Supplementary Material for Optimal Subsampling for Large-scale Network-based Logistic Regression

Jue Liu[*] Zhuocheng Lu[†]

1 Proofs

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

1.1 Proof of Theorem 3.1 We begin by establishing a lemma that will be used in the proof of Theorems 3.1.

LEMMA 1.1. *If Conditions 3.1-3.4 hold, then conditional on $\mathcal{F}_n$ in probability,*

$$\tilde{\boldsymbol{M}}(\hat{\boldsymbol{\theta}}_{MLE}) - \boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE}) = O_{P|\mathcal{F}_n}(r^{-1/2}), \tag{1.1}$$

$$\frac{\partial \ell^*_{adj}(\hat{\boldsymbol{\theta}}_{MLE})}{\partial \boldsymbol{\theta}} = O_{P|\mathcal{F}_n}(r^{-1/2}), \tag{1.2}$$

where

$$\tilde{\boldsymbol{M}}(\hat{\boldsymbol{\theta}}_{MLE}) = \begin{pmatrix} \tilde{\boldsymbol{M}}_1(\hat{\boldsymbol{\beta}}_{MLE}) & 0 \\ 0 & \tilde{\boldsymbol{M}}_2(\hat{\boldsymbol{\gamma}}_{MLE}) \end{pmatrix} \tag{1.3}$$

where

$$\tilde{\boldsymbol{M}}_1(\hat{\boldsymbol{\beta}}_{MLE}) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\omega_i}p_i(\hat{\boldsymbol{\beta}}_{MLE})(1-p_i(\hat{\boldsymbol{\beta}}_{MLE}))\boldsymbol{x}_i\boldsymbol{x}_i^T, \tag{1.4}$$

$$\tilde{\boldsymbol{M}}_2(\hat{\boldsymbol{\gamma}}_{MLE}) = -\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{1\le j\le n\\ j\ne i}}\frac{\delta_i\delta_j}{\omega_i\omega_j}\pi_{ij}^{y_iy_j}(\hat{\boldsymbol{\gamma}}_{MLE})(1-\pi_{ij}^{y_iy_j}(\hat{\boldsymbol{\gamma}}_{MLE}))\boldsymbol{z}_{ij}\boldsymbol{z}_{ij}^T, \tag{1.5}$$

and

$$\frac{\partial \ell^*_{adj}(\hat{\boldsymbol{\theta}}_{MLE})}{\partial \boldsymbol{\theta}} = (\frac{\partial \ell_1^*(\hat{\boldsymbol{\beta}}_{MLE})}{\partial \boldsymbol{\beta}^T}, \frac{\partial \ell_2^*(\hat{\boldsymbol{\gamma}}_{MLE})}{\partial \boldsymbol{\gamma}^T})^T \tag{1.6}$$

Proof. Direct calculation shows that

$$E(\tilde{\boldsymbol{M}}(\hat{\boldsymbol{\theta}}_{MLE})|\mathcal{F}_n) = \boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE}). \tag{1.7}$$

For any element $\tilde{\boldsymbol{M}}_1(\hat{\boldsymbol{\beta}}_{MLE})^{j_1j_2}$ of $\tilde{\boldsymbol{M}}_1(\hat{\boldsymbol{\beta}}_{MLE})$, where $1\le j_1,j_2\le p$, we have

$$\begin{aligned} Var\big(\tilde{\boldsymbol{M}}_1(\hat{\boldsymbol{\beta}}_{MLE})^{j_1j_2}|\mathcal{F}_n\big) &= \frac{1}{n^2}\sum_{i=1}^{n}(\frac{1}{\omega_i}-1)p_i(\hat{\boldsymbol{\beta}}_{MLE})^2(1-p_i(\hat{\boldsymbol{\beta}}_{MLE}))^2x_{ij_1}^2x_{ij_2}^2 \\ &\le \frac{1}{16n^2}\max_{1\le j\le n}\frac{1}{n\omega_j}\sum_{i=1}^{n}||\boldsymbol{x}_i||^4 \\ &= O_P(r^{-1}) \end{aligned} \tag{1.8}$$

[*]School of Software and Microelectronics, Peking University, Beijing, China. Email: liujue@pku.edu.cn.
[†]Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. Email: lzc1997@pku.edu.cn.

where the last equality follows from Condition 3.4. Moreover, let $b_{ij} = \pi_{ij}^{y_i y_j}(\hat{\boldsymbol{\gamma}}_{MLE})(1 - \pi_{ij}^{y_i y_j}(\hat{\boldsymbol{\gamma}}_{MLE}))$, then for any element of $\tilde{\boldsymbol{M}}_2(\hat{\boldsymbol{\gamma}}_{MLE})^{j_1 j_2}$ of $\tilde{\boldsymbol{M}}_2(\hat{\boldsymbol{\gamma}}_{MLE})$, where $1 \le j_1, j_2 \le 5$, we have

