Proof for "A comparative study on sampling with replacement vs Poisson sampling in optimal subsampling"

A.1 Proofs

In this section, we prove all the theorems in the paper.

A.1.1 Proof for Theorem 1

To prove Theorem 1, we first establish Lemma 1 and Lemma 2 in the following. Recall that

$$M_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} m(Z_i, \boldsymbol{\theta}).$$

For the sampling with replacement estimator in (2), let

$$M_s^*(\boldsymbol{\theta}) = \frac{1}{s}\sum_{i=1}^{s} \frac{m(Z_i^*, \boldsymbol{\theta})}{n\pi_i^*}.$$

Lemma 1. *Under Assumptions 3 and 5, if* $\|\tilde{\boldsymbol{\theta}}_R - \hat{\boldsymbol{\theta}}\| = o_P(1)$*, then conditional on* $\mathcal{D}_n$*,*

$$B_s - \ddot{M}_n(\hat{\boldsymbol{\theta}}) = o_P(1),$$

where

$$\ddot{M}_n(\hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{i=1}^{n} \ddot{m}(Z_i, \hat{\boldsymbol{\theta}}),$$
$$B_s = \int_0^1 \frac{1}{s}\sum_{i=1}^{s} \frac{\ddot{m}\{Z_i^*, \hat{\boldsymbol{\theta}} + \lambda(\tilde{\boldsymbol{\theta}}_R - \hat{\boldsymbol{\theta}})\}}{n\pi_i^*}\, \mathrm{d}\lambda.$$

In Lemma 1, the notation $o_P(1)$ means convergence to 0 in probability. Here the probability is conditional probability. From Xiong and Li (2008); Cheng and Huang (2010), a sequence converges to 0 in conditional probability is equivalent to the fact that it converges to 0 in unconditional probability. Thus we use $o_P(1)$ to indicate convergence to 0 either in unconditional or conditional probability.

Proof. Firstly, note that

$$\mathbb{E}\left(\frac{1}{s}\sum_{i=1}^{s} \frac{\psi(Z_i^*)}{n\pi_i^*}\Big|\mathcal{D}_n\right) = \frac{1}{n}\sum_{i=1}^{n} \psi(Z_i) = \mathbb{E}\psi(Z) + o_P(1), \quad \text{and}$$
$$\mathbb{V}\left(\frac{1}{s}\sum_{i=1}^{s} \frac{\psi(Z_i^*)}{n\pi_i^*}\Big|\mathcal{D}_n\right) = \frac{1}{s}\sum_{i=1}^{n} \frac{\psi^2(Z_i)}{n^2\pi_i} \le \max_{i=1,\dots,n}\left(\frac{1}{n\pi_i}\right)\frac{1}{sn}\sum_{i=1}^{n} \psi^2(Z_i) = O_P(s^{-1}).$$

Thus,

$$\frac{1}{s}\sum_{i=1}^{s} \frac{\psi(Z_i^*)}{n\pi_i^*} = O_{P|\mathcal{D}_n}(1).$$

For every $k, l = 1, 2, ..., d$, from Lipschitz continuity, for $\lambda \in (0, 1)$, we have

$$\begin{aligned}
&\left|\frac{1}{s}\sum_{i=1}^{s}\frac{\ddot{m}_{k,l}\{Z_i^*,\hat{\boldsymbol{\theta}}+\lambda(\tilde{\boldsymbol{\theta}}_R-\hat{\boldsymbol{\theta}})\}}{n\pi_i^*}-\frac{1}{s}\sum_{i=1}^{s}\frac{\ddot{m}_{k,l}(Z_i^*,\hat{\boldsymbol{\theta}})}{n\pi_i^*}\right| \\
&=\lambda\|\tilde{\boldsymbol{\theta}}_R-\hat{\boldsymbol{\theta}}\|\frac{1}{s}\sum_{i=1}^{s}\frac{\psi(Z_i^*)}{n\pi_i^*}=o_P(1),
\end{aligned} \tag{A.1}$$

and for any fixed $\boldsymbol{\theta}$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\ddot{m}_{k,l}^2(Z_i,\hat{\boldsymbol{\theta}})\le\frac{2}{n}\sum_{i=1}^{n}\ddot{m}_{k,l}^2(Z_i,\boldsymbol{\theta})+\frac{2\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|^2}{n}\sum_{i=1}^{n}\psi^2(Z_i)=O_P(1). \tag{A.2}$$

In addition, according to (A.2),

$$\begin{aligned}
\mathbb{E}\left\{\frac{1}{s}\sum_{i=1}^{s}\frac{\ddot{m}_{k,l}(Z_i^*,\hat{\boldsymbol{\theta}})}{n\pi_i^*}\Big|\mathcal{D}_n\right\}&=\frac{1}{n}\sum_{i=1}^{n}\ddot{m}_{k,l}(Z_i,\hat{\boldsymbol{\theta}}),\\
\mathbb{V}\left\{\frac{1}{s}\sum_{i=1}^{s}\frac{\ddot{m}_{k,l}(Z_i^*,\hat{\boldsymbol{\theta}})}{n\pi_i^*}\Big|\mathcal{D}_n\right\}&=\frac{1}{s}\sum_{i=1}^{n}\frac{\ddot{m}_{k,l}^2(Z_i,\hat{\boldsymbol{\theta}})}{n^2\pi_i}\\
&\le\max_{i=1,2,\ldots,n}\left(\frac{1}{n\pi_i}\right)\frac{1}{sn}\sum_{i=1}^{n}\ddot{m}_{k,l}^2(Z_i,\hat{\boldsymbol{\theta}})=O_P(s^{-1}).
\end{aligned}$$

Thus, by Chebyshev's inequality, we have

$$\left|\frac{1}{s}\sum_{i=1}^{s}\frac{\ddot{m}_{k,l}(Z_i^*,\hat{\boldsymbol{\theta}})}{n\pi_i^*}-\frac{1}{n}\sum_{i=1}^{n}\ddot{m}_{k,l}(Z_i,\hat{\boldsymbol{\theta}})\right|=O_{P|\mathcal{D}_n}(s^{-1/2})=o_{P|\mathcal{D}_n}(1). \tag{A.3}$$

Combining (A.1) and (A.3), we have

$$\begin{aligned}
\left|B_s-\frac{1}{n}\sum_{i=1}^{n}\ddot{m}_{k,l}(Z_i,\hat{\boldsymbol{\theta}})\right|&\le\int_0^1\left|\frac{1}{s}\sum_{i=1}^{s}\frac{\ddot{m}_{k,l}\{Z_i^*,\hat{\boldsymbol{\theta}}+\lambda(\tilde{\boldsymbol{\theta}}_R-\hat{\boldsymbol{\theta}})\}}{n\pi_i^*}-\frac{1}{n}\sum_{i=1}^{n}\ddot{m}_{k,l}(Z_i,\hat{\boldsymbol{\theta}})\right|\mathrm{d}\lambda\\
&\le\int_0^1\left[\left|\frac{1}{s}\sum_{i=1}^{s}\frac{\ddot{m}_{k,l}\{Z_i^*,\hat{\boldsymbol{\theta}}+\lambda(\tilde{\boldsymbol{\theta}}_R-\hat{\boldsymbol{\theta}})\}}{n\pi_i^*}-\frac{1}{s}\sum_{i=1}^{s}\frac{\ddot{m}_{k,l}(Z_i^*,\hat{\boldsymbol{\theta}})}{n\pi_i^*}\right|\right]\mathrm{d}\lambda\\
&\quad+\left|\frac{1}{s}\sum_{i=1}^{s}\frac{\ddot{m}_{k,l}(Z_i^*,\hat{\boldsymbol{\theta}})}{n\pi_i^*}-\frac{1}{n}\sum_{i=1}^{n}\ddot{m}_{k,l}(Z_i,\hat{\boldsymbol{\theta}})\right|\\
&=o_{P|\mathcal{D}_n}(1).
\end{aligned}$$

□

Lemma 2. *Under Assumptions 4-5, given $\mathcal{D}_n$,*

$$\sqrt{s}\{\Lambda_R(\hat{\boldsymbol{\theta}})\}^{-1/2}\dot{M}_s^*(\hat{\boldsymbol{\theta}})\to\mathbb{N}\left(\mathbf{0},\boldsymbol{I}\right), \tag{A.4}$$

in conditional distribution.

Proof. Note that

$$\sqrt{s}\dot{M}_s^*(\hat{\boldsymbol{\theta}})=\frac{1}{\sqrt{s}}\sum_{i=1}^{s}\frac{\dot{m}(Z_i^*,\hat{\boldsymbol{\theta}})}{n\pi_i^*}\equiv\frac{1}{\sqrt{s}}\sum_{i=1}^{s}\boldsymbol{\eta}_i \tag{A.5}$$

Given $\mathcal{D}_n,\boldsymbol{\eta}_1,\ldots,\boldsymbol{\eta}_s$ are i.i.d, with

$$\mathbb{E}(\boldsymbol{\eta}|\mathcal{D}_n)=\frac{1}{n}\sum_{i=1}^{n}\dot{m}(Z_i,\hat{\boldsymbol{\theta}})=\mathbf{0},\ \text{and} \tag{A.6}$$

$$\mathbb{V}(\boldsymbol{\eta}_i|\mathcal{D}_n) = \Lambda_R(\hat{\boldsymbol{\theta}}) = \frac{1}{n^2}\sum_{i=1}^{n}\frac{\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\dot{m}^{\mathrm{T}}(Z_i,\hat{\boldsymbol{\theta}})}{\pi_i}$$
$$\le \max_{i=1,\ldots,n}\left(\frac{1}{n\pi_i}\right)\frac{1}{n}\sum_{i=1}^{n}\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\dot{m}^{\mathrm{T}}(Z_i,\hat{\boldsymbol{\theta}}) = O_P(1). \tag{A.7}$$

Meanwhile, for every $\varepsilon > 0$ and some $\delta \in (0, 2]$,

$$\frac{1}{s}\sum_{i=1}^{s}\mathbb{E}\left\{\|\boldsymbol{\eta}_i\|^2 I(\|\boldsymbol{\eta}_i\|> s^{1/2}\varepsilon)|\mathcal{D}_n\right\} \le \frac{1}{s^{1+\delta/2}\varepsilon^{\delta}}\sum_{i=1}^{s}\mathbb{E}\left\{\|\boldsymbol{\eta}_i\|^{2+\delta} I(\|\boldsymbol{\eta}_i\|> s^{1/2}\varepsilon)|\mathcal{D}_n\right\}$$
$$\le \frac{1}{s^{1+\delta/2}\varepsilon^{\delta}}\sum_{i=1}^{s}\mathbb{E}\left(\|\boldsymbol{\eta}_i\|^{2+\delta}|\mathcal{D}_n\right) \le \frac{1}{s^{\delta/2}n^{2+\delta}\varepsilon^{\delta}}\sum_{i=1}^{n}\frac{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^{2+\delta}}{\pi_i^{1+\delta}}$$
$$= \max_{i=1,\ldots,n}\left(\frac{1}{n\pi_i}\right)^{1+\delta}\frac{1}{ns^{\delta/2}\varepsilon^{\delta}}\sum_{i=1}^{n}\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^{2+\delta} = O_P(s^{-\delta/2}).$$

This shows that Lindeberg's condition is satisfied in probability. From (A.5), (A.6) and (A.7), by the Lindeberg-Feller central limit theorem (Proposition 2.27 of van der Vaart (1998)), conditionally on $\mathcal{D}_n$, (A.4) follows. □

Proof of Theorem 1. Based on Lemma 1 and Lemma 2, now we are ready to prove Theorem 1. By direct calculation, we have that for any $\boldsymbol{\theta}$,

$$\mathbb{E}\left(M_s^*(\boldsymbol{\theta})|\mathcal{D}_n\right) = M_n(\boldsymbol{\theta}).$$

By Chebyshev's inequality, for any $\varepsilon > 0$,

$$\mathbb{P}\left\{|M_s^*(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta})| \ge \varepsilon|\mathcal{D}_n\right\} \le \frac{\mathbb{V}\{M_s^*(\boldsymbol{\theta})|\mathcal{D}_n\}}{\varepsilon^2} = \frac{1}{\varepsilon^2 sn^2}\sum_{i=1}^{n}\frac{m^2(Z_i,\boldsymbol{\theta})}{\pi_i}$$
$$\le \frac{1}{\varepsilon^2 s}\max_{i=1,\ldots,n}\left(\frac{1}{n\pi_i}\right)\frac{1}{n}\sum_{i=1}^{n}m^2(Z_i,\boldsymbol{\theta}) = O_P\left(s^{-1}\right).$$

Thus, for every $\boldsymbol{\theta}$,

$$M_s^*(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}) = o_{P|\mathcal{D}_n}(1). \tag{A.8}$$

Note that under Assumptions 1, 2, the parameter space is compact and $\hat{\boldsymbol{\theta}}$ is the unique global maximum of the continuous concave function $M_n(\boldsymbol{\theta})$. Thus from Theorem 5.9 and its remark of van der Vaart (1998), conditionally on $\mathcal{D}_n$,

$$\|\tilde{\boldsymbol{\theta}}_R - \hat{\boldsymbol{\theta}}\| = o_{P|\mathcal{D}_n}(1) = o_P(1). \tag{A.9}$$

