0231)	0.2237(0.0317)	0.1876(0.0298)
1200	3.5824(0.0000)	0.0600(0.0184)	0.2301(0.0248)	0.2376(0.0261)	0.2046(0.0335)	0.0604(0.0152)	0.2290(0.0234)	0.2422(0.0312)	0.2101(0.0370)
1400	3.5824(0.0000)	0.0610(0.0168)	0.2409(0.0262)	0.2485(0.0284)	0.2128(0.0343)	0.0619(0.0149)	0.2388(0.0227)	0.2503(0.0309)	0.2162(0.0353)
Scenario 3: $\mu(\boldsymbol{\theta}) = \exp(\boldsymbol{a}_1^\top \boldsymbol{\theta})$.									
300	3.5824(0.0000)	0.0305(0.0118)	0.1862(0.0199)	0.1952(0.0201)	0.1612(0.0190)	0.0326(0.0089)	0.1875(0.0190)	0.1980(0.0240)	0.1633(0.0185)
500	3.5824(0.0000)	0.0302(0.0102)	0.1841(0.0190)	0.1950(0.0200)	0.1617(0.0193)	0.0314(0.0071)	0.1861(0.0189)	0.1969(0.0238)	0.1632(0.0190)
700	3.5824(0.0000)	0.0379(0.0138)	0.1830(0.0178)	0.1922(0.0185)	0.1592(0.0173)	0.0376(0.0106)	0.1834(0.0165)	0.1936(0.0225)	0.1614(0.0179)
1000	3.5824(0.0000)	0.0549(0.0177)	0.1991(0.0187)	0.2064(0.0196)	0.1739(0.0227)	0.0563(0.0158)	0.2069(0.0220)	0.2223(0.0303)	0.1889(0.0325)
1200	3.5824(0.0000)	0.0590(0.0173)	0.2283(0.0229)	0.2358(0.0269)	0.2027(0.0333)	0.0610(0.0179)	0.2296(0.0237)	0.2423(0.0309)	0.2070(0.0336)
1400	3.5824(0.0000)	0.0609(0.0164)	0.2410(0.0252)	0.2491(0.0303)	0.2132(0.0334)	0.0629(0.0152)	0.2406(0.0246)	0.2532(0.0329)	0.2165(0.0353)
Scenario 4: $\mu(\boldsymbol{\theta}) = \boldsymbol{a}_2^\top \boldsymbol{\theta}$, where $\boldsymbol{a}_2 = (1, 0, 0, 0, 0, 0, 0)^\top$.									
300	3.5824(0.0000)	0.0303(0.0117)	0.1842(0.0187)	0.1944(0.0198)	0.1612(0.0181)	0.0323(0.0086)	0.1866(0.0188)	0.1968(0.0242)	0.1644(0.0189)
500	3.5824(0.0000)	0.0301(0.0119)	0.1842(0.0183)	0.1946(0.0210)	0.1612(0.0198)	0.0316(0.0076)	0.1847(0.0186)	0.1950(0.0250)	0.1637(0.0206)
700	3.5824(0.0000)	0.0377(0.0133)	0.1829(0.0175)	0.1929(0.0198)	0.1589(0.0178)	0.0374(0.0106)	0.1839(0.0176)	0.1933(0.0228)	0.1615(0.0181)
1000	3.5824(0.0000)	0.0552(0.0169)	0.1999(0.0191)	0.2074(0.0211)	0.1744(0.0229)	0.0566(0.0179)	0.2084(0.0239)	0.2231(0.0308)	0.1908(0.0353)
1200	3.5824(0.0000)	0.0593(0.0179)	0.2291(0.0230)	0.2374(0.0274)	0.2035(0.0327)	0.0593(0.0139)	0.2288(0.0216)	0.2427(0.0310)	0.2108(0.0400)
1400	3.5824(0.0000)	0.0612(0.0173)	0.2398(0.0234)	0.2487(0.0306)	0.2153(0.0372)	0.0617(0.0152)	0.2383(0.0235)	0.2501(0.0321)	0.2171(0.0380)
Scenario 5: $\mu(\boldsymbol{\theta}) = \exp(\boldsymbol{a}_2^\top \boldsymbol{\theta})$.									