(1.9)

$$
\begin{aligned}
Var\big(\tilde{\boldsymbol{M}}_2(\hat{\boldsymbol{\gamma}}_{MLE})^{j_1 j_2}|\mathcal{F}_n\big) &= \frac{1}{n^2(n-1)^2}\Bigg\{\sum_{i=1}^{n}\sum_{\substack{1\le j\le n\\ j\neq i}}(\frac{1}{\omega_i\omega_j}-1)b_{ij}^2 z_{ijj_1}^2 z_{ijj_2}^2 + b_{ij}b_{ji}z_{ijj_1}z_{ijj_2}z_{jij_1}z_{jij_2} \\
&\quad + \sum_{i=1}^{n}(\frac{1}{\omega_i}-1)\sum_{\substack{1\le j,k\le n\\ j\neq i\\ k\neq i j\neq k}}[b_{ij}z_{ijj_1}z_{ijj_2} + b_{ji}z_{jij_1}z_{jij_2}][b_{ik}z_{ikj_1}z_{ikj_2} + b_{ki}z_{kij_1}z_{kij_2}]\Bigg\} \\
&\le \frac{1}{16(n-1)^2}\max_{1\le l\le n}(\frac{1}{n\omega_l})^2\cdot n(n-1) + \frac{1}{16n(n-1)^2}\max_{1\le l\le n}(\frac{1}{n\omega_l})\cdot 4n(n-1)(n-2) \\
&= O_P(r^{-1})
\end{aligned}
$$

where the inequality follows from the fact that all elements of z_{ij} are smaller than 1. By 1.8 and 1.9, (1.1) holds.

To prove 1.2, direct calculation yields,

$$
E\big[\frac{\partial \ell_{adj}^*(\hat{\boldsymbol{\theta}}_{MLE})}{\partial \boldsymbol{\theta}}|\mathcal{F}_n\big] = \frac{\partial \ell_{adj}(\hat{\boldsymbol{\theta}}_{MLE})}{\partial \boldsymbol{\theta}} = 0. \tag{1.10}
$$

From condition 3.2,

$$
\begin{aligned}
Var\big[\frac{\partial \ell_1^*(\hat{\boldsymbol{\beta}}_{MLE})}{\partial \boldsymbol{\beta}}|\mathcal{F}_n\big] &= \frac{1}{n^2}\sum_{i=1}^{n}(\frac{1}{\omega_i}-1)(y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE}))^2\boldsymbol{x}_i\boldsymbol{x}_i^T \\
&\le \frac{1}{n^2}\max_{1\le j\le n}\frac{1}{n\omega_j}\cdot\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^T \\
&= O_P(r^{-1}).
\end{aligned} \tag{1.11}
$$

Moreover, let $c_{ij} = a_{ij} - \pi_{ij}^{y_i y_j}(\hat{\boldsymbol{\gamma}}_{MLE})$, we have

(1.12)

$$
\begin{aligned}
Var\big[\frac{\partial \ell_2^*(\hat{\boldsymbol{\gamma}}_{MLE})}{\partial \boldsymbol{\gamma}}|\mathcal{F}_n\big] &= \frac{1}{n^2(n-1)^2}\Bigg\{\sum_{i=1}^{n}\sum_{\substack{1\le j\le n\\ j\neq i}}(\frac{1}{\omega_i\omega_j}-1)c_{ij}^2\boldsymbol{z}_{ij}\boldsymbol{z}_{ij}^T + c_{ij}c_{ji}\boldsymbol{z}_{ij}\boldsymbol{z}_{ji}^T \\
&\quad + \sum_{i=1}^{n}(\frac{1}{\omega_i}-1)\sum_{\substack{1\le j,k\le n\\ j\neq i\\ k\neq i\\ j\neq k}}[c_{ij}\boldsymbol{z}_{ij} + c_{ji}\boldsymbol{z}_{ji}][c_{ij}\boldsymbol{z}_{ik}^T + c_{ij}\boldsymbol{z}_{ki}^T]\Bigg\} \\
&\le \max_{1\le l\le n}(\frac{1}{n\omega_l})^2\cdot\frac{2}{(n-1)^2}\sum_{i=1}^{n}\sum_{\substack{1\le j\le n\\ j\neq i}}\boldsymbol{z}_{ij}\boldsymbol{z}_{ij}^T + \max_{1\le l\le n}(\frac{1}{n\omega_l})\cdot\frac{1}{n(n-1)^2}\sum_{i=1}^{n}\sum_{\substack{1\le j,k\le n\\ j\neq i\\ k\neq i\\ j\neq k}}(\boldsymbol{z}_{ij}+\boldsymbol{z}_{ji})(\boldsymbol{z}_{ik}+\boldsymbol{z}_{ki}) \\
&= O_P(r^{-1}).
\end{aligned}
$$

Finally, we calculate the covariance between $\frac{\partial \ell_1^*(\hat{\boldsymbol{\beta}}_{MLE})}{\partial \boldsymbol{\beta}}$ and $\frac{\partial \ell_2^*(\hat{\gamma}_{MLE})}{\partial \boldsymbol{\gamma}^T}$. Direct calculation yields,