The consistency ensures that $\tilde{\boldsymbol{\theta}}_R$ is close to $\hat{\boldsymbol{\theta}}$ as long as s is large. By Taylor expansion,

$$0 = \dot{M}_s^*(\tilde{\boldsymbol{\theta}}_R) = \dot{M}_s^*(\hat{\boldsymbol{\theta}}) + B_s(\tilde{\boldsymbol{\theta}}_R - \hat{\boldsymbol{\theta}}), \tag{A.10}$$

where

$$B_s = \int_0^1 \frac{1}{s}\sum_{i=1}^{s}\frac{\ddot{m}\{Z_i^*, \hat{\boldsymbol{\theta}} + \lambda(\tilde{\boldsymbol{\theta}}_R - \hat{\boldsymbol{\theta}})\}}{n\pi_i^*}\,\mathrm{d}\lambda.$$

From (A.10) and Lemma 1,

$$0 = \dot{M}_s^*(\tilde{\boldsymbol{\theta}}_R) = \dot{M}_s^*(\hat{\boldsymbol{\theta}}) + \{\ddot{M}_n(\hat{\boldsymbol{\theta}}) + o_P(1)\}(\tilde{\boldsymbol{\theta}}_R - \hat{\boldsymbol{\theta}}), \tag{A.11}$$

which shows that

$$\tilde{\boldsymbol{\theta}}_R - \hat{\boldsymbol{\theta}} = -\{\ddot{M}_n(\hat{\boldsymbol{\theta}}) + o_P(1)\}^{-1}\dot{M}_s^*(\hat{\boldsymbol{\theta}})$$

$$= -\frac{1}{\sqrt{s}}\{\ddot{M}_n(\hat{\boldsymbol{\theta}}) + o_P(1)\}^{-1}\{\Lambda_R(\hat{\boldsymbol{\theta}})\}^{1/2}\sqrt{s}\{\Lambda_R(\hat{\boldsymbol{\theta}})\}^{-1/2}\dot{M}_s^*(\hat{\boldsymbol{\theta}}). \tag{A.12}$$

By Lemma 2 and Slutsky's theorem, we can obtain that, given full data $\mathcal{D}_n$ in probability,

$$\sqrt{s}\{V_R(\hat{\boldsymbol{\theta}})\}^{-1/2}(\tilde{\boldsymbol{\theta}}_R - \hat{\boldsymbol{\theta}}) \to \mathbb{N}\left(\mathbf{0}, \boldsymbol{I}\right), \tag{A.13}$$

in conditional distribution. This means that for any $\boldsymbol{x}$, as s and n get large,

$$\mathbb{P}\left[\sqrt{s}\{V_R(\hat{\boldsymbol{\theta}})\}^{-1/2}(\tilde{\boldsymbol{\theta}}_R - \hat{\boldsymbol{\theta}}) < \boldsymbol{x}\middle|\mathcal{D}_n\right] \to \Phi(\boldsymbol{x})$$

in probability, where $\Phi(\boldsymbol{x})$ is the cumulative distribution function of the standard multivariate normal distribution. A conditional probability is a bounded random variable, for which convergence in probability to a constant implies convergence in the mean. Therefore, the convergence in (A.13) also holds in unconditional distribution. □

A.1.2 Proof for Theorem 2

To prove Theorem 2, we first establish the following Lemmas 3 and 4. Let $\nu_i = 1$ if the i-th data point is selected in the subsample and $\nu_i = 0$ otherwise. The estimator in (2) is the same as the maximizer of

$$M_P^*(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{s^*}\frac{m(Z_i^*, \boldsymbol{\theta})}{s\pi_i^*} = \frac{1}{n}\sum_{i=1}^{n}\frac{\nu_i m(Z_i, \boldsymbol{\theta})}{s\pi_i}.$$

Here, we use s to replace s^* in (2) for convenience, and the resulting estimator is identical to $\tilde{\boldsymbol{\theta}}_P$.

Lemma 3. *If Assumptions 4-5 hold, then, given $\mathcal{D}_n$,*

$$\sqrt{s}\{\Lambda_P(\hat{\boldsymbol{\theta}})\}^{-1/2}\dot{M}_P^*(\hat{\boldsymbol{\theta}}) \to \mathbb{N}(\mathbf{0}, \boldsymbol{I}),$$

in conditional distribution, where

$$\Lambda_P(\hat{\boldsymbol{\theta}}) = \frac{1}{n^2}\sum_{i=1}^{n}\frac{(1 - s\pi_i)\dot{m}(Z_i, \hat{\boldsymbol{\theta}})\dot{m}^{\mathrm{T}}(Z_i, \hat{\boldsymbol{\theta}})}{\pi_i}.$$

Proof. Write

$$\sqrt{s}\dot{M}_P^*(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n}\frac{\nu_i\dot{m}(Z_i, \hat{\boldsymbol{\theta}})}{n\sqrt{s}\pi_i} \equiv \sum_{i=1}^{n}\boldsymbol{\eta}_{Pi}.$$

By direct calculation and according to the definition of $\hat{\boldsymbol{\theta}}$,

$$\mathbb{E}\left(\sum_{i=1}^{n}\boldsymbol{\eta}_{Pi}\middle|\mathcal{D}_n\right) = \sqrt{s}\sum_{i=1}^{n}\frac{\dot{m}(Z_i, \hat{\boldsymbol{\theta}})}{n} = \mathbf{0},$$

and

$$\begin{aligned}
\mathbb{V}\left(\sum_{i=1}^{n}\boldsymbol{\eta}_{Pi}\middle|\mathcal{D}_n\right) &= \frac{1}{n^2}\sum_{i=1}^{n}\frac{\mathbb{V}(\nu_i|\mathcal{D}_n)\dot{m}(Z_i, \hat{\boldsymbol{\theta}})\dot{m}^{\mathrm{T}}(Z_i, \hat{\boldsymbol{\theta}})}{r\pi_i^2} \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\frac{(1 - s\pi_i)\dot{m}(Z_i, \hat{\boldsymbol{\theta}})\dot{m}^{\mathrm{T}}(Z_i, \hat{\boldsymbol{\theta}})}{\pi_i} = \Lambda_P(\hat{\boldsymbol{\theta}}) \\
&\leq \left(\max_{i=1,2,\ldots,n}\frac{1}{n\pi_i}\right)\frac{1}{n}\sum_{i=1}^{n}\dot{m}(Z_i, \hat{\boldsymbol{\theta}})\dot{m}^{\mathrm{T}}(Z_i, \hat{\boldsymbol{\theta}}) = O_P(1).
\end{aligned}$$

Next, we check Lindeberg's condition in conditional distribution. Note that for $\rho \in (0, 2]$ and any $\varepsilon > 0$,

$$\sum_{i=1}^{n} \mathbb{E}\left\{ \|\boldsymbol{\eta}_{Pi}\| I(\|\boldsymbol{\eta}_{Pi}\| > \varepsilon) \Big| \mathcal{D}_n \right\} \le \frac{1}{\varepsilon^\rho} \sum_{i=1}^{n} \mathbb{E}\left\{ \|\boldsymbol{\eta}_{Pi}\|^{2+\rho} I(\|\boldsymbol{\eta}_{Pi}\| > \varepsilon) \Big| \mathcal{D}_n \right\}$$
$$\le \frac{1}{\varepsilon^\rho} \sum_{i=1}^{n} \mathbb{E}\left(\|\boldsymbol{\eta}_{Pi}\|^{2+\rho} \Big| \mathcal{D}_n \right) = \frac{1}{\varepsilon^\rho} \mathbb{E}\left\{ \sum_{i=1}^{n} \frac{\nu_i^{2+\rho} \|\dot{m}(Z_i, \hat{\boldsymbol{\theta}})\|^{2+\rho}}{n^{2+\rho} s^{1+\rho/2} \pi_i^{2+\rho}} \Bigg| \mathcal{D}_n \right\}$$
$$= \frac{1}{\varepsilon^\rho} \sum_{i=1}^{n} \frac{\|\dot{m}(Z_i, \hat{\boldsymbol{\theta}})\|^{2+\rho}}{n^{2+\rho} s^{\rho/2} \pi_i^{1+\rho}}$$
$$\le \max_{i=1,2,\ldots,n} \left(\frac{1}{n\pi_i} \right)^{1+\rho} \frac{1}{s^{\rho/2} \varepsilon^\rho n} \sum_{i=1}^{n} \|\dot{m}(Z_i, \hat{\boldsymbol{\theta}})\|^{2+\rho} = O_P(s^{-\rho/2}) = o_P(1).$$

According to the Lindeberg-Feller Central Limit Theorem (van der Vaart, 1998, cf.), given $\mathcal{D}_n$,

$$\sqrt{s}\{\Lambda_P(\hat{\boldsymbol{\theta}})\}^{-1/2} \dot{M}_P^*(\hat{\boldsymbol{\theta}}) \to \mathbb{N}(\mathbf{0}, \boldsymbol{I}),$$

in conditional distribution. □

Lemma 4. *Under Assumptions 3 and 5, for any* $\boldsymbol{u}_s = o_P(1)$*, conditional on* $\mathcal{D}_n$*,*

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\nu_i \ddot{m}(Z_i, \hat{\boldsymbol{\theta}} + \boldsymbol{u}_s)}{s\pi_i} - \frac{1}{n} \sum_{i=1}^{n} \ddot{m}(Z_i, \hat{\boldsymbol{\theta}}) = o_P(1).$$

Proof. First, note that

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\nu_i \psi(Z_i)}{s\pi_i} = O_{P|\mathcal{D}_n}(1), \tag{A.14}$$

by Chebyshev's inequality and the fact that

$$\mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} \frac{\nu_i \psi(Z_i)}{s\pi_i} \Bigg| \mathcal{D}_n \right) = \frac{1}{n} \sum_{i=1}^{n} \frac{\psi(Z_i)\mathbb{E}(\nu_i|\mathcal{D}_n)}{s\pi_i} = \frac{1}{n} \sum_{i=1}^{n} \psi(Z_i) = \mathbb{E}\{\psi(Z_i)\} + o_P(1),$$
$$\mathbb{V}\left(\frac{1}{n} \sum_{i=1}^{n} \frac{\nu_i \psi(Z_i)}{s\pi_i} \Bigg| \mathcal{D}_n \right) = \frac{1}{n^2} \sum_{i=1}^{n} \frac{\psi^2(Z_i)\mathbb{V}(\nu_i|\mathcal{D}_n)}{s^2\pi_i^2} \le \frac{1}{n^2} \sum_{i=1}^{n} \frac{\psi^2(Z_i)\mathbb{E}(\nu_i^2)}{s^2\pi_i^2}$$
$$= \frac{1}{n^2} \sum_{i=1}^{n} \frac{\psi^2(Z_i)}{s\pi_i} \le \frac{1}{sn} \sum_{i=1}^{n} \psi^2(Z_i) \max_{i=1,2,\ldots,n} \frac{1}{n\pi_i} = O_P(s^{-1}).$$

Thus, for every $k, l = 1, 2, ..., d$, from Assumption 3, we have

$$\left| \frac{1}{n} \sum_{i=1}^{n} \frac{\nu_i \ddot{m}_{k,l}(Z_i, \hat{\boldsymbol{\theta}} + \boldsymbol{u}_s)}{s\pi_i} - \frac{1}{n} \sum_{i=1}^{n} \frac{\nu_i \ddot{m}_{k,l}(Z_i, \hat{\boldsymbol{\theta}})}{s\pi_i} \right| \le \frac{\|\boldsymbol{u}_s\|}{n} \sum_{i=1}^{n} \frac{\nu_i \psi(Z_i)}{s\pi_i} = o_P(1).$$

which shows that

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\nu_i \ddot{m}(Z_i, \hat{\boldsymbol{\theta}} + \boldsymbol{u}_s)}{s\pi_i} - \frac{1}{n} \sum_{i=1}^{n} \frac{\nu_i \ddot{m}(Z_i, \hat{\boldsymbol{\theta}})}{s\pi_i} = o_P(1). \tag{A.15}$$

According to (A.2), for every $k, l = 1, 2, ..., d$

$$\mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} \frac{\nu_i \ddot{m}_{k,l}(Z_i, \hat{\boldsymbol{\theta}})}{s\pi_i} \Bigg| \mathcal{D}_n \right) = \frac{1}{n} \sum_{i=1}^{n} \ddot{m}_{k,l}(Z_i, \hat{\boldsymbol{\theta}}),$$
$$\mathbb{V}\left(\frac{1}{n} \sum_{i=1}^{n} \frac{\nu_i \ddot{m}_{k,l}(Z_i, \hat{\boldsymbol{\theta}})}{s\pi_i} \Bigg| \mathcal{D}_n \right) = \frac{1}{sn^2} \sum_{i=1}^{n} \frac{(1 - s\pi_i)\ddot{m}_{k,l}^2(Z_i, \hat{\boldsymbol{\theta}})}{\pi_i} \le \frac{1}{sn^2} \sum_{i=1}^{n} \frac{\ddot{m}_{k,l}^2(Z_i, \hat{\boldsymbol{\theta}})}{\pi_i}$$

$$\leq \max_{i=1,2,\ldots,n}\left(\frac{1}{n\pi_i}\right)\frac{1}{sn}\sum_{i=1}^{n}\ddot{m}_{k,l}^2(Z_i,\hat{\boldsymbol{\theta}}) = O_P(s^{-1}).$$