300	3.5824(0.0000)	0.0304(0.0124)	0.1863(0.0211)	0.1960(0.0221)	0.1631(0.0203)	0.0324(0.0085)	0.1874(0.0197)	0.1990(0.0260)	0.1648(0.0201)
500	3.5824(0.0000)	0.0302(0.0109)	0.1854(0.0200)	0.1955(0.0218)	0.1619(0.0198)	0.0318(0.0085)	0.1867(0.0192)	0.1965(0.0233)	0.1646(0.0200)
700	3.5824(0.0000)	0.0375(0.0124)	0.1838(0.0185)	0.1944(0.0195)	0.1610(0.0190)	0.0379(0.0104)	0.1847(0.0175)	0.1953(0.0226)	0.1641(0.0194)
1000	3.5824(0.0000)	0.0553(0.0176)	0.1979(0.0186)	0.2057(0.0200)	0.1744(0.0233)	0.0557(0.0157)	0.2082(0.0229)	0.2220(0.0309)	0.1893(0.0324)
1200	3.5824(0.0000)	0.0599(0.0180)	0.2311(0.0237)	0.2393(0.0277)	0.2110(0.0392)	0.0617(0.0170)	0.2309(0.0237)	0.2459(0.0314)	0.2104(0.0354)
1400	3.5824(0.0000)	0.0608(0.0160)	0.2422(0.0248)	0.2511(0.0301)	0.2213(0.0413)	0.0632(0.0162)	0.2406(0.0238)	0.2543(0.0336)	0.2209(0.0398)

Table 4: The average computation time per repetition (measured in seconds) and their standard deviations of Poisson regression for different subsampling methods under Case 4 settings when r varies.

r Values	Full	uniR	L-optR	A-optR	TOSR	uniP	L-optP	A-optP	TOSP
Scenario 1: $\mu(\boldsymbol{\theta}) = \boldsymbol{a}_0^\top \boldsymbol{\theta}$, where $\boldsymbol{a}_0 = I_7$ and I_p is a p-dimensional identical matrix.									
300	4.5492(0.0000)	0.0322(0.0100)	0.1928(0.0140)	0.2022(0.0152)	0.2208(0.0153)	0.0350(0.0050)	0.1925(0.0141)	0.2029(0.0208)	0.2224(0.0169)
500	4.5492(0.0000)	0.0337(0.0094)	0.1932(0.0137)	0.2031(0.0142)	0.2215(0.0154)	0.0362(0.0052)	0.1938(0.0131)	0.2039(0.0210)	0.2242(0.0154)
700	4.5492(0.0000)	0.0448(0.0108)	0.1959(0.0141)	0.2042(0.0144)	0.2228(0.0152)	0.0447(0.0071)	0.1958(0.0135)	0.2058(0.0207)	0.2244(0.0164)
1000	4.5492(0.0000)	0.0726(0.0122)	0.2204(0.0123)	0.2263(0.0121)	0.2491(0.0127)	0.0728(0.0090)	0.2320(0.0164)	0.2448(0.0231)	0.2643(0.0182)
1200	4.5492(0.0000)	0.0850(0.0116)	0.2761(0.0173)	0.2808(0.0172)	0.3073(0.0174)	0.0856(0.0102)	0.2762(0.0155)	0.2878(0.0255)	0.3088(0.0187)
1400	4.5492(0.0000)	0.0923(0.0120)	0.3097(0.0209)	0.3139(0.0203)	0.3449(0.0218)	0.0931(0.0110)	0.3081(0.0192)	0.3200(0.0292)	0.3459(0.0235)
Scenario 2: $\mu(\boldsymbol{\theta}) = \boldsymbol{a}_1^\top \boldsymbol{\theta}$, where $\boldsymbol{a}_1 = (1, 0, -1, 0.5, 1.5, -0.5, -1.5)^\top$.									
300	4.5492(0.0000)	0.0308(0.0125)	0.1794(0.0130)	0.1883(0.0136)	0.1533(0.0117)	0.0324(0.0050)	0.1818(0.0134)	0.1901(0.0191)	0.1554(0.0116)
500	4.5492(0.0000)	0.0321(0.0110)	0.1821(0.0151)	0.1908(0.0147)	0.1552(0.0122)	0.0333(0.0050)	0.1838(0.0133)	0.1932(0.0198)	0.1569(0.0130)
700	4.5492(0.0000)	0.0435(0.0110)	0.1824(0.0135)	0.1910(0.0137)	0.1547(0.0114)	0.0424(0.0071)	0.1839(0.0131)	0.1924(0.0189)	0.1576(0.0118)
1000	4.5492(0.0000)	0.0670(0.0103)	0.2062(0.0118)	0.2108(0.0114)	0.1746(0.0104)	0.0683(0.0080)	0.2182(0.0163)	0.2313(0.0219)	0.1898(0.0164)
1200	4.5492(0.0000)	0.0793(0.0132)	0.2524(0.0148)	0.2565(0.0142)	0.2141(0.0134)	0.0811(0.0092)	0.2522(0.0138)	0.2632(0.0210)	0.2194(0.0155)
1400	4.5492(0.0000)	0.0876(0.0107)	0.2819(0.0175)	0.2871(0.0172)	0.2401(0.0168)	0.0878(0.0097)	0.2815(0.0167)	0.2918(0.0234)	0.2445(0.0172)
Scenario 3: $\mu(\boldsymbol{\theta}) = \exp(\boldsymbol{a}_1^\top \boldsymbol{\theta})$.									