(1.13)

$$
\begin{aligned}
&Cov(\frac{\partial \ell_1^*(\hat{\boldsymbol{\beta}}_{MLE})}{\partial \boldsymbol{\beta}}, \frac{\partial \ell_2^*(\hat{\gamma}_{MLE})}{\partial \boldsymbol{\gamma}^T}|\mathcal{F}_n)\\
&= \frac{1}{n^2(n-1)}\sum_{i=1}^{n}(\frac{1}{\omega_i}-1)[y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE})]\boldsymbol{x}_i \sum_{\substack{1\le j\le n\\ j\neq i}}\big\{[a_{ij}-\pi_{ij}^{y_iy_j}(\hat{\boldsymbol{\gamma}}_{MLE})]\boldsymbol{z}_{ij}^T + [a_{ji}-\pi_{ji}^{y_jy_i}(\hat{\boldsymbol{\gamma}}_{MLE})]\boldsymbol{z}_{ji}^T\big\}\\
&\le \max_{1\le l\le n}\frac{1}{n\omega_l}\cdot\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{1\le j\le n\\ j\neq i}}\big\{[y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE})][a_{ij}-\pi_{ij}^{y_iy_j}(\hat{\boldsymbol{\gamma}}_{MLE})]\boldsymbol{x}_i + [y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE})][a_{ji}-\pi_{ji}^{y_jy_i}(\hat{\boldsymbol{\gamma}}_{MLE})]\boldsymbol{x}_i\big\}\\
&= O_P(r^{-1})
\end{aligned}
$$

Therfore, combining 1.11,1.12 and 1.13 and using Markov's inequality, 1.2 holds. $\square$

Now we prove Theorem 3.1.

Following the same procedure as the proof of Lemma 1.2, we can prove that

$$E\big[\frac{\partial \ell_{adj}^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|\mathcal{F}_n\big] = \frac{\partial \ell_{adj}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \tag{1.14}$$

$$Var\big[\frac{\partial \ell_{adj}^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|\mathcal{F}_n\big] = O_P(r^{-1}). \tag{1.15}$$

Therefore, as $r\to\infty$, $\frac{\partial \ell_{adj}^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \ell_{adj}^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \to 0$ for all $\theta\in\Lambda$ in conditional probability given $\mathcal{F}_n$. Note that the parameter pace is compact and $\hat{\boldsymbol{\theta}}_{MLE}$ is the unique solution of $\frac{\partial \ell_{adj}^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0$ under Condition 2.3 [3]. Thus, from Theorem 5.9 and its remark of [2], we have

$$||\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{MLE}|| = o_{P|\mathcal{F}_n}(1), \tag{1.16}$$

as $n\to\infty$, $r\to\infty$, conditionally on $\mathcal{F}_n$ in probability.

By Taylor's theorem (c.f. Chapter 4 of [1]),

$$0 = \dot{\ell}_{adj,k}^*(\tilde{\boldsymbol{\theta}}) = \dot{\ell}_{adj,k}^*(\hat{\boldsymbol{\theta}}_{MLE}) + \frac{\partial \dot{\ell}_{adj,k}^*(\hat{\boldsymbol{\theta}}_{MLE})}{\partial \boldsymbol{\theta}^T}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{MLE}) + R_k, \tag{1.17}$$

where $\dot{\ell}_{adj,k}^*(\hat{\boldsymbol{\theta}}_{MLE})$ is the partial derivative of $\ell_{adj}^*(\boldsymbol{\theta})$ with respect to the kth element of $\boldsymbol{\theta}$, and

$$R_k = (\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{MLE})^T\int_0^1\int_0^1 \frac{\partial^2 \dot{\ell}_{adj,k}^*(\hat{\boldsymbol{\theta}}_{MLE} + uv(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{MLE}))}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} v du dv(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{MLE}) \tag{1.18}$$

with

(1.19)

$$\frac{\partial^2 \dot{\ell}_{adj,k}^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} = \begin{cases} \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\pi_i} p_i(\boldsymbol{\beta})(1-p_i(\boldsymbol{\beta}))(1-2p_i(\boldsymbol{\beta}))x_{ik}\boldsymbol{x}_i\boldsymbol{x}_i^T & 1\le k\le p\\ \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{\substack{1\le j\le n\\ j\neq i}} \frac{\delta_i\delta_j}{\pi_i\pi_j}\pi_{ij}^{y_iy_j}(\boldsymbol{\gamma})(1-\pi_{ij}^{y_iy_j}(\boldsymbol{\gamma}))(1-2\pi_{ij}^{y_iy_j}(\boldsymbol{\gamma}))z_{ijk}\boldsymbol{z}_{ij}\boldsymbol{z}_{ij}^T & p+1\le k\le p+5. \end{cases}$$

By using similar arguments in the proof of 1.1 in Lemma 1.1 under condition 3.2 and the fact that every element of $\boldsymbol{z}_{ij}$ is smaller than 1, we have

$$\frac{\partial^2 \dot{\ell}_{adj,k}^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} - E\big[\frac{\partial^2 \dot{\ell}_{adj,k}^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}|\mathcal{F}_n\big] = O_{P|\mathcal{F}_n}(r^{-1/2}). \tag{1.20}$$

Under condition 3.2, $E\big[\frac{\partial^2 \dot{\ell}^*_{adj,k}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}|\mathcal{F}_n\big] = O_P(1)$, therefore,

$$||\frac{\partial^2 \dot{\ell}^*_{adj,k}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}|| = O_{P|\mathcal{F}_n}(1), \tag{1.21}$$

and

$$R_k = O_{P|\mathcal{F}_n}(||\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{MLE}||^2). \tag{1.22}$$