Thus, Chebyshev's inequality tells us that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\nu_i\ddot{m}(Z_i,\hat{\boldsymbol{\theta}})}{s\pi_i} - \frac{1}{n}\sum_{i=1}^{n}\ddot{m}(Z_i,\hat{\boldsymbol{\theta}}) = O_{P|\mathcal{D}_n}(s^{-1/2}) = o_P(1). \tag{A.16}$$

Therefore, combining (A.15) and (A.16), we have

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\nu_i\ddot{m}(Z_i,\hat{\boldsymbol{\theta}}+\boldsymbol{u}_s)}{s\pi_i} - \frac{1}{n}\sum_{i=1}^{n}\ddot{m}(Z_i,\hat{\boldsymbol{\theta}}) = o_P(1).$$

□

Proof of Theorem 2. Denote

$$\gamma_P(\boldsymbol{u}) = sM_P^*(\hat{\boldsymbol{\theta}}+\boldsymbol{u}/\sqrt{s}) - sM_P^*(\hat{\boldsymbol{\theta}}).$$

Under Assumption 2, $\sqrt{s}(\tilde{\boldsymbol{\theta}}_P-\hat{\boldsymbol{\theta}})$ is the unique maximizer of $\gamma_P(\boldsymbol{u})$ as $\tilde{\boldsymbol{\theta}}_P$ is the unique maximizer of $M_P^*(\hat{\boldsymbol{\theta}}+\boldsymbol{u}/\sqrt{s})$. By Taylor's expansion,

$$\gamma_P(\boldsymbol{u}) = \sqrt{s}\boldsymbol{u}^T\dot{M}_P^*(\hat{\boldsymbol{\theta}}) + \frac{1}{2}\boldsymbol{u}^T\ddot{M}_P^*(\hat{\boldsymbol{\theta}}+\acute{\boldsymbol{u}}/\sqrt{s})\boldsymbol{u}$$

where $\acute{\boldsymbol{u}}$ lies between $\mathbf{0}$ and $\boldsymbol{u}/\sqrt{s}$. From Lemma 3 $\sqrt{s}\dot{M}_P^*(\hat{\boldsymbol{\theta}})$ is stochastically bounded in conditional probability given $\mathcal{D}_n$. From Lemma 4, conditional on $\mathcal{D}_n$, $\ddot{M}_P^*(\hat{\boldsymbol{\theta}}+\acute{\boldsymbol{u}}/\sqrt{s}) - \ddot{M}_n(\hat{\boldsymbol{\theta}}) = o_P(1)$ and $\ddot{M}_n(\hat{\boldsymbol{\theta}})$ converges to a positive-definite matrix.

Thus from the Basic Corollary in page 2 of Hjort and Pollard Hjort and Pollard (2011), the maximizer of $s\gamma_P(\boldsymbol{u})$, $\sqrt{s}(\tilde{\boldsymbol{\theta}}_P-\hat{\boldsymbol{\theta}})$, satisfies that

$$\sqrt{s}(\tilde{\boldsymbol{\theta}}_P-\hat{\boldsymbol{\theta}}) = \ddot{M}_n^{-1}(\hat{\boldsymbol{\theta}})\sqrt{s}\dot{M}_P^*(\hat{\boldsymbol{\theta}}) + o_P(1), \tag{A.17}$$

which implies that

$$\sqrt{s}\{V_P(\hat{\boldsymbol{\theta}})\}^{-1/2}(\tilde{\boldsymbol{\theta}}_P-\hat{\boldsymbol{\theta}}) \to \mathbb{N}(\mathbf{0},\boldsymbol{I}), \tag{A.18}$$

in conditional distribution. Thus, by argumentation similar to that used in the proof of Theorem 1, we know that the convergence also holds in unconditional distribution, and this finishes the proof. □

A.1.3 Proof of Theorem 3

Proof of Theorem 3. For the result in (8),

$$\mathrm{tr}\{\Lambda_R(\hat{\boldsymbol{\theta}})\} = \frac{1}{n^2}\sum_{i=1}^{n}\frac{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2}{\pi_i} = \frac{1}{n^2}\sum_{i=1}^{n}\pi_i\sum_{i=1}^{n}\frac{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2}{\pi_i} \geq \frac{1}{n^2}\left\{\sum_{i=1}^{n}\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|\right\}^2.$$

Here, the last step is from the Cauchy-Schwarz inequality and the equality holds if and only if $\pi_i \propto \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|$.

□

A.1.4 Proof of Theorem 4

Proof. Note that

$$\mathrm{tr}\{\Lambda_P(\hat{\boldsymbol{\theta}})\} = \mathrm{tr}\left\{\frac{1}{n^2}\sum_{i=1}^{n}\frac{(1-s\pi_i)\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\dot{m}^{\mathrm{T}}(Z_i,\hat{\boldsymbol{\theta}})}{\pi_i}\right\}$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{n}\frac{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2}{\pi_i} - s\sum_{i=1}^{n}\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2\right].$$

Thus, minimizing $\mathrm{tr}\{\Lambda_P(\hat{\boldsymbol{\theta}})\}$ is equal to minimizing $\sum_{i=1}^{n}\frac{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2}{\pi_i}$. For $i = 1, ..., n$, let $t_i = \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|$ and let $t_{(i)}$ denote the order statistics of $\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|$, i.e., $t_{(i)} = \|\dot{m}(Z,\hat{\boldsymbol{\theta}})\|_{(i)}$. The optimization problem of minimizing $\mathrm{tr}\{\Lambda_P(\hat{\boldsymbol{\theta}})\}$ subject to the constrains on π_i can be presented as minimizing

$$T(\pi_1,\pi_2,...,\pi_n) = \sum_{i=1}^{n}\frac{t_{(i)}^2}{\pi_i}, \tag{A.19}$$

$$\text{subject to } \sum_{i=1}^{n}\pi_i = 1 \quad \text{and} \quad 0 \le \pi_i \le \frac{1}{s}, i = 1, 2, ..., n.$$

Defining slack variables $\omega_1^2, \omega_2^2, ..., \omega_n^2$, to use Lagrangian multiplier method, we can construct

$$H(\pi_1,...,\pi_n,\tau,\mu_1,...,\mu_n,\omega_1,...,\omega_n) = \sum_{i=1}^{n}\frac{t_{(i)}^2}{\pi_i} + \tau\left(\sum_{i=1}^{n}\pi_i - 1\right) + \sum_{i=1}^{n}\mu_i\left(\pi_i + \omega_i^2 - \frac{1}{s}\right).$$

By taking the derivatives, the Karush–Kuhn–Tucker (KKT) conditions (Nocedal and Wright, 1999) are

$$\frac{\partial H}{\partial \pi_i} = -\frac{t_{(i)}^2}{\pi_i^2} + \tau + \mu_i = 0, \qquad i = 1, 2, ..., n. \tag{A.20}$$

$$\frac{\partial H}{\partial \tau} = \sum_{i=1}^{n}\pi_i - 1 = 0, \tag{A.21}$$

$$\frac{\partial H}{\partial \mu_i} = \pi_i + \omega_i^2 = \frac{1}{s}, \qquad i = 1, 2, ..., n. \tag{A.22}$$

$$\frac{\partial H}{\partial \omega_i} = 2\mu_i\omega_i = 0, \qquad i = 1, 2, ..., n. \tag{A.23}$$

$$\mu_i \ge 0, \qquad i = 1, 2, ..., n. \tag{A.24}$$

From (A.20), we have

$$\pi_i = \frac{t_{(i)}}{\sqrt{\tau + \mu_i}}, \quad i = 1, 2, ..., n. \tag{A.25}$$

Combining it with (A.22), we have

$$\frac{t_{(i)}}{\sqrt{\tau + \mu_i}} + \omega_i^2 = \frac{1}{s}, \quad i = 1, 2, ..., n. \tag{A.26}$$

According to (A.23), at least one of μ_i and ω_i must be 0. From (A.25) and (A.26),

$$\text{if } t_{(i)} < \frac{\sqrt{\tau}}{s}, \qquad \mu = 0 \text{ and } \pi_i = \frac{t_{(i)}}{\sqrt{\tau}} < \frac{1}{s}; \tag{A.27}$$

$$\text{if } t_{(i)} \ge \frac{\sqrt{\tau}}{s}, \qquad \omega_i = 0 \text{ and } \pi_i = \frac{t_{(i)}}{\sqrt{\tau + \mu_i}} = \frac{1}{s}. \tag{A.28}$$

Thus, letting g be the number of cases that $t_{(i)} \ge \frac{\sqrt{\tau}}{s}$, from (A.21) and the fact that $t_{(i)}$ is non-decreasing in i,

$$1 = \sum_{i=1}^{n}\pi_i = \sum_{i=1}^{n-g}\frac{t_{(i)}}{\sqrt{\tau}} + \sum_{i=n-g+1}^{n}\frac{1}{s} = \frac{\sum_{i=1}^{n-g}t_{(i)}}{\sqrt{\tau}} + \frac{g}{s}, \tag{A.29}$$

which shows that

$$\sqrt{\tau} = \frac{s}{s-g}\sum_{i=1}^{n-g}t_{(i)}. \tag{A.30}$$

Combining(A.27), (A.28), and (A.30),

$$\pi_i = \begin{cases} \dfrac{t_{(i)}(s-g)}{s\sum_{i=1}^{n-g} t_{(i)}}, & \text{for } i = 1, 2, ..., n-g; \quad (A.31) \\ \dfrac{1}{s}, & \text{for } i = n-g+1, ..., n. \quad (A.32) \end{cases}$$

From (A.30),

$$H = \frac{\sum_{i=1}^{n-g} t_{(i)}}{s-g} = \frac{\sqrt{\tau}}{s}, \tag{A.33}$$

Thus, from (A.27) and (A.28), we know $t_{(i)} < H$ for $i = 1, 2, ..., n-g$, and $t_{(i)} \geq H$, for $i = n-g+1, ..., n$. Therefore

$$\sum_{i=1}^{n}(t_{(i)} \wedge H) = \sum_{i=1}^{n-g} t_{(i)} + \sum_{i=n-g+1}^{n} H = sH \tag{A.34}$$

Thus, from (A.31), for $i = 1, 2, ..., n-g$,

$$\pi_i = \frac{t_{(i)}}{sH} = \frac{t_{(i)} \wedge H}{\sum_{i=1}^{n}(t_{(i)} \wedge H)}; \tag{A.35}$$

from (A.32), for $i = n-g+1, ..., n$,

$$\pi_i = \frac{H}{sH} = \frac{t_{(i)} \wedge H}{\sum_{i=1}^{n}(t_{(i)} \wedge H)}. \tag{A.36}$$

For the result under the A-optimality, define $t_{(i)} = \|\ddot{M}_n^{-1}(\hat{\boldsymbol{\theta}})\dot{m}(Z, \hat{\boldsymbol{\theta}})\|_{(i)}$ and the proof is the same as the used for the L-optimality. □

A.1.5 Proof of Theorem 5

We prove Theorem 5 from establishing the following lemmas.

Lemma 5. *Under Assumption 3, if* $\|\tilde{\boldsymbol{\theta}}_R^{\alpha} - \hat{\boldsymbol{\theta}}\| = o_P(1)$*, then conditional on* $\mathcal{D}_n$ *and* $\tilde{\boldsymbol{\theta}}_R^{0*}$*,*

$$B_s^{\tilde{\boldsymbol{\theta}}_R^{0*}} - \ddot{M}_n(\hat{\boldsymbol{\theta}}) = o_P(1), \tag{A.37}$$

where

$$B_s^{\tilde{\boldsymbol{\theta}}_R^{0*}} = \int_0^1 \frac{1}{s}\sum_{i=1}^{s} \frac{\ddot{m}\left\{Z_i^*, \hat{\boldsymbol{\theta}} + \lambda(\tilde{\boldsymbol{\theta}}_R^{\alpha} - \hat{\boldsymbol{\theta}})\right\}}{n\tilde{\pi}_{R\alpha i}^{\text{opt}*}} \, \mathrm{d}\lambda.$$

Proof. For every $k, l = 1, 2, ..., d$, from Lipschitz continuity, we have

$$\begin{aligned} &\left|\frac{1}{s}\sum_{i=1}^{s} \frac{\ddot{m}_{k,l}\{Z_i^*, \hat{\boldsymbol{\theta}} + \lambda(\tilde{\boldsymbol{\theta}}_R^{\alpha} - \hat{\boldsymbol{\theta}})\}}{n\tilde{\pi}_{R\alpha i}^{\text{opt}*}} - \frac{1}{s}\sum_{i=1}^{s} \frac{\ddot{m}_{k,l}(Z_i^*, \hat{\boldsymbol{\theta}})}{n\tilde{\pi}_{R\alpha i}^{\text{opt}*}}\right| \\ \leq &\frac{1}{s}\sum_{i=1}^{s} \frac{\psi(Z_i^*)\|\lambda(\tilde{\boldsymbol{\theta}}_R^{\alpha} - \hat{\boldsymbol{\theta}})\|}{n\tilde{\pi}_{R\alpha i}^{\text{opt}*}} \leq \lambda\|\tilde{\boldsymbol{\theta}}_R^{\alpha} - \hat{\boldsymbol{\theta}}\|\frac{1}{s}\sum_{i=1}^{s}\frac{\psi(Z_i^*)}{\alpha} = \|\tilde{\boldsymbol{\theta}}_R^{\alpha} - \hat{\boldsymbol{\theta}}\|O_P(1) = o_P(1). \end{aligned} \tag{A.38}$$