300	4.5492(0.0000)	0.0302(0.0071)	0.1801(0.0142)	0.1892(0.0148)	0.1536(0.0123)	0.0325(0.0050)	0.1817(0.0140)	0.1908(0.0195)	0.1558(0.0123)
500	4.5492(0.0000)	0.0319(0.0099)	0.1819(0.0138)	0.1911(0.0139)	0.1557(0.0121)	0.0336(0.0053)	0.1839(0.0135)	0.1930(0.0194)	0.1573(0.0125)
700	4.5492(0.0000)	0.0444(0.0103)	0.1845(0.0156)	0.1924(0.0156)	0.1573(0.0129)	0.0430(0.0072)	0.1860(0.0154)	0.1948(0.0204)	0.1593(0.0129)
1000	4.5492(0.0000)	0.0677(0.0110)	0.2068(0.0129)	0.2112(0.0123)	0.1750(0.0108)	0.0691(0.0084)	0.2185(0.0159)	0.2313(0.0220)	0.1897(0.0165)
1200	4.5492(0.0000)	0.0789(0.0115)	0.2533(0.0151)	0.2578(0.0139)	0.2150(0.0132)	0.0816(0.0087)	0.2545(0.0133)	0.2650(0.0206)	0.2208(0.0149)
1400	4.5492(0.0000)	0.0878(0.0118)	0.2830(0.0179)	0.2874(0.0178)	0.2423(0.0161)	0.0878(0.0092)	0.2833(0.0165)	0.2928(0.0239)	0.2451(0.0169)
Scenario 4: $\mu(\boldsymbol{\theta}) = \boldsymbol{a}_2^\top \boldsymbol{\theta}$, where $\boldsymbol{a}_2 = (1, 0, 0, 0, 0, 0, 0)^\top$.									
300	4.5492(0.0000)	0.0304(0.0100)	0.1808(0.0145)	0.1893(0.0151)	0.1528(0.0123)	0.0324(0.0049)	0.1824(0.0144)	0.1906(0.0203)	0.1557(0.0135)
500	4.5492(0.0000)	0.0319(0.0099)	0.1810(0.0155)	0.1905(0.0157)	0.1541(0.0128)	0.0333(0.0050)	0.1833(0.0150)	0.1918(0.0209)	0.1567(0.0131)
700	4.5492(0.0000)	0.0440(0.0106)	0.1839(0.0140)	0.1918(0.0136)	0.1556(0.0112)	0.0430(0.0071)	0.1852(0.0140)	0.1932(0.0198)	0.1582(0.0124)
1000	4.5492(0.0000)	0.0674(0.0108)	0.2075(0.0136)	0.2118(0.0122)	0.1755(0.0109)	0.0683(0.0085)	0.2184(0.0167)	0.2324(0.0219)	0.1910(0.0163)
1200	4.5492(0.0000)	0.0813(0.0110)	0.2597(0.0152)	0.2639(0.0157)	0.2196(0.0140)	0.0832(0.0090)	0.2602(0.0152)	0.2711(0.0233)	0.2246(0.0158)
1400	4.5492(0.0000)	0.0878(0.0122)	0.2820(0.0182)	0.2884(0.0175)	0.2413(0.0157)	0.0879(0.0103)	0.2825(0.0163)	0.2931(0.0242)	0.2448(0.0162)
Scenario 5: $\mu(\boldsymbol{\theta}) = \exp(\boldsymbol{a}_2^\top \boldsymbol{\theta})$.									
300	4.5492(0.0000)	0.0308(0.0096)	0.1823(0.0158)	0.1912(0.0158)	0.1550(0.0132)	0.0331(0.0051)	0.1833(0.0143)	0.1929(0.0208)	0.1573(0.0134)
500	4.5492(0.0000)	0.0319(0.0086)	0.1827(0.0156)	0.1919(0.0167)	0.1560(0.0132)	0.0339(0.0050)	0.1857(0.0156)	0.1940(0.0213)	0.1579(0.0138)
700	4.5492(0.0000)	0.0436(0.0096)	0.1829(0.0137)	0.1910(0.0146)	0.1551(0.0110)	0.0426(0.0071)	0.1842(0.0135)	0.1921(0.0194)	0.1576(0.0121)
1000	4.5492(0.0000)	0.0676(0.0122)	0.2074(0.0130)	0.2124(0.0130)	0.1758(0.0106)	0.0689(0.0086)	0.2196(0.0170)	0.2320(0.0236)	0.1916(0.0174)
1200	4.5492(0.0000)	0.0787(0.0120)	0.2535(0.0144)	0.2570(0.0130)	0.2147(0.0129)	0.0817(0.0087)	0.2539(0.0137)	0.2651(0.0222)	0.2201(0.0148)
1400	4.5492(0.0000)	0.0890(0.0124)	0.2876(0.0192)	0.2923(0.0175)	0.2470(0.0158)	0.0888(0.0101)	0.2871(0.0170)	0.2981(0.0247)	0.2477(0.0168)