From 1.17 and 1.22,

$$\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{MLE} = -\tilde{\boldsymbol{M}}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\left\{\frac{\partial \ell^*_{adj}(\hat{\boldsymbol{\theta}}_{MLE})}{\partial\boldsymbol{\theta}} + O_{P|\mathcal{F}_n}(||\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{MLE}||^2)\right\} \tag{1.23}$$

By Lemma 1.1, $\tilde{\boldsymbol{M}}(\hat{\boldsymbol{\theta}}_{MLE})^{-1} = O_{P|\mathcal{F}_n}(1)$. Combining this with (1.2), (1.16), and (1.23), we have

$$\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{MLE} = O_{P|\mathcal{F}_n}(r^{-1/2}) + o_{P|\mathcal{F}_n}(||\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{MLE}||), \tag{1.24}$$

which implies that

$$\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{MLE} = O_{P|\mathcal{F}_n}(r^{-1/2}). \tag{1.25}$$

1.2 Proof of Proposition 3.1 Note that

$$\frac{\partial \ell^*_1(\hat{\boldsymbol{\beta}}_{MLE})}{\partial\boldsymbol{\beta}} = \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\omega_i}[y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE})]\boldsymbol{x}_i \equiv \sum_{i=1}^n \boldsymbol{\eta}_i \tag{1.26}$$

Direct calculation shows that

$$E\big[\frac{\partial \ell^*_1(\hat{\boldsymbol{\beta}}_{MLE})}{\partial\boldsymbol{\beta}}|\mathcal{F}_n\big] = \frac{\partial \ell_1(\hat{\boldsymbol{\beta}}_{MLE})}{\partial\boldsymbol{\beta}} = 0 \tag{1.27}$$

$$Var(\frac{\partial \ell^*_1(\hat{\boldsymbol{\beta}}_{MLE})}{\partial\boldsymbol{\beta}}|\mathcal{F}_n) \equiv \boldsymbol{V}_0 = \frac{1}{n^2}\sum_{i=1}^n \frac{(1-\omega_i)[y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE})]^2\boldsymbol{x}_i\boldsymbol{x}_i^T}{\omega_i} = O_P(r^{-1}). \tag{1.28}$$

Now we check the Lindeberg-Feller condition under the conditional distribution. For every $\epsilon > 0$,

$$\begin{aligned}\sum_{i=1}^n E[||\boldsymbol{\eta}_i||I_{(||\boldsymbol{\eta}_i||>\epsilon)}|\mathcal{F}_n] &\le \frac{1}{\epsilon}\sum_{i=1}^n E(||\boldsymbol{\eta}_i||^3|\mathcal{F}_n) = \frac{1}{n^3\epsilon}\sum_{i=1}^n \frac{|y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE})|^3||\boldsymbol{x}_i||^3}{\omega_i^2} \\ &\le \frac{1}{\epsilon}\max_{1\le i\le n}\frac{1}{(n\omega_i)^2}\sum_{i=1}^n \frac{|y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE})|^3||\boldsymbol{x}_i||^3}{n} = O_P(r^{-2}) = o_P(r^{-1}),\end{aligned} \tag{1.29}$$

This and (1.26) show that the Lindeberg-Feller conditions are satisfied in probability. Therefore, by the Linderberg-Feller central limit theorem [2], conditional on $\mathcal{F}_n$,

$$\boldsymbol{V}_0^{-1/2}\frac{\partial \ell^*_1(\hat{\boldsymbol{\beta}}_{MLE})}{\partial\boldsymbol{\beta}} \to N(\boldsymbol{0}, \boldsymbol{I}), \tag{1.30}$$

in distribution. By using the similar argument in the proof of 1.23 and 1.25 in the proof of Theorem 2.1,

$$\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{MLE} = -\tilde{\boldsymbol{M}}_1(\hat{\boldsymbol{\beta}}_{MLE})^{-1}\frac{\partial \ell^*_1(\hat{\boldsymbol{\beta}}_{MLE})}{\partial\boldsymbol{\beta}} + O_{P|\mathcal{F}_n}(r^{-1}). \tag{1.31}$$

Moreover, from Lemma 1.1, $\tilde{\boldsymbol{M}}_1(\hat{\boldsymbol{\beta}}_{MLE}) - \boldsymbol{M}_1(\hat{\boldsymbol{\beta}}_{MLE}) = O_{P|\mathcal{F}_n}(r^{-1/2})$ and therefore,

$$\tilde{\boldsymbol{M}}_1(\hat{\boldsymbol{\beta}}_{MLE})^{-1} - \boldsymbol{M}_1(\hat{\boldsymbol{\beta}}_{MLE})^{-1} = -\boldsymbol{M}_1(\hat{\boldsymbol{\beta}}_{MLE})^{-1}(\tilde{\boldsymbol{M}}_1(\hat{\boldsymbol{\beta}}_{MLE}) - \boldsymbol{M}_1(\hat{\boldsymbol{\beta}}_{MLE}))\tilde{\boldsymbol{M}}_1(\hat{\boldsymbol{\beta}}_{MLE})^{-1} = O_{P|\mathcal{F}_n}(r^{-1/2}). \tag{1.32}$$