According to (A.2), we have

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

$$\mathbb{V}\left(\frac{1}{s}\sum_{i=1}^{s}\frac{\ddot{m}_{k,l}(Z_i^*,\hat{\boldsymbol{\theta}})}{n\tilde{\pi}_{R\alpha i}^{\mathrm{opt}*}}\middle|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_R^{0*}\right)=\frac{1}{s}\sum_{i=1}^{n}\frac{\ddot{m}_{k,l}^2(Z_i,\hat{\boldsymbol{\theta}})}{n^2\pi_{R\alpha i}^{\mathrm{opt}}}\leq\frac{1}{\alpha sn}\sum_{i=1}^{n}\ddot{m}_{k,l}^2(Z_i,\hat{\boldsymbol{\theta}})=O_P(s^{-1}).$$

Thus, by Chebyshev's inequality, similar to (A.3), we have

$$\left\|\frac{1}{s}\sum_{i=1}^{s}\frac{\ddot{m}(Z_i^*,\hat{\boldsymbol{\theta}})}{n\tilde{\pi}_{R\alpha i}^{\mathrm{opt}*}}-\frac{1}{n}\sum_{i=1}^{n}\ddot{m}(Z_i,\hat{\boldsymbol{\theta}})\right\|=o_{P|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_R^{0*}}(1). \tag{A.39}$$

Combining (A.38) and (A.39), we have

$$\begin{aligned}\left\|B_s-\frac{1}{n}\sum_{i=1}^{n}\ddot{m}(Z_i,\hat{\boldsymbol{\theta}})\right\|&\leq\int_0^1\left\|\frac{1}{s}\sum_{i=1}^{s}\frac{\ddot{m}\{Z_i^*,\hat{\boldsymbol{\theta}}+\lambda(\tilde{\boldsymbol{\theta}}_R^{\alpha}-\hat{\boldsymbol{\theta}})\}}{n\tilde{\pi}_{R\alpha i}^{\mathrm{opt}*}}-\frac{1}{n}\sum_{i=1}^{n}\ddot{m}(Z_i,\hat{\boldsymbol{\theta}})\right\|\mathrm{d}\lambda\\&\leq\int_0^1\left[\left\|\frac{1}{s}\sum_{i=1}^{s}\frac{\ddot{m}\{Z_i^*,\hat{\boldsymbol{\theta}}+\lambda(\tilde{\boldsymbol{\theta}}_R^{\alpha}-\hat{\boldsymbol{\theta}})\}}{n\tilde{\pi}_{R\alpha i}^{\mathrm{opt}*}}-\frac{1}{s}\sum_{i=1}^{s}\frac{\ddot{m}(Z_i^*,\hat{\boldsymbol{\theta}})}{n\tilde{\pi}_{R\alpha i}^{\mathrm{opt}*}}\right\|\right.\\&\qquad\left.+\left\|\frac{1}{s}\sum_{i=1}^{s}\frac{\ddot{m}(Z_i^*,\hat{\boldsymbol{\theta}})}{n\tilde{\pi}_{R\alpha i}^{\mathrm{opt}*}}-\frac{1}{n}\sum_{i=1}^{n}\ddot{m}(Z_i,\hat{\boldsymbol{\theta}})\right\|\right]\mathrm{d}\lambda=o_P(1),\end{aligned}$$

which finishes the proof. □

Lemma 6. *If Assumption 4 hold, then given $\mathcal{D}_n$ and $\tilde{\boldsymbol{\theta}}_R^{0*}$,*

$$\sqrt{s}\{\Lambda_R^{\alpha}(\tilde{\boldsymbol{\theta}}_R^{0*})\}^{-1/2}\dot{M}_{R\alpha}^*(\hat{\boldsymbol{\theta}})\to\mathbb{N}\left(\mathbf{0},\boldsymbol{I}\right), \tag{A.40}$$

in conditional distribution, where

$$\dot{M}_{R\alpha}^*(\hat{\boldsymbol{\theta}})=\frac{1}{ns}\sum_{i=1}^{s}\frac{\dot{m}(Z_i^*,\hat{\boldsymbol{\theta}})}{\tilde{\pi}_{R\alpha i}^{\mathrm{opt}*}},\quad \textit{and}\quad \Lambda_R^{\alpha}(\tilde{\boldsymbol{\theta}}_R^{0*})=\frac{1}{n^2}\sum_{i=1}^{n}\frac{\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\dot{m}^{\mathrm{T}}(Z_i,\hat{\boldsymbol{\theta}})}{\tilde{\pi}_{R\alpha i}^{\mathrm{opt}}}.$$

Proof. Note that

$$\sqrt{s}\dot{M}_{R\alpha}^*(\hat{\boldsymbol{\theta}})=\frac{1}{\sqrt{s}}\sum_{i=1}^{s}\frac{\dot{m}(Z_i^*,\hat{\boldsymbol{\theta}})}{n\tilde{\pi}_{R\alpha i}^{\mathrm{opt}*}}\equiv\frac{1}{\sqrt{s}}\sum_{i=1}^{s}\boldsymbol{\eta}_i^{\tilde{\boldsymbol{\theta}}_R^{0*}}. \tag{A.41}$$

Given $\mathcal{D}_n$ and $\tilde{\boldsymbol{\theta}}_R^{0*}$, $\boldsymbol{\eta}_1^{\tilde{\boldsymbol{\theta}}_R^{0*}},...,\boldsymbol{\eta}_s^{\tilde{\boldsymbol{\theta}}_R^{0*}}$ are i.i.d, with

$$\mathbb{E}(\boldsymbol{\eta}_i^{\tilde{\boldsymbol{\theta}}_R^{0*}}|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_R^{0*})=\frac{1}{n}\sum_{i=1}^{n}\dot{m}(Z_i,\hat{\boldsymbol{\theta}})=\mathbf{0},\quad\text{and} \tag{A.42}$$

$$\mathbb{V}(\boldsymbol{\eta}_i^{\tilde{\boldsymbol{\theta}}_R^{0*}}|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_R^{0*})=\mathbb{E}\left\{\frac{\dot{m}(Z_i^*,\hat{\boldsymbol{\theta}})\dot{m}^{\mathrm{T}}(Z_i^*,\hat{\boldsymbol{\theta}})}{n^2(\tilde{\pi}_{R\alpha i}^{\mathrm{opt}*})^2}\middle|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_R^{0*}\right\} \tag{A.43}$$

$$=\frac{1}{n^2}\sum_{i=1}^{n}\frac{\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\dot{m}^{\mathrm{T}}(Z_i,\hat{\boldsymbol{\theta}})}{\tilde{\pi}_{R\alpha i}^{\mathrm{opt}*}}=\Lambda_R^{\alpha}(\tilde{\boldsymbol{\theta}}_R^{0*}). \tag{A.44}$$

Meanwhile, for every $\varepsilon>0$ and some $\delta\in(0,2]$,

$$\begin{aligned}&\frac{1}{s}\sum_{i=1}^{s}\mathbb{E}\left\{\|\boldsymbol{\eta}_i^{\tilde{\boldsymbol{\theta}}_R^{0*}}\|^2I(\|\boldsymbol{\eta}_i^{\tilde{\boldsymbol{\theta}}_R^{0*}}\|>s^{1/2}\varepsilon)\middle|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_R^{0*}\right\}\\&\leq\frac{1}{s^{1+\delta/2}\varepsilon^{\delta}}\sum_{i=1}^{s}\mathbb{E}\left\{\|\boldsymbol{\eta}_i^{\tilde{\boldsymbol{\theta}}_R^{0*}}\|^{2+\delta}I(\|\boldsymbol{\eta}_i^{\tilde{\boldsymbol{\theta}}_R^{0*}}\|>s^{1/2}\varepsilon)\middle|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_R^{0*}\right\}\\&\leq\frac{1}{s^{1+\delta/2}\varepsilon^{\delta}}\sum_{i=1}^{s}\mathbb{E}\left(\|\boldsymbol{\eta}_i^{\tilde{\boldsymbol{\theta}}_R^{0*}}\|^{2+\delta}\middle|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_R^{0*}\right)\end{aligned}$$

$$\leq \frac{1}{s^{\delta/2} n^{2+\delta} \varepsilon^{\delta}} \sum_{i=1}^{n} \frac{\|\dot{m}(Z_i, \hat{\boldsymbol{\theta}})\|^{2+\delta}}{(\tilde{\pi}_{R\alpha i}^{\text{opt}*})^{1+\delta}}$$
$$\leq \frac{1}{s^{\delta/2} \alpha^{1+\delta} \varepsilon^{\delta}} \frac{1}{n} \sum_{i=1}^{n} \|\dot{m}(Z_i, \hat{\boldsymbol{\theta}})\|^{2+\delta} = O_P(s^{-\delta/2}) = o_P(1).$$

where the second last equality is from Assumption 4. This show that Lindeberg's condition is satisfied in probability. From (A.41), (A.42) and (A.44), by the Lindeberg-Feller central limit theorem (Proposition 2.27 of van der Vaart (1998)), conditional on $\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_R^{0*}$, we obtain (A.40). □

Proof of Theorem 5. By direct calculation, we have

$$\mathbb{E}\left\{ M_{R\alpha}^*(\boldsymbol{\theta}) \middle| \mathcal{D}_n, \tilde{\boldsymbol{\theta}}_R^{0*} \right\} = M_n(\boldsymbol{\theta}),$$
$$\mathbb{V}\left\{ M_{R\alpha}^*(\boldsymbol{\theta}) \middle| \mathcal{D}_n, \tilde{\boldsymbol{\theta}}_R^{0*} \right\} \leq \frac{1}{sn^2} \sum_{i=1}^{n} \frac{m^2(Z_i, \boldsymbol{\theta})}{\tilde{\pi}_{R\alpha i}^{\text{opt}*}} \leq \frac{1}{sn} \sum_{i=1}^{n} \frac{m^2(Z_i, \boldsymbol{\theta})}{\alpha} = O_P(s^{-1}).$$

By Chebyshev's inequality, for each $\boldsymbol{\theta}$, we have

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

Under Assumptions 1 and 2, the parameter space is compact and $\hat{\boldsymbol{\theta}}$ is the unique global maximum of the continuous concave function $M_n(\boldsymbol{\theta})$. Thus from Theorem 5.9 and its remark of van der Vaart (1998), conditionally on $\mathcal{D}_n$ and $\tilde{\boldsymbol{\theta}}_R^{0*}$,

$$\|\tilde{\boldsymbol{\theta}}_R^{\alpha} - \hat{\boldsymbol{\theta}}\| = o_P(1).$$

By Taylor expansion

$$0 = \dot{M}_{R\alpha}^*(\tilde{\boldsymbol{\theta}}_R^{\alpha}) = \dot{M}_{R\alpha}^*(\hat{\boldsymbol{\theta}}) + B_s^{\tilde{\boldsymbol{\theta}}_R^{0*}}(\tilde{\boldsymbol{\theta}}_R^{\alpha} - \hat{\boldsymbol{\theta}}),$$

so

$$\tilde{\boldsymbol{\theta}}_R^{\alpha} - \hat{\boldsymbol{\theta}} = -\left(B_s^{\tilde{\boldsymbol{\theta}}_R^{0*}}\right)^{-1} \dot{M}_{R\alpha}^*(\hat{\boldsymbol{\theta}}) = -\frac{1}{\sqrt{s}} \left(B_s^{\tilde{\boldsymbol{\theta}}_R^{0*}}\right)^{-1} \{\Lambda_R^{\alpha}(\tilde{\boldsymbol{\theta}}_R^{0*})\}^{1/2} \sqrt{s} \{\Lambda_R^{\alpha}(\tilde{\boldsymbol{\theta}}_R^{0*})\}^{-1/2} \dot{M}_{R\alpha}^*(\hat{\boldsymbol{\theta}}).$$

Therefore, from Lemma 5 and Lemma 6, conditional on $\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_R^{0*}$, by Slutsky's theorem

$$\sqrt{s}\{\Lambda_R^{\alpha}(\tilde{\boldsymbol{\theta}}_R^{0*})\}^{-1/2} \ddot{M}_n(\hat{\boldsymbol{\theta}})(\tilde{\boldsymbol{\theta}}_R^{\alpha} - \hat{\boldsymbol{\theta}}) \to \mathbb{N}(\mathbf{0}, \boldsymbol{I}), \tag{A.45}$$

in conditional distribution.

Next, we check the distance bewtween $\Lambda_R^{\alpha}(\tilde{\boldsymbol{\theta}}_R^{0*})$ and $\Lambda_R^{\alpha}(\hat{\boldsymbol{\theta}})$.

$$\|\Lambda_R^{\alpha}(\tilde{\boldsymbol{\theta}}_R^{0*}) - \Lambda_R^{\alpha}(\hat{\boldsymbol{\theta}})\|$$
$$= \left\| \frac{1}{n^2} \sum_{i=1}^{n} \frac{\dot{m}(Z_i, \hat{\boldsymbol{\theta}}) \dot{m}^{\mathrm{T}}(Z_i, \hat{\boldsymbol{\theta}})}{(1-\alpha)\pi_{Ri}^{\text{opt}}(\tilde{\boldsymbol{\theta}}_R^{0*}) + \alpha \frac{1}{n}} - \frac{1}{n^2} \sum_{i=1}^{n} \frac{\dot{m}(Z_i, \hat{\boldsymbol{\theta}}) \dot{m}^{\mathrm{T}}(Z_i, \hat{\boldsymbol{\theta}})}{(1-\alpha)\pi_{Ri}^{\text{opt}}(\hat{\boldsymbol{\theta}}) + \alpha \frac{1}{n}} \right\|$$
$$\leq \frac{1}{n^2} \sum_{i=1}^{n} \|\dot{m}(Z_i, \hat{\boldsymbol{\theta}})\|^2 \left| \frac{1}{(1-\alpha)\pi_{Ri}^{\text{opt}}(\tilde{\boldsymbol{\theta}}_R^{0*}) + \alpha \frac{1}{n}} - \frac{1}{(1-\alpha)\pi_{Ri}^{\text{opt}}(\hat{\boldsymbol{\theta}}) + \alpha \frac{1}{n}} \right|$$
$$< \frac{1}{\alpha^2} \sum_{i=1}^{n} \|\dot{m}(Z_i, \hat{\boldsymbol{\theta}})\|^2 \left| \pi_{Ri}^{\text{opt}}(\tilde{\boldsymbol{\theta}}_R^{0*}) - \pi_{Ri}^{\text{opt}}(\hat{\boldsymbol{\theta}}) \right|$$
$$\leq \frac{1}{\alpha^2} \sum_{i=1}^{n} \|\dot{m}(Z_i, \hat{\boldsymbol{\theta}})\|^2 \left\{ \frac{\left| \|\dot{m}(Z_i, \tilde{\boldsymbol{\theta}}_R^{0*})\| - \|\dot{m}(Z_i, \hat{\boldsymbol{\theta}})\| \right|}{\sum_{j=1}^{n} \|\dot{m}(Z_j, \tilde{\boldsymbol{\theta}}_R^{0*})\|} \right.$$

$$+\|\dot{m}(Z_i,\hat{\boldsymbol\theta})\|\frac{\sum_{j=1}^n\left|\|\dot{m}(Z_j,\tilde{\boldsymbol\theta}_R^{0*})\|-\|\dot{m}(Z_j,\hat{\boldsymbol\theta})\|\right|}{\sum_{j=1}^n\|\dot{m}(Z_j,\hat{\boldsymbol\theta})\|\sum_{j=1}^n\|\dot{m}(Z_j,\tilde{\boldsymbol\theta}_R^{0*})\|}\Bigg\}$$

$$\equiv\frac{1}{\alpha^2}\sum_{i=1}^n\|\dot{m}(Z_i,\hat{\boldsymbol\theta})\|^2(\Delta_{1i}+\Delta_{2i})\,. \tag{A.46}$$

Under Assumption 3, for any $j=1,2,...,n$

$$\begin{aligned}&\left|\|\dot{m}(Z_j,\hat{\boldsymbol\theta})\|-\|\dot{m}(Z_j,\tilde{\boldsymbol\theta}_R^{0*})\|\right|\le\|\dot{m}(Z_j,\hat{\boldsymbol\theta})-\dot{m}(Z_j,\tilde{\boldsymbol\theta}_R^{0*})\|\\&\le\sqrt{\sum_{k=1}^d\{\dot{m}_k(Z_j,\hat{\boldsymbol\theta})-\dot{m}_k(Z_j,\tilde{\boldsymbol\theta}_R^{0*})\}^2}\le\sum_{k=1}^d\left|\dot{m}_k(Z_j,\hat{\boldsymbol\theta})-\dot{m}_k(Z_j,\tilde{\boldsymbol\theta}_R^{0*})\right|\\&\le\sum_{k=1}^d\left|\ddot{m}_k^T(Z_j,\xi_k)(\hat{\boldsymbol\theta}-\tilde{\boldsymbol\theta}_R^{0*})\right|\le\|\hat{\boldsymbol\theta}-\tilde{\boldsymbol\theta}_R^{0*}\|\sum_{k=1}^d\|\ddot{m}_k(Z_j,\xi_k)\|\equiv\|\hat{\boldsymbol\theta}-\tilde{\boldsymbol\theta}_R^{0*}\|h(Z_j),\end{aligned} \tag{A.47}$$

where $\dot{m}_k(Z_j,\hat{\boldsymbol\theta})$ is the kth element of $\dot{m}(Z_j,\hat{\boldsymbol\theta})$, $\ddot{m}_k(Z_j,\hat{\boldsymbol\theta})$ is the kth column of $\ddot{m}(Z_j,\hat{\boldsymbol\theta})$, and all ξ_k are between $\hat{\boldsymbol\theta}$ and $\tilde{\boldsymbol\theta}_R^{0*}$. Thus,

$$\Delta_{1i}\le\frac{\|\hat{\boldsymbol\theta}-\tilde{\boldsymbol\theta}_R^{0*}\|h(Z_i)}{\sum_{j=1}^n\|\dot{m}(Z_j,\tilde{\boldsymbol\theta}_R^{0*})\|}, \tag{A.48}$$

and

$$\Delta_{2i}\le\frac{\|\dot{m}(Z_i,\hat{\boldsymbol\theta})\|\|\hat{\boldsymbol\theta}-\tilde{\boldsymbol\theta}_R^{0*}\|\sum_{j=1}^n h(Z_j)}{\sum_{j=1}^n\|\dot{m}(Z_j,\hat{\boldsymbol\theta})\|\sum_{j=1}^n\|\dot{m}(Z_j,\tilde{\boldsymbol\theta}_R^{0*})\|} \tag{A.49}$$

From (A.2) and Assumption 3

$$\begin{aligned}&\frac{1}{n}\sum_{j=1}^n h^2(Z_j)\le d\frac{1}{n}\sum_{j=1}^n\sum_{k=1}^d\|\ddot{m}_k(Z_j,\xi_k)\|^2=d\frac{1}{n}\sum_{j=1}^n\sum_{k=1}^d\sum_{l=1}^d\ddot{m}_{k,l}^2(Z_j,\xi_k)\\&\le d\frac{1}{n}\sum_{j=1}^n\sum_{k=1}^d\sum_{l=1}^d\left(2\ddot{m}_{k,l}^2(Z_j,\hat{\boldsymbol\theta})+2\psi^2(Z_j)\|\tilde{\boldsymbol\theta}_R^{0*}-\hat{\boldsymbol\theta}\|^2\right)=O_P(1)\end{aligned} \tag{A.50}$$

which also implies that $\frac{1}{n}\sum_{j=1}^n h(Z_j)=O_P(1)$. Thus,

$$\begin{aligned}\sum_{i=1}^n\|\dot{m}(Z_i,\hat{\boldsymbol\theta})\|\Delta_{1i}&\le\frac{O_P(\|\hat{\boldsymbol\theta}-\tilde{\boldsymbol\theta}_R^{0*}\|)}{n}\sum_{i=1}^n\|\dot{m}(Z_i,\hat{\boldsymbol\theta})\|^2h(Z_i)\\&\le O_P(\|\hat{\boldsymbol\theta}-\tilde{\boldsymbol\theta}_R^{0*}\|)\left\{\frac{1}{n}\sum_{i=1}^n\|\dot{m}(Z_i,\hat{\boldsymbol\theta})\|^4\right\}^{\frac12}\left\{\frac{1}{n}\sum_{i=1}^n h^2(Z_i)\right\}^{\frac12},\end{aligned} \tag{A.51}$$

and

$$\sum_{i=1}^n\|\dot{m}(Z_i,\hat{\boldsymbol\theta})\|\Delta_{2i}=O_P(\|\hat{\boldsymbol\theta}-\tilde{\boldsymbol\theta}_R^{0*}\|)\frac{1}{n}\sum_{i=1}^n\|\dot{m}(Z_i,\hat{\boldsymbol\theta})\|^2 \tag{A.52}$$

Combining (A.46), (A.51), and (A.52), we obtain that for large s_0, s and n,

$$\|\Lambda_R^\alpha(\tilde{\boldsymbol\theta}_R^{0*})-\Lambda_R^\alpha(\hat{\boldsymbol\theta})\|=\|\hat{\boldsymbol\theta}-\tilde{\boldsymbol\theta}_R^{0*}\|O_P(1)=o_P(1).$$

Thus, Slutsky's theorem and (A.45) indicate that given $\mathcal{D}_n$ and $\tilde{\boldsymbol{\theta}}_R^{0*}$, as s_0, s and $n \to \infty$

$$\sqrt{s}\{V_R^\alpha(\hat{\boldsymbol{\theta}})\}^{-1/2}(\tilde{\boldsymbol{\theta}}_R^\alpha - \hat{\boldsymbol{\theta}}) \to \mathbb{N}\left(\mathbf{0}, \boldsymbol{I}\right),$$

in conditional distribution. From similar arguments used in the proof of Theorem 1, we know that the convergence also holds in unconditional distribution, and this finishes the proof.

□

A.1.6 Proof of Theorem 6

To prove Theorem 6, we begin with Lemmas 7, 8 and 9.

Lemma 7. *Under Assumptions 4, conditional on $\mathcal{D}_n$ and $\tilde{\boldsymbol{\theta}}_R^{0*}$, then*

$$\sqrt{s}\{\Lambda_P^\alpha(\tilde{\boldsymbol{\theta}}_P^{0*})\}^{-1/2}\dot{M}_{P\alpha}^*(\hat{\boldsymbol{\theta}}) \to \mathbb{N}(\mathbf{0}, \boldsymbol{I}),$$

in conditional distribution, where

$$\Lambda_P^\alpha(\tilde{\boldsymbol{\theta}}_P^{0*}) = \frac{s}{n^2}\sum_{i=1}^n \frac{\{1-(s\tilde{\pi}_{P\alpha i}^{\mathrm{opt}})\wedge 1\}\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\dot{m}^{\mathrm{T}}(Z_i,\hat{\boldsymbol{\theta}})}{(s\tilde{\pi}_{P\alpha i}^{\mathrm{opt}})\wedge 1}.$$

Proof. Here, for writing convenience, we redefined the notation $\nu_i = I(u_i \le s\pi_{P\alpha i}^{\mathrm{opt}})$ and let

$$\sqrt{s}\dot{M}_{P\alpha}^*(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \frac{\nu_i\sqrt{s}\dot{m}(Z_i,\hat{\boldsymbol{\theta}})}{n\{(s\tilde{\pi}_{P\alpha i}^{\mathrm{opt}})\wedge 1\}} \equiv \sum_{i=1}^n \boldsymbol{\eta}_i^{\tilde{\boldsymbol{\theta}}_P^{0*}}. \tag{A.53}$$

From direct calculation and the definition of $\hat{\boldsymbol{\theta}}$, we have

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

and

$$\begin{aligned}
\mathbb{V}\Big(\sqrt{s}\dot{M}_{P\alpha}^*(\hat{\boldsymbol{\theta}})\Big|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_P^{0*}\Big) &= \frac{s}{n^2}\sum_{i=1}^n \frac{\mathbb{V}(\nu_i|\mathcal{D}_n,\tilde{\boldsymbol{\theta}}_P^{0*})\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\dot{m}^{\mathrm{T}}(Z_i,\hat{\boldsymbol{\theta}})}{\{(s\tilde{\pi}_{P\alpha i}^{\mathrm{opt}})\wedge 1\}^2} \\
&= \frac{s}{n^2}\sum_{i=1}^n \frac{\{1-(s\tilde{\pi}_{P\alpha i}^{\mathrm{opt}})\wedge 1\}\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\dot{m}^{\mathrm{T}}(Z_i,\hat{\boldsymbol{\theta}})}{(s\tilde{\pi}_{P\alpha i}^{\mathrm{opt}})\wedge 1} \\
&\le \frac{1}{\alpha n}\sum_{i=1}^n \dot{m}(Z_i,\hat{\boldsymbol{\theta}})\dot{m}^{\mathrm{T}}(Z_i,\hat{\boldsymbol{\theta}}) = O_P(1).
\end{aligned}$$