Based on Lemma 1.1 and (1.28), it is verified that,

$$\boldsymbol{V}_{\boldsymbol{\beta}} = \boldsymbol{M}_1(\hat{\boldsymbol{\beta}}_{MLE})^{-1}\boldsymbol{V}_0\boldsymbol{M}_1(\hat{\boldsymbol{\beta}}_{MLE})^{-1} = O_P(r^{-1}). \tag{1.33}$$

Combining (1.31), (1.32) and (1.33) yield

$$\begin{aligned}
\boldsymbol{V}_{\boldsymbol{\beta}}^{-1/2}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{MLE}) &= -\boldsymbol{V}_{\boldsymbol{\beta}}^{-1/2}\tilde{\boldsymbol{M}}_1(\hat{\boldsymbol{\beta}}_{MLE})^{-1}\frac{\partial \ell_1^*(\hat{\boldsymbol{\beta}}_{MLE})}{\partial \boldsymbol{\beta}} + O_{P|\mathcal{F}_n}(r^{-1/2}) \\
&= -\boldsymbol{V}_{\boldsymbol{\beta}}^{-1/2}\boldsymbol{M}_1(\hat{\boldsymbol{\beta}}_{MLE})^{-1}\frac{\partial \ell_1^*(\hat{\boldsymbol{\beta}}_{MLE})}{\partial \boldsymbol{\beta}} - \boldsymbol{V}_{\boldsymbol{\beta}}^{-1/2}(\tilde{\boldsymbol{M}}_1(\hat{\boldsymbol{\beta}}_{MLE})^{-1} - \boldsymbol{M}_1(\hat{\boldsymbol{\beta}}_{MLE})^{-1})\frac{\partial \ell_1^*(\hat{\boldsymbol{\beta}}_{MLE})}{\partial \boldsymbol{\beta}} + O_{P|\mathcal{F}_n}(r^{-1/2}) \\
&= -\boldsymbol{V}_{\boldsymbol{\beta}}^{-1/2}\boldsymbol{M}_1(\hat{\boldsymbol{\beta}})^{-1}\boldsymbol{V}_0^{1/2}\boldsymbol{V}_0^{-1/2}\frac{\partial \ell_1^*(\hat{\boldsymbol{\beta}}_{MLE})}{\partial \boldsymbol{\beta}} + O_{P|\mathcal{F}_n}(r^{-1/2}).
\end{aligned} \tag{1.34}$$

The result in Proposition 3.1 follows from Sluksky's Theorem (Theorem 6 of [1]) and the fact that

$$\boldsymbol{V}_{\boldsymbol{\beta}}^{-1/2}\boldsymbol{M}_1(\hat{\boldsymbol{\beta}}_{MLE})^{-1}\boldsymbol{V}_0^{1/2}(\boldsymbol{V}_{\boldsymbol{\beta}}^{-1/2}\boldsymbol{M}_1(\hat{\boldsymbol{\beta}}_{MLE})^{-1}\boldsymbol{V}_0^{1/2})^T = \boldsymbol{V}_{\boldsymbol{\beta}}^{-1/2}\boldsymbol{M}_1(\hat{\boldsymbol{\beta}}_{MLE})^{-1}\boldsymbol{V}_0^{1/2}\boldsymbol{V}_0^{1/2}\boldsymbol{M}_1(\hat{\boldsymbol{\beta}}_{MLE})^{-1}\boldsymbol{V}_{\boldsymbol{\beta}}^{-1/2} = \boldsymbol{I}. \tag{1.35}$$

1.3 Proof of Theorem 4.1. Since $\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{MLE} = O_{P|\mathcal{F}_n}(r^{-1/2})$, by 1.23,

$$\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{MLE} = \tilde{\boldsymbol{M}}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\frac{\partial \ell_{adj}^*(\hat{\boldsymbol{\theta}}_{MLE})}{\partial \boldsymbol{\theta}} + O_{P|\mathcal{F}_n}(r^{-1}). \tag{1.36}$$

By 1.1 of lemma 1.1,

$$\tilde{\boldsymbol{M}}(\hat{\boldsymbol{\theta}}_{MLE})^{-1} - \boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1} = -\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}[\tilde{\boldsymbol{M}}(\hat{\boldsymbol{\theta}}_{MLE}) - \boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})]\tilde{\boldsymbol{M}}(\hat{\boldsymbol{\theta}}_{MLE})^{-1} = O_{P|\mathcal{F}_n}(r^{-1/2}). \tag{1.37}$$

By 1.11,1.12, 1.37, the fact that $\tilde{\boldsymbol{M}}(\hat{\boldsymbol{\theta}}_{MLE})^{-1} = O_{P|\mathcal{F}_n}(1)$ and the uniform intergrability of $\tilde{\boldsymbol{\theta}}$, we have

$$r\left\{E\big[(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{MLE})(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{MLE})^T\big|\mathcal{F}_n\big] - \boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}E\big[\frac{\partial \ell_{adj}^*(\hat{\boldsymbol{\theta}}_{MLE})}{\partial \boldsymbol{\theta}}\frac{\partial \ell_{adj}^*(\hat{\boldsymbol{\theta}}_{MLE})}{\partial \boldsymbol{\theta}}^T\big|\mathcal{F}_n\big]\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\right\} = o_P(1). \tag{1.38}$$