Next, we check Lindeberg's condition. For any $epsilon > 0$ and $rho \in (0,2]$,

$$\begin{aligned}
&\mathbb{E}\left\{\sum_{i=1}^n \|\boldsymbol{\eta}_i^{\tilde{\boldsymbol{\theta}}_P^{0*}}\| I(\|\boldsymbol{\eta}_i^{\tilde{\boldsymbol{\theta}}_P^{0*}}\| > \varepsilon)\middle|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_P^{0*}\right\} \\
&\le \frac{1}{\varepsilon^\rho}\sum_{i=1}^n \mathbb{E}\left\{\|\boldsymbol{\eta}_i^{\tilde{\boldsymbol{\theta}}_P^{0*}}\|^{2+\rho} I(\|\boldsymbol{\eta}_i^{\tilde{\boldsymbol{\theta}}_P^{0*}}\| > \varepsilon)\middle|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_P^{0*}\right\} \\
&\le \frac{1}{\varepsilon^\rho}\sum_{i=1}^n \mathbb{E}\left(\|\boldsymbol{\eta}_i^{\tilde{\boldsymbol{\theta}}_P^{0*}}\|^{2+\rho}\middle|\mathcal{D}_n, \tilde{\boldsymbol{\theta}}_P^{0*}\right) = \frac{s^{1+\rho/2}}{\varepsilon^\rho n^{2+\rho}}\sum_{i=1}^n \frac{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^{2+\rho}}{\{(s\tilde{\pi}_{P\alpha i}^{\mathrm{opt}})\wedge 1\}^{1+\rho}} \\
&\le \frac{s^{1+\rho/2}}{\varepsilon^\rho n^{2+\rho}}\sum_{i=1}^n \frac{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^{2+\rho}}{(s\alpha/n)^{1+\rho}} = \frac{1}{\alpha^{1+\rho}\varepsilon^\rho s^{\rho/2}}\frac{1}{n}\sum_{i=1}^n \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^{2+\rho} = O_P(s^{-\rho/2}).
\end{aligned}$$

Thus, from the Lindeberg-Feller Central Limit Theorem (cf. van der Vaart, 1998), Lemma 7 follows. □

Lemma 8. *Under Assumption 3, for any $\boldsymbol{u}_s = o_P(1)$, conditional on $\mathcal{D}_n$ and $\tilde{\boldsymbol{\theta}}_P^{0*}$,*

$$\frac{1}{n}\sum_{i=1}^n \frac{\nu_i \ddot{m}(Z_i, \hat{\boldsymbol{\theta}} + \boldsymbol{u}_s)}{(s\tilde{\pi}_{P\alpha i}^{\text{opt}}) \wedge 1} - \frac{1}{n}\sum_{i=1}^n \ddot{m}(Z_i, \hat{\boldsymbol{\theta}}) = o_P(1).$$

Proof. First, using an approach similar to prove (A.14), we can show that given $\mathcal{D}_n$ and $\tilde{\boldsymbol{\theta}}_P^{0*}$,

$$\frac{1}{n}\sum_{i=1}^n \frac{\nu_i \psi(Z_i)}{(s\tilde{\pi}_{P\alpha i}^{\text{opt}}) \wedge 1} = O_P(1). \tag{A.54}$$

For every $k, l = 1, 2, ..., d$, from Lipschitz continuity, we have

$$\left|\frac{1}{n}\sum_{i=1}^n \frac{\nu_i \ddot{m}_{k,l}(Z_i, \hat{\boldsymbol{\theta}} + \boldsymbol{u}_s)}{(s\tilde{\pi}_{P\alpha i}^{\text{opt}}) \wedge 1} - \frac{1}{n}\sum_{i=1}^n \frac{\nu_i \ddot{m}_{k,l}(Z_i, \hat{\boldsymbol{\theta}})}{(s\tilde{\pi}_{P\alpha i}^{\text{opt}}) \wedge 1}\right| \le \frac{1}{n}\sum_{i=1}^n \frac{\nu_i \psi(Z_i)\|\boldsymbol{u}_s\|}{(s\tilde{\pi}_{P\alpha i}^{\text{opt}}) \wedge 1} = o_P(1). \tag{A.55}$$

For each $k, l = 1, 2, ..., d$, direct calculations show that

$$\mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^n \frac{\nu_i \ddot{m}_{k,l}(Z_i, \hat{\boldsymbol{\theta}})}{(s\tilde{\pi}_{P\alpha i}^{\text{opt}}) \wedge 1}\middle| \mathcal{D}_n, \tilde{\boldsymbol{\theta}}_P^{0*}\right\} = \frac{1}{n}\sum_{i=1}^n \ddot{m}_{k,l}(Z_i, \hat{\boldsymbol{\theta}}),$$

$$\mathbb{V}\left\{\frac{1}{n}\sum_{i=1}^n \frac{\nu_i \ddot{m}_{k,l}(Z_i, \hat{\boldsymbol{\theta}})}{(s\tilde{\pi}_{P\alpha i}^{\text{opt}}) \wedge 1}\middle| \mathcal{D}_n, \tilde{\boldsymbol{\theta}}_P^{0*}\right\} \le \frac{1}{sn^2}\sum_{i=1}^n \frac{\ddot{m}_{k,l}^2(Z_i, \hat{\boldsymbol{\theta}})}{(s\tilde{\pi}_{P\alpha i}^{\text{opt}}) \wedge 1} \le \frac{1}{\alpha sn}\sum_{i=1}^n h^2(Z_i) = O_P(s^{-1}).$$

According to Chebyshev's inequality, we obtain

$$\frac{1}{n}\sum_{i=1}^n \frac{\nu_i \ddot{m}(Z_i, \hat{\boldsymbol{\theta}})}{(s\tilde{\pi}_{P\alpha i}^{\text{opt}}) \wedge 1} - \frac{1}{n}\sum_{i=1}^n \ddot{m}(Z_i, \hat{\boldsymbol{\theta}}) = O_P(s^{-1/2}). \tag{A.56}$$

Therefore, combining (A.55) and (A.56), we have

$$\frac{1}{n}\sum_{i=1}^n \frac{\nu_i \ddot{m}(Z_i, \hat{\boldsymbol{\theta}} + \boldsymbol{u}_s)}{s\pi_{P\alpha i}^{\text{opt}}} - \frac{1}{n}\sum_{i=1}^n \ddot{m}(Z_i, \hat{\boldsymbol{\theta}}) = o_P(1).$$

□

Lemma 9. *Under Assumptions 3 and 4,*

1) if $\varrho_n = s/(bn) \to \varrho \in (0, 1)$, then $H^{0*} - H_{\varrho_n} = o_P(1)$;

2) $\Psi^{0*} - \Psi_{\varrho_n} = o_P(1)$, where

$$\Psi_{\varrho_n} = \frac{1}{n}\sum_{i=1}^n \{\|\dot{m}(Z_i, \hat{\boldsymbol{\theta}})\| \wedge H_{\varrho_n}\}; \tag{A.57}$$

3) if $s/(bn) \to \varrho = 0$, then $\Psi^{0*} - \Psi_\infty = o_P(1)$.

Proof. Note that H^{0*} is the $\lceil s_0^* - s_0^* s/b/n \rceil$-th order statistics of $\|\dot{m}(Z_i^{0*}, \tilde{\boldsymbol{\theta}}_P^{0*})\|$, $i = 1, ..., s_0^*$. For any $\rho > 0$, let $\tilde{H}_\rho$ be the $\lceil n(1-\rho) \rceil$-th order statistics of $\|\dot{m}(Z_1, \tilde{\boldsymbol{\theta}}_P^{0*})\|, ..., \|\dot{m}(Z_n, \tilde{\boldsymbol{\theta}}_P^{0*})\|$. Let $\nu_{(i)}^0 = 1$ if $\|\dot{m}(Z, \tilde{\boldsymbol{\theta}}_P^{0*})\|_{(i)}$ is included in $\|\dot{m}(Z_1^{0*}, \tilde{\boldsymbol{\theta}}_P^{0*})\|, ..., \|\dot{m}(Z_{s_0^*}^{0*}, \tilde{\boldsymbol{\theta}}_P^{0*})\|$. For any $\varrho_+ > \varrho$,

$$\mathbb{P}(H^{0*} \le \tilde{H}_{\varrho_+}) = \mathbb{P}\left(\sum_{i=1}^{\lceil n(1-\varrho_+) \rceil} \nu_{(i)}^0 \ge \lceil s_0^* - s_0^* s/b/n \rceil\right). \tag{A.58}$$

Note that

$$\frac{1}{s_0}\sum_{i=1}^{\lceil n(1-\varrho_+)\rceil}\nu^0_{(i)} = 1-\varrho_+ + o_P(1) \quad \text{and} \quad \frac{\lceil s_0^* - s_0^* s/b/n\rceil}{s_0} = 1-\varrho+o_P(1). \tag{A.59}$$

Thus,

$$\mathbb{P}(H^{0*} \le \tilde{H}_{\varrho_+}) \to 0. \tag{A.60}$$

Similarly, we obtain that for any $\varrho_- < \varrho$,

$$\mathbb{P}(H^{0*} \le \tilde{H}_{\varrho_-}) \to 1. \tag{A.61}$$

Note that $\tilde{H}_{\varrho+}$ is between the $\lceil n(1-\varrho_+)\rceil - s_0^*$-th and the $\lceil n(1-\varrho_+)\rceil$-th order statistics of $\|\dot{m}(Z_i, \tilde{\boldsymbol{\theta}}_P^{0*})\|$'s that are not included in $\|\dot{m}(Z_1^{0*}, \tilde{\boldsymbol{\theta}}_P^{0*})\|, ..., \|\dot{m}(Z_{s_0^*}^{0*}, \tilde{\boldsymbol{\theta}}_P^{0*})\|$. The joint distribution of these $\|\dot{m}(Z_i, \tilde{\boldsymbol{\theta}}_P^{0*})\|$'s are exchangeable, and $s_0^*/n \to 0$ in probability. Therefore, both the $\lceil n(1-\varrho_+)\rceil - s_0^*$-th and the $\lceil n(1-\varrho_+)\rceil$-th order statistics of these $\|\dot{m}(Z_i, \tilde{\boldsymbol{\theta}}_P^{0*})\|$'s converge to the ϱ_+-quantile of the distribution of $\|\dot{m}(Z_i, \boldsymbol{\theta}_0)\|$ in probability (Chanda, 1971), where $\boldsymbol{\theta}_0 = \arg\max_{\boldsymbol{\theta}} \mathbb{E}\{m(Z, \boldsymbol{\theta})\}$. As a result, $\tilde{H}_{\varrho+}$ converge in probability to the ϱ_+-quantile of the distribution of $\|\dot{m}(Z, \boldsymbol{\theta}_0)\|$, say $\zeta_{\varrho+}$. Similarly, $\tilde{H}_{\varrho_-}$ converge in probability to the ϱ_--quantile of the distribution of $\|\dot{m}(Z, \boldsymbol{\theta}_0)\|$, say $\zeta_{\varrho-}$. Thus, (A.60) and (A.61) together imply that for any $\epsilon > 0$,

$$\mathbb{P}(\zeta_{\varrho_+} - \epsilon < H^{0*} < \zeta_{\varrho_-} + \epsilon) \to 1. \tag{A.62}$$

Since the distribution of Z is continuous and so is that of $\|\dot{m}(Z, \boldsymbol{\theta}_0)\|$, we can choose ϱ_+ and ϱ_- close to ϱ enough such that $\zeta_{\varrho_-} - \zeta_\varrho < \epsilon$ and $\zeta_\varrho - \zeta_{\varrho_+} < \epsilon$, which implies that

$$\mathbb{P}(\zeta_\varrho - 2\epsilon < H^{0*} < \zeta_\varrho + 2\epsilon) \to 1, \tag{A.63}$$

for any ϵ. Thus, $H^{0*} = \zeta_\varrho + o_P(1)$. Since $\|\dot{m}(Z_1, \hat{\boldsymbol{\theta}})\|, ..., \|\dot{m}(Z_n, \hat{\boldsymbol{\theta}})\|$ are exchangeable, $H_{\varrho_n} = \zeta_\varrho + o_P(1)$, where H_{ϱ_n} is the $\lceil n(1-\varrho_n)\rceil$-th order statistics of $\|\dot{m}(Z_1, \hat{\boldsymbol{\theta}})\|, ..., \|\dot{m}(Z_n, \hat{\boldsymbol{\theta}})\|$. Therefore, $H^{0*} - H_{\varrho_n} = o_P(1)$.

Now we prove 2) of Lemma 9. If $\varrho = 0$ and $\|\dot{m}(Z, \boldsymbol{\theta})\|$ is bounded, then

$$\Psi^{0*} = \sum_{i=1}^{\lceil s_0^* - s_0^* s/b/n\rceil} \frac{\|\dot{m}(Z^{0*}, \tilde{\boldsymbol{\theta}}_P^{0*})\|_{(i)}}{s_0^*} + \frac{s_0^* - \lceil s_0^* - s_0^* s/b/n\rceil}{s_0^*} H^{0*} = \sum_{i=1}^{s_0^*} \frac{\|\dot{m}(Z_i^{0*}, \tilde{\boldsymbol{\theta}}_P^{0*})\|}{s_0^*} + o_P(1),$$

and similarly,

$$\Psi_{\varrho_n} = \frac{1}{n}\sum_{i=1}^{n}\|\dot{m}(Z_i, \hat{\boldsymbol{\theta}})\| + o_P(1).$$

Thus the proof reduce to prove that

$$\sum_{i=1}^{s_0^*} \frac{\|\dot{m}(Z_i^{0*}, \tilde{\boldsymbol{\theta}}_P^{0*})\|}{s_0^*} = \frac{1}{n}\sum_{i=1}^{n}\|\dot{m}(Z_i, \hat{\boldsymbol{\theta}})\| + o_P(1),$$

which can be proved by Taylor's expansion and Markov's inequality. To prove other cases, let $\nu_i^0 = 1$ if the i-th observation is included in the pilot subsample and $\nu_i^0 = 0$ otherwise; then Ψ^{0*} can be written as

$$\Psi^{0*} = \frac{1}{s_0^*}\sum_{i=1}^{n} \nu_i^0 \{\|\dot{m}(Z_i, \tilde{\boldsymbol{\theta}}_P^{0*})\| \wedge H^{0*}\}.$$

Define

$$\Psi^{0*}_{H_{\varrho_n}} = \frac{1}{s_0^*}\sum_{i=1}^{n} \nu_i^0 \{\|\dot{m}(Z_i, \tilde{\boldsymbol{\theta}}_P^{0*})\| \wedge H_{\varrho_n}\} \quad \text{and} \quad \Psi^{0*}_{\hat{\boldsymbol{\theta}}} = \frac{1}{s_0^*}\sum_{i=1}^{n} \nu_i^0 \{\|\dot{m}(Z_i, \hat{\boldsymbol{\theta}})\| \wedge H_{\varrho_n}\}.$$

If $\varrho > 0$, then

$$
\begin{aligned}
|\Psi^{0*} - \Psi^{0*}_{H_{\varrho n}}| &= \frac{1}{s_0^*}\sum_{i=1}^{n}\nu_i^0\Big|\|\dot m(Z_i,\tilde{\boldsymbol\theta}_P^{0*})\|\wedge H^{0*} - \|\dot m(Z_i,\tilde{\boldsymbol\theta}_P^{0*})\|\wedge H_{\varrho n}\Big| \\
&\le \frac{|H^{0*} - H_{\varrho n}|}{s_0^*}\sum_{i=1}^{n}\nu_i^0 I\Big\{\|\dot m(Z_i,\tilde{\boldsymbol\theta}_P^{0*})\|\ge H^{0*}\wedge H_{\varrho n}\Big\} \le |H^{0*} - H_{\varrho n}| = o_P(1).
\end{aligned}
$$

If $\varrho = 0$ and $\|\dot m(Z,\boldsymbol\theta)\|$ is unbounded, then $H^{0*}\wedge H_{\varrho n} \to \infty$ in probability. Under Assumptions 3 and 4, it can be shown that $\frac{1}{s_0^*}\sum_{i=1}^{n}\nu_i^0\|\dot m(Z_i,\tilde{\boldsymbol\theta}_P^{0*})\|^2 = O_{P|\mathcal{D}_n}(1)$. Thus,

$$
\begin{aligned}
|\Psi^{0*} - \Psi^{0*}_{H_{\varrho n}}| \le{}& \frac{1}{s_0^*}\sum_{i=1}^{n}\nu_i^0\|\dot m(Z_i,\tilde{\boldsymbol\theta}_P^{0*})\| I\Big\{\|\dot m(Z_i,\tilde{\boldsymbol\theta}_P^{0*})\|\ge H^{0*}\wedge H_{\varrho n}\Big\} \\
&+ \frac{H^{0*}}{s_0^*}\sum_{i=1}^{n}\nu_i^0 I\Big\{\|\dot m(Z_i,\tilde{\boldsymbol\theta}_P^{0*})\|\ge H^{0*}\Big\} + \frac{H_{\varrho n}}{s_0^*}\sum_{i=1}^{n}\nu_i^0 I\Big\{\|\dot m(Z_i,\tilde{\boldsymbol\theta}_P^{0*})\|\ge H_{\varrho n}\Big\} \\
\le{}& \left\{\frac{1}{H^{0*}\wedge H_{\varrho n}} + \frac{1}{H^{0*}} + \frac{1}{H_{\varrho n}}\right\}\frac{1}{s_0^*}\sum_{i=1}^{n}\nu_i^0\|\dot m(Z_i,\tilde{\boldsymbol\theta}_P^{0*})\|^2 = o_P(1). \qquad \text{(A.64)}
\end{aligned}
$$

Furthermore, we can show that

$$
\begin{aligned}
|\Psi^{0*}_{H_{\varrho n}} - \Psi^{0*}_{\hat{\boldsymbol\theta}}| &\le \frac{1}{s_0^*}\sum_{i=1}^{n}\nu_i^0\{\|\dot m(Z_i,\tilde{\boldsymbol\theta}_P^{0*})\| - \|\dot m(Z_i,\hat{\boldsymbol\theta})\|\} \\
&\le \frac{\|\hat{\boldsymbol\theta} - \tilde{\boldsymbol\theta}_P^{0*}\|}{s_0^*}\sum_{i=1}^{n}\nu_i^0 h(Z_i) = o_P(1)
\end{aligned}
$$

and

$$
|\Psi^{0*}_{\hat{\boldsymbol\theta}} - \Psi_{\varrho n}| = o_P(1),
$$

where the last two $o_P(1)$ are obtained by mean and variance calculations under the conditional distribution of ν_i^0's. Thus, we have that

$$
|\Psi^{0*} - \Psi_{\varrho n}| = o_P(1). \qquad \text{(A.65)}
$$

With 2) of Lemma 9 proved, in order to prove 3), we only need to show that $\Psi_\infty - \Psi_{\varrho n} = o_P(1)$ if $s/(bn) \to \varrho = 0$. This is true because if $\|\dot m(Z,\boldsymbol\theta)\|$ is bounded, then

$$
\begin{aligned}
|\Psi_\infty - \Psi_{\varrho n}| &\le \frac{1}{n}\sum_{i=1}^{n}\Big|\|\dot m(Z_i,\hat{\boldsymbol\theta})\| - \|\dot m(Z_i,\hat{\boldsymbol\theta})\|\wedge H_{\varrho n}\Big| \\
&\le \frac{n - \lceil n(1-\varrho_n)\rceil}{n}\|\dot m(Z,\hat{\boldsymbol\theta})\|_{(n)} = o_P(1);
\end{aligned}
$$

otherwise,

$$
\begin{aligned}
|\Psi_\infty - \Psi_{\varrho n}| &\le \frac{1}{n}\sum_{i=1}^{n}\Big|\|\dot m(Z_i,\hat{\boldsymbol\theta})\| - \|\dot m(Z_i,\hat{\boldsymbol\theta})\|\wedge H_{\varrho n}\Big| \\
&\le \frac{1}{n}\sum_{i=1}^{n}\|\dot m(Z_i,\hat{\boldsymbol\theta})\| I\Big\{\|\dot m(Z_i,\hat{\boldsymbol\theta})\|\ge H_{\varrho n}\Big\} \\
&\le \frac{1}{nH_{\varrho n}}\sum_{i=1}^{n}\|\dot m(Z_i,\hat{\boldsymbol\theta})\|^2 = o_P(1).
\end{aligned}
$$

□

Proof of Theorem 6. For Algorithm 2, $M^*_{P\alpha}(\boldsymbol{\theta})$ can be written as

$$M^*_{P\alpha}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\nu_i m(Z_i,\boldsymbol{\theta})}{(s\tilde{\pi}^{\rm opt}_{P\alpha i})\wedge 1}.$$

Denote

$$\gamma_{\tilde{\boldsymbol{\theta}}^{0*}_P P}(\boldsymbol{u}) = sM^*_{P\alpha}(\hat{\boldsymbol{\theta}} + \boldsymbol{u}/\sqrt{s}) - sM^*_{P\alpha}(\hat{\boldsymbol{\theta}}).$$

Under Assumption 2, $\sqrt{s}(\tilde{\boldsymbol{\theta}}^{\alpha}_P - \hat{\boldsymbol{\theta}})$ is the unique maximizer of $\gamma_{\tilde{\boldsymbol{\theta}}^{0*}_P P}(\boldsymbol{u})$. By Taylor's expansion,

$$\gamma_{\tilde{\boldsymbol{\theta}}^{0*}_P P}(\boldsymbol{u}) = \sqrt{s}\boldsymbol{u}^T\dot{M}^*_{P\alpha}(\hat{\boldsymbol{\theta}}) + \frac{\boldsymbol{u}^T\ddot{M}^*_{P\alpha}(\hat{\boldsymbol{\theta}} + \acute{\boldsymbol{u}}/\sqrt{s})\boldsymbol{u}}{2}$$

where $\acute{\boldsymbol{u}}$ lies between $\mathbf{0}$ and $\boldsymbol{u}/\sqrt{s}$. From Lemma 7, $\sqrt{s}\dot{M}^*_{P\alpha}(\hat{\boldsymbol{\theta}})$ is stochastically bounded in conditional probability given $\mathcal{D}_n$ and $\tilde{\boldsymbol{\theta}}^{0*}_P$; from Lemma 8, conditional on $\mathcal{D}_n$ and $\tilde{\boldsymbol{\theta}}^{0*}_P$, $\ddot{M}^*_{P\alpha}(\hat{\boldsymbol{\theta}} + \acute{\boldsymbol{u}}/\sqrt{s}) - \ddot{M}_n(\hat{\boldsymbol{\theta}}) = o_P(1)$ and $\ddot{M}_n(\hat{\boldsymbol{\theta}})$ converges to a positive-definite matrix. Thus, from the Basic Corollary in page 2 of Hjort and Pollard Hjort and Pollard (2011), the minimizer of $s\gamma(\boldsymbol{u})$, $\sqrt{s}(\tilde{\boldsymbol{\theta}}^{\alpha}_P - \hat{\boldsymbol{\theta}})$, satisfies that

$$\sqrt{s}(\tilde{\boldsymbol{\theta}}^{\alpha}_P - \hat{\boldsymbol{\theta}}) = \ddot{M}^{-1}_n(\hat{\boldsymbol{\theta}})\sqrt{s}\dot{M}^*_P(\hat{\boldsymbol{\theta}}) + o_P(1), \tag{A.66}$$

which implies that

$$\sqrt{s}\{\Lambda^{\alpha}_P(\tilde{\boldsymbol{\theta}}^{0*}_P)\}^{-1/2}\ddot{M}_n(\hat{\boldsymbol{\theta}})(\tilde{\boldsymbol{\theta}}^{\alpha}_P - \hat{\boldsymbol{\theta}}) \to \mathbb{N}(\mathbf{0}, \boldsymbol{I}), \tag{A.67}$$

in conditional distribution given $\mathcal{D}_n$ and $\tilde{\boldsymbol{\theta}}^{0*}_P$.

Next, we will check the distance between $\Lambda^{\alpha}_P(\tilde{\boldsymbol{\theta}}^{0*}_P)$ and $\Lambda^{\alpha}_P(\hat{\boldsymbol{\theta}})$. Let $\Lambda^{\alpha}_{P\varrho_n}(\hat{\boldsymbol{\theta}})$ have the same expression as $\Lambda^{\alpha}_P(\hat{\boldsymbol{\theta}})$ in Theorem 6 except that $\pi^{\rm opt}_{Pi}(\hat{\boldsymbol{\theta}})$ in the denominator is replaced by

$$\pi^{\varrho_n}_{P\alpha i}(\hat{\boldsymbol{\theta}}) = (1-\alpha)\pi^{\varrho_n}_{Pi}(\hat{\boldsymbol{\theta}}) + \alpha\frac{1}{n} \quad \text{with} \quad \pi^{\varrho_n}_{Pi} = \frac{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|\wedge H_{\varrho_n}}{\sum_{j=1}^{n}\{\|\dot{m}(Z_j,\hat{\boldsymbol{\theta}})\|\wedge H_{\varrho_n}\}}.$$

We have that

$$\begin{aligned}
\|\Lambda^{\alpha}_P(\tilde{\boldsymbol{\theta}}^{0*}_P) - \Lambda^{\alpha}_{P\varrho_n}(\hat{\boldsymbol{\theta}})\| &\le \frac{s}{n^2}\sum_{i=1}^{n}\left|\frac{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2}{\{s\tilde{\pi}^{\rm opt}_{P\alpha i}(\tilde{\boldsymbol{\theta}}^{0*}_P)\}\wedge 1} - \frac{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2}{\{s\pi^{\varrho_n}_{P\alpha i}(\hat{\boldsymbol{\theta}})\}\wedge 1}\right| \\
&= \frac{s}{n^2}\sum_{i=1}^{n}\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2\left|\frac{1}{\{s\tilde{\pi}^{\rm opt}_{P\alpha i}(\tilde{\boldsymbol{\theta}}^{0*}_P)\}\wedge 1} - \frac{1}{\{s\pi^{\varrho_n}_{P\alpha i}(\hat{\boldsymbol{\theta}})\}\wedge 1}\right| \\
&= \frac{s}{n^2}\sum_{i=1}^{n}\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2\left|\frac{\{s\tilde{\pi}^{\rm opt}_{P\alpha i}(\tilde{\boldsymbol{\theta}}^{0*}_P)\}\wedge 1 - \{s\pi^{\varrho_n}_{P\alpha i}(\hat{\boldsymbol{\theta}})\}\wedge 1}{[\{s\tilde{\pi}^{\rm opt}_{P\alpha i}(\tilde{\boldsymbol{\theta}}^{0*}_P)\}\wedge 1][\{s\pi^{\varrho_n}_{P\alpha i}(\hat{\boldsymbol{\theta}})\}\wedge 1]}\right| \\
&< \frac{1}{\alpha^2}\sum_{i=1}^{n}\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2\left|\tilde{\pi}^{\rm opt}_{Pi}(\tilde{\boldsymbol{\theta}}^{0*}_P) - \pi^{\varrho_n}_{Pi}(\hat{\boldsymbol{\theta}})\right|
\end{aligned} \tag{A.68}$$