By 1.11, 1.12, 1.13 and the fact that $E\big[\frac{\partial \ell_{adj}^*(\hat{\boldsymbol{\theta}}_{MLE})}{\partial \boldsymbol{\theta}}|\mathcal{F}_n\big] = 0$, it remains to show that

$$r^{1/2}E[(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{MLE})|\mathcal{F}_n] = o_P(1) \tag{1.39}$$

and

$$\frac{1}{n^2(n-1)^2}\sum_{i=1}^{n}\sum_{\substack{1\le j\le n\\ j\ne i}}(\frac{1}{\omega_i\omega_j}-1)[a_{ij}-\pi_{ij}^{y_iy_j}(\hat{\boldsymbol{\gamma}}_{MLE})]^2\boldsymbol{z}_{ij}\boldsymbol{z}_{ij}^T+[a_{ij}-\pi_{ij}^{y_iy_j}(\hat{\boldsymbol{\gamma}}_{MLE})][a_{ji}-\pi_{ji}^{y_jy_i}(\hat{\boldsymbol{\gamma}}_{MLE})]\boldsymbol{z}_{ij}\boldsymbol{z}_{ji}^T = o_P(r^{-1}). \tag{1.40}$$

Since $\tilde{\boldsymbol{M}}(\hat{\boldsymbol{\theta}}_{MLE})^{-1} = O_{P|\mathcal{F}_n}(1)$ and $E\big[\frac{\partial \ell^*_{adj}(\hat{\boldsymbol{\theta}}_{MLE})}{\partial \boldsymbol{\theta}}|\mathcal{F}_n\big] = 0$, 1.39 holds directly from 1.37 and the uniform intergrability of $\tilde{\boldsymbol{\theta}}$. Moreover, by condition 3.4 and simple algebra, 1.40 is $O_P(r^{-2})$ and therefore the desired result follows.

1.4 Proof of Theorem 4.2 and 4.3. Without loss of generality, we asssume $||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_1(\hat{\boldsymbol{\theta}}_{MLE})|| \leq ... \leq ||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})-1\boldsymbol{s}_n(\hat{\boldsymbol{\theta}}_{MLE})||$. Minimizing $tr(\boldsymbol{V})$ is equivalent to solve the following problem:

$$\begin{aligned} min \quad & \sum_{i=1}^{n} tr\big(\frac{1}{\omega_i}||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||^2\big) \\ s.t. \quad & \sum_{i=1}^{n}\omega_i = r \quad 0 \leq \pi_i \leq 1 \quad i = 1, ...n. \end{aligned} \tag{1.41}$$

From Cauchy-Schwarz inequality,

$$\sum_{i=1}^{n}\frac{||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||^2}{\omega_i} = \frac{1}{r}(\sum_{i=1}^{n}\omega_i)(\sum_{j=1}^{n}\omega_j||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_j(\hat{\boldsymbol{\theta}}_{MLE})||^2) \geq \frac{1}{r}\sum_{i=1}^{n}(||\boldsymbol{M}_(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||)^2, \tag{1.42}$$

where the equality holds if and only if $\omega_i \propto ||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||$. Therefore, when $\omega_i = \frac{r||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||}{\sum_{j=1}^{n}||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_j(\hat{\boldsymbol{\theta}}_{MLE})||}$ satisfied that $\omega_i \leq 1$ for all $i = 1, ..., n$, ω_i is the optimal solution.

Otherwise, we can easily see that $\omega_n = 1$ when $\frac{r||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_n(\hat{\boldsymbol{\theta}}_{MLE})||}{\sum_{j=1}^{n}||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_j(\hat{\boldsymbol{\theta}}_{MLE})||} > 1$. Thus, the original problem can be turned into

$$\begin{aligned} min \quad & \sum_{i=1}^{n-1}\frac{||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||}{\omega_i} \\ s.t. \quad & \sum_{i=1}^{n-1}\pi_i = r-1 \quad 0 \leq \pi_i \leq 1 \quad 1 \leq i \leq n-1. \end{aligned} \tag{1.43}$$

Obviously, this is a recursion problem and the optimal solution is

$$\omega_i = \begin{cases} 1 & n-k+1 \leq i \leq n \\ \frac{(r-k)||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||}{\sum_{j=1}^{n-k}||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_j(\hat{\boldsymbol{\theta}}_{MLE})||} & 1 \leq i \leq n-k, \end{cases} \tag{1.44}$$

where k satisfies

$$\frac{(r-k+1)||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_{n-k+1}(\hat{\boldsymbol{\theta}}_{MLE})||}{\sum_{i=1}^{n-k+1}||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||} \geq 1 \quad and \quad \frac{(r-k)||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_{n-k}(\hat{\boldsymbol{\theta}}_{MLE})||}{\sum_{i=1}^{n-k}||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||} < 1. \tag{1.45}$$

Now we prove that

$$\omega_i = r\frac{||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})|| \wedge N}{\sum_{j=1}^{n}(||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_j(\hat{\boldsymbol{\theta}}_{MLE})|| \wedge N)}, \tag{1.46}$$

where $N = \frac{1}{r-k}\sum_{i=1}^{n-k}||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||$. By (1.45) we see that

$$||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_{n-k}(\hat{\boldsymbol{\theta}}_{MLE})|| < N \leq ||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_{n-k+1}(\hat{\boldsymbol{\theta}}_{MLE})||, \tag{1.47}$$

which indicates that

$$||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})|| \wedge N = \begin{cases} ||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})|| & 1 \leq i \leq n-k \\ N & n-k+1 \leq i \leq n. \end{cases} \tag{1.48}$$

Therefore, we have

$$\omega_i = \begin{cases} r\frac{N}{kN+\sum_{j=1}^{n-k}||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_j(\hat{\boldsymbol{\theta}}_{MLE})||} = r\frac{N}{(k+r-k)N} = 1 & n-k+1 \leq i \leq n \\ r\frac{||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||}{kN+\sum_{j=1}^{n-k}||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_j(\hat{\boldsymbol{\theta}}_{MLE})||} = \frac{||\boldsymbol{M}_X^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||}{\sum_{j=1}^{n}||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_j(\hat{\boldsymbol{\theta}}_{MLE})||} & n-k \leq i \leq n, \end{cases} \tag{1.49}$$

which completes the proof.

The proof of Theorem 4.3 is similar, so we omit it.

1.5 Proofs of Theorem 4.4 and Proposition 4.5 Since $\tilde{\omega}_i^{MV} \geq \rho r/n$ and $\tilde{\omega}_i^{MV_c} \geq \rho r/n$, Theorem 3.1 indicates Theorem 4.4, it remains to show that Proposition 4.5 holds.

Since $r_0 r^{-1/2} \to 0$, the contribution of the first step subsample to the estimation equation is $o_{P|\mathcal{F}_n}(r^{-1/2})$. Thus, we can focus on the subsamples drawn from the second step only. For ease of presentation, we use the unified notations ω_i^{sos} to denote $\omega_i^{MV} \wedge 1$ or $\omega_i^{MV_c} \wedge 1$, $\tilde{\omega}_i^{sos}$ to denote $\tilde{\omega}_i^{MV} \wedge 1$ or $\tilde{\omega}_i^{MV_c} \wedge 1$, h_i^{sos} to denote $||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||$ or $||s_i(\hat{\boldsymbol{\theta}}_{MLE})||$ and $\tilde{h}_i^{sos}$ to denote $||\boldsymbol{M}(\tilde{\boldsymbol{\theta}}_0)^{-1}\boldsymbol{s}_i(\tilde{\boldsymbol{\theta}}_0)||$ or $||s_i(\tilde{\boldsymbol{\theta}}_0)||$. Let

$$\frac{\partial \ell^*_{\tilde{\boldsymbol{\theta}}_0}(\hat{\boldsymbol{\beta}}_{MLE})}{\partial \boldsymbol{\beta}} = \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\tilde{\omega}_i^{sos}}|y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE})|\boldsymbol{x}_i \equiv \sum_{i=1}^n \boldsymbol{\eta}_i^{\tilde{\boldsymbol{\theta}}_0}. \tag{1.50}$$

Direct calculation shows that

$$E\Big[\frac{\partial \ell^*_{\tilde{\boldsymbol{\theta}}_0}(\hat{\boldsymbol{\beta}}_{MLE})}{\partial \boldsymbol{\beta}}|\mathcal{F}_n\Big] = 0 \tag{1.51}$$

and

$$Var(\frac{\partial \ell^*_{\tilde{\boldsymbol{\theta}}_0}(\hat{\boldsymbol{\beta}}_{MLE})}{\partial \boldsymbol{\beta}}|\mathcal{F}_n) \equiv \boldsymbol{V}_0^{\tilde{\boldsymbol{\theta}}_0} = \frac{1}{n^2}\sum_{i=1}^n \frac{(1-\omega_i^{sos})[y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE})]^2\boldsymbol{x}_i\boldsymbol{x}_i^T}{\tilde{\omega}_i^{sos}} = O_P(r^{-1}). \tag{1.52}$$

Meanwhile, for every $\epsilon > 0$,

$$\begin{aligned}\sum_{i=1}^n E[||\boldsymbol{\eta}_i^{\tilde{\boldsymbol{\theta}}_0}||I_{(||\boldsymbol{\eta}_i^{\tilde{\boldsymbol{\theta}}_0}||>\epsilon)}|\mathcal{F}_n] &\leq \frac{1}{\epsilon}\sum_{i=1}^n E(||\boldsymbol{\eta}_i^{\tilde{\boldsymbol{\theta}}_0}||^3|\mathcal{F}_n) = \frac{1}{n^3\epsilon}\sum_{i=1}^n \frac{|y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE})|^3||\boldsymbol{x}_i||^3}{(\tilde{\omega}_i^{sos})^2} \\ &\leq \frac{1}{\epsilon}\max_{1\leq i\leq n}\frac{1}{(n\tilde{\omega}_i^{sos})^2}\sum_{i=1}^n \frac{|y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE})|^3||\boldsymbol{x}_i||^3}{n} = O_P(r^{-2}) = o_P(r^{-1}),\end{aligned} \tag{1.53}$$