If $\varrho > 0$, then from

$$\begin{aligned}
& n\left|\tilde{\pi}^{\rm opt}_{Pi}(\tilde{\boldsymbol{\theta}}^{0*}_P) - \pi^{\varrho_n}_{Pi}(\hat{\boldsymbol{\theta}})\right| \\
&= \left|\frac{\|\dot{m}(Z_i,\tilde{\boldsymbol{\theta}}^{0*}_P)\|\wedge H^{0*}}{\Psi^{0*}_{\varrho_n}} - \frac{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|\wedge H_{\varrho_n}}{\Psi_{\varrho_n}}\right| \\
&\le \left|\frac{\|\dot{m}(Z_i,\tilde{\boldsymbol{\theta}}^{0*}_P)\|\wedge H^{0*} - \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|\wedge H_{\varrho_n}}{\Psi^{0*}_{\varrho_n}}\right| + \{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|\wedge H_{\varrho_n}\}\left|\frac{\Psi^{0*}_{\varrho_n} - \Psi_{\varrho_n}}{\Psi^{0*}_{\varrho_n}\Psi_{\varrho_n}}\right| \\
&\le \frac{\left|\|\dot{m}(Z_i,\tilde{\boldsymbol{\theta}}^{0*}_P)\| - \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|\right|}{\Psi^{0*}_{\varrho_n}} + \frac{|H^{0*} - H_{\varrho_n}|}{\Psi^{0*}_{\varrho_n}} + \{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|\wedge H_{\varrho_n}\}\frac{|\Psi^{0*}_{\varrho_n} - \Psi_{\varrho_n}|}{\Psi^{0*}_{\varrho_n}\Psi_{\varrho_n}},
\end{aligned}$$

we have that

$$
\begin{aligned}
&\|\Lambda_P^\alpha(\tilde{\boldsymbol{\theta}}_P^{0*}) - \Lambda_{P\varrho_n}^\alpha(\hat{\boldsymbol{\theta}})\| \\
&< \frac{1}{\alpha^2}\sum_{i=1}^n \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2 \left|\tilde{\pi}_{Pi}^{\text{opt}}(\tilde{\boldsymbol{\theta}}_P^{0*}) - \pi_{Pi}^{\varrho_n}(\hat{\boldsymbol{\theta}})\right| \\
&\le \frac{1}{\alpha^2\Psi_{\varrho_n}^{0*}}\sum_{i=1}^n \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2 \left|\|\dot{m}(Z_i,\tilde{\boldsymbol{\theta}}_P^{0*})\| - \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|\right| \\
&\quad + \frac{|H^{0*}-H_{\varrho_n}|}{\alpha^2\Psi_{\varrho_n}^{0*}}\sum_{i=1}^n \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2 + \frac{|\Psi_{\varrho_n}^{0*}-\Psi_{\varrho_n}|}{\alpha^2\Psi_{\varrho_n}^{0*}\Psi_{\varrho_n}}\sum_{i=1}^n \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^3 = o_P(1),
\end{aligned}
$$

by (A.51) and Lemma 9.

If $\varrho = 0$, then,

$$
\begin{aligned}
& n\left|\tilde{\pi}_{Pi}^{\text{opt}}(\tilde{\boldsymbol{\theta}}_P^{0*}) - \pi_{Pi}^{\varrho_n}(\hat{\boldsymbol{\theta}})\right| \\
&\le \frac{\left|\|\dot{m}(Z_i,\tilde{\boldsymbol{\theta}}_P^{0*})\| \wedge H^{0*} - \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\| \wedge H_{\varrho_n}\right|}{\Psi_{\varrho_n}^{0*}} + \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\| \left|\frac{\Psi_{\varrho_n}^{0*}-\Psi_{\varrho_n}}{\Psi_{\varrho_n}^{0*}\Psi_{\varrho_n}}\right| \\
&\le \frac{\left|\|\dot{m}(Z_i,\tilde{\boldsymbol{\theta}}_P^{0*})\| - \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|\right|}{\Psi_{\varrho_n}^{0*}} + \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\| \frac{|\Psi_{\varrho_n}^{0*}-\Psi_{\varrho_n}|}{\Psi_{\varrho_n}^{0*}\Psi_{\varrho_n}} \\
&\quad + \frac{\|\dot{m}(Z_i,\tilde{\boldsymbol{\theta}}_P^{0*})\|}{\Psi_{\varrho_n}^{0*}} I\Big\{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\| \ge H_{\varrho_n}\Big\} + \frac{H_{\varrho_n}}{\Psi_{\varrho_n}^{0*}} I\Big\{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\| \ge H_{\varrho_n}\Big\} \\
&\quad + \frac{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|}{\Psi_{\varrho_n}^{0*}} I\Big\{\|\dot{m}(Z_i,\tilde{\boldsymbol{\theta}}_P^{0*})\| \ge H^{0*}\Big\} + \frac{H^{0*}}{\Psi_{\varrho_n}^{0*}} I\Big\{\|\dot{m}(Z_i,\tilde{\boldsymbol{\theta}}_P^{0*})\| \ge H^{0*}\Big\} \\
&\equiv \Delta_{3i} + \Delta_{4i} + \Delta_{5i} + \Delta_{6i} + \Delta_{7i} + \Delta_{8i}.
\end{aligned} \tag{A.69}
$$

From (A.51) and Lemma 9, we know that

$$
\frac{1}{n}\sum_{i=1}^n \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2 \Delta_{3i} = o_P(1) \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2 \Delta_{4i} = o_P(1). \tag{A.70}
$$

Note that

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2 \|\dot{m}(Z_i,\tilde{\boldsymbol{\theta}}_P^{0*})\| I\Big\{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\| \ge H_{\varrho_n}\Big\} \\
&\le \left\{\frac{1}{n}\sum_{i=1}^n \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^4\right\}^{\frac{1}{2}} \left\{\frac{1}{n}\sum_{i=1}^n \|\dot{m}(Z_i,\tilde{\boldsymbol{\theta}}_P^{0*})\|^4\right\}^{\frac{1}{4}} \left[\frac{1}{n}\sum_{i=1}^n I\Big\{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\| \ge H_{\varrho_n}\Big\}\right]^{\frac{1}{4}} \\
&= o_P(1),
\end{aligned}
$$

because

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n I\{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\| \ge H_{\varrho_n}\} = o_P(1), \quad \frac{1}{n}\sum_{i=1}^n \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^4 = O_P(1), \\
&\quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n \|\dot{m}(Z_i,\tilde{\boldsymbol{\theta}}_P^{0*})\|^4 = O_P(1).
\end{aligned}
$$

Thus,

$$
\frac{1}{n}\sum_{i=1}^n \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2 \Delta_{5i} = o_P(1). \tag{A.71}
$$

If $\|\dot{m}(Z,\boldsymbol{\theta})\|$ is bounded, then

$$\frac{H_{\varrho_n}}{n}\sum_{i=1}^{n}\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2 I\Big\{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|\geq H_{\varrho_n}\Big\} \leq \frac{n-\lceil n(1-\varrho_n)\rceil}{n}\|\dot{m}(Z,\hat{\boldsymbol{\theta}})\|^3_{(n)}= o_P(1); \tag{A.72}$$

otherwise

$$\frac{H_{\varrho_n}}{n}\sum_{i=1}^{n}\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2 I\Big\{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|\geq H_{\varrho_n}\Big\} \leq \frac{1}{nH_{\varrho_n}}\sum_{i=1}^{n}\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^4= o_P(1). \tag{A.73}$$

Thus we know that

$$\frac{1}{n}\sum_{i=1}^{n}\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2\Delta_{6i} = o_P(1). \tag{A.74}$$

Similarly, we can obtain that

$$\frac{1}{n}\sum_{i=1}^{n}\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2\Delta_{7i} = o_P(1) \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n}\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2\Delta_{8i} = o_P(1). \tag{A.75}$$

Combining (A.68), (A.69), (A.70), (A.71), (A.74), and (A.75), we know that

$$\|\Lambda_P^{\alpha}(\tilde{\boldsymbol{\theta}}_P^{0*}) - \Lambda_{P_{\varrho_n}}^{\alpha}(\hat{\boldsymbol{\theta}})\|= o_P(1).$$

To finish the proof for the case of $\varrho = 0$, we only need to show that $\|\Lambda_{P_{\varrho_n}}^{\alpha}(\hat{\boldsymbol{\theta}}) - \Lambda_R^{\alpha}(\hat{\boldsymbol{\theta}})\|= o_P(1)$. Let $\Psi_\infty = \frac{1}{n}\sum_{i=1}^{n}\{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|\}$. We notice that

$$\begin{aligned}
& n\left|\pi_{Pi}^{\varrho_n}(\hat{\boldsymbol{\theta}}) - \pi_{Ri}^{\text{opt}}(\hat{\boldsymbol{\theta}})\right| \\
&\leq \frac{\left|\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|\wedge H_{\varrho_n} - \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|\right|}{\Psi_{\varrho_n}} + \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|\left|\frac{\Psi_{\varrho_n}-\Psi_\infty}{\Psi_{\varrho_n}\Psi_\infty}\right| \\
&\leq \frac{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|}{\Psi_{\varrho_n}} I\Big\{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|\geq H_{\varrho_n}\Big\} + \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|\frac{|\Psi_{\varrho_n}-\Psi_\infty|}{\Psi_{\varrho_n}\Psi_\infty} \\
&\leq \frac{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2}{\Psi_{\varrho_n}H_{\varrho_n}} + \|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|\frac{|\Psi_{\varrho_n}-\Psi_\infty|}{\Psi_{\varrho_n}\Psi_\infty} \equiv \Delta_{9i} + \Delta_{10i}.
\end{aligned} \tag{A.76}$$

With this result, it can be shown that

$$\frac{1}{n}\sum_{i=1}^{n}\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2\Delta_{9i} = o_P(1) \quad \text{and} \qquad \frac{1}{n}\sum_{i=1}^{n}\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2\Delta_{10i} = o_P(1),$$

which indicates that $\|\Lambda_{P_{\varrho_n}}^{\alpha}(\hat{\boldsymbol{\theta}}) - \Lambda_R^{\alpha}(\hat{\boldsymbol{\theta}})\|= o_P(1)$.

From Slutsky's theorem, we know that give $\mathcal{D}_n$, $\tilde{\boldsymbol{\theta}}_P^{0*}$, as s_0, s, and n goes to infinity,

$$\sqrt{s}\{V_P^{\alpha}(\hat{\boldsymbol{\theta}})\}^{-1/2}(\tilde{\boldsymbol{\theta}}_P^{\alpha} - \hat{\boldsymbol{\theta}}) \to \mathbb{N}\left(\mathbf{0},\boldsymbol{I}\right),$$

in conditional distribution. From similar arguments used in the proof of Theorem 1, we know that the convergence also holds in unconditional distribution, and this finishes the proof.

□

Proof of Remark 8. Since $\Lambda_R^{\text{opt}}(\hat{\boldsymbol{\theta}})$ has the minimum trace among all choices of sampling probabilities, if $\alpha \neq 0$ then $\text{tr}\{\Lambda_R^{\text{opt}}(\hat{\boldsymbol{\theta}})\} < \text{tr}\{\Lambda_R^{\alpha}(\hat{\boldsymbol{\theta}})\}$. On the other hand,

$$\text{tr}\{\Lambda_R^{\alpha}(\hat{\boldsymbol{\theta}})\} = \frac{1}{n^2}\sum_{i=1}^{n}\frac{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2}{(1-\alpha)\pi_i^{\text{Ropt}} + \alpha\frac{1}{n}} < \frac{1}{n^2}\sum_{i=1}^{n}\frac{\|\dot{m}(Z_i,\hat{\boldsymbol{\theta}})\|^2}{(1-\alpha)\pi_i^{\text{Ropt}}}$$

$$= \frac{1}{(1-\alpha)n^2} \left\{ \sum_{i=1}^{n} \|\dot{m}(Z_i, \hat{\boldsymbol{\theta}})\| \right\}^2 = \frac{\mathrm{tr}_{opt}\{\Lambda_R(\hat{\boldsymbol{\theta}})\}}{1-\alpha},$$

and this finishes the proof for $\Lambda_R^\alpha(\hat{\boldsymbol{\theta}})$ from subsampling with replacement. For $\Lambda_P^\alpha(\hat{\boldsymbol{\theta}})$ from Poisson subsampling, the proof is similar.

□

References

Chanda, K. (1971). Asymptotic distribution of sample quantiles for exchangeable random variables. *Calcutta Statistical Association Bulletin*, 20(4):135–142.

Cheng, G. and Huang, J. (2010). Bootstrap consistency for general semiparametric m-estimation. *The Annals of Statistics*, 38(5):2884–2915.

Hjort, N. L. and Pollard, D. (2011). Asymptotics for minimisers of convex processes. *arXiv preprint arXiv:1107.3806*.

Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization*. Springer.

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press, London.

Xiong, S. and Li, G. (2008). Some results on the convergence of conditional distributions. *Statistics & Probability Letters*, 78(18):3249–3253.