This and (1.53) show that the Lindeberg-Feller conditions are satisfied in probability. Therefore, by the Linderberg-Feller central limit theorem [2], conditional on $\mathcal{F}_n$,

$$(\boldsymbol{V}_0^{\tilde{\boldsymbol{\theta}}_0})^{-1/2}\frac{\partial \ell^*_{\tilde{\boldsymbol{\theta}}_0}(\hat{\boldsymbol{\beta}}_{MLE})}{\partial \boldsymbol{\beta}} \to N(\boldsymbol{0}, \boldsymbol{I}), \tag{1.54}$$

in distribution. Now we examine the distance between $\boldsymbol{V}_0^{\tilde{\boldsymbol{\theta}}_0}$ and $\boldsymbol{V}_0$. First,

$$\begin{aligned}||\boldsymbol{V}_0^{\tilde{\boldsymbol{\theta}}_0} - \boldsymbol{V}_0|| &= \left|\left|\frac{1}{n^2}\sum_{i=1}^n \frac{1}{\omega_i^{sos}}|y_i - p_i(\hat{\boldsymbol{\beta}})|^2\boldsymbol{x}_i\boldsymbol{x}_i^T(\frac{\omega_i^{sos}}{\tilde{\omega}_i^{sos}} - 1)\right|\right| \\ &= \frac{1}{n^2}\sum_{i=1}^n \frac{1}{\omega_i^{sos}}\left|\frac{\omega_i^{sos}}{\tilde{\omega}_i^{sos}} - 1\right|[y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE})]^2||\boldsymbol{x}_i||^2 \\ &\leq \max_{1\leq i\leq n}\frac{1}{n\omega_i^{sos}}\sum_{i=1}^n \frac{1}{n}\left|\frac{\omega_i^{sos}}{\tilde{\omega}_i^{sos}} - 1\right|[y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE})]^2||\boldsymbol{x}_i||^2 \\ &\leq (\rho r)^{-1}\sum_{i=1}^n \frac{1}{n}\left|\frac{\omega_i^{sos}}{\tilde{\omega}_i^{sos}} - 1\right|[y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE})]^2||\boldsymbol{x}_i||^2\end{aligned} \tag{1.55}$$

Simple calculation yields,

$$\begin{aligned}&\sum_{i=1}^n \frac{1}{n}\left|\frac{\omega_i^{sos}}{\tilde{\omega}_i^{sos}} - 1\right|[y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE})]^2||\boldsymbol{x}_i||^2 \\ &\leq \sum_{i=1}^n \frac{\frac{\sum_{l=1}^n h_l^{sos}}{\sum_{l=1}^n \tilde{h}_l^{sos}}|\tilde{h}_i^{sos} - h_i^{sos}| + |\frac{\sum_{l=1}^n h_l^{sos}}{\sum_{l=1}^n \tilde{h}_l^{sos}} - 1|h_i^{sos}}{\rho r n^{-1}\sum_{l=1}^n h_l^{sos}}\frac{[y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE})]^2||\boldsymbol{x}_i||^2}{n} \\ &= \frac{\sum_{l=1}^n h_l^{sos}}{\sum_{l=1}^n \tilde{h}_l^{sos}}\sum_{i=1}^n \frac{|\tilde{h}_i^{sos} - h_i^{sos}|[y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE})]^2||\boldsymbol{x}_i||^2}{n\rho r n^{-1}\sum_{l=1}^n h_l^{sos}} + |\frac{\sum_{l=1}^n h_l^{sos}}{\sum_{l=1}^n \tilde{h}_l^{sos}} - 1|\sum_{i=1}^n \frac{h_i^{sos}[y_i - p_i(\hat{\boldsymbol{\beta}}_{MLE})]^2||\boldsymbol{x}_i||^2}{n\rho r n^{-1}\sum_{l=1}^n h_l^{sos}}\end{aligned} \tag{1.56}$$

where the first inequality comes from the facts

$$\left|\frac{\omega_i^{sos}}{\tilde{\omega}_i^{sos}}-1\right| \le \left|\frac{(1-\rho)\frac{h_i^{sos}}{\sum_{l=1}^n h_l^{sos}}+\rho r/n-(1-\rho)\frac{\tilde{h}_i^{sos}}{\sum_{l=1}^n \tilde{h}_l^{sos}}-\rho r/n}{[(1-\rho)\frac{h_i^{sos}}{\sum_{l=1}^n h_l^{sos}}+\rho r/n]\wedge 1}\right|$$

$$\le (1-\rho)\left|\frac{\frac{\tilde{h}_i^{sos}}{\sum_{l=1}^n \tilde{h}_l^{sos}}-\frac{h_i^{sos}}{\sum_{l=1}^n h_l^{sos}}}{\rho r/n}\right|$$

$$\le \left|\frac{\left|\frac{\tilde{h}_i^{sos}}{\sum_{l=1}^n \tilde{h}_l^{sos}}-\frac{h_i^{sos}}{\sum_{l=1}^n \tilde{h}_l^{sos}}\right|+\left|\frac{h_i^{sos}}{\sum_{l=1}^n \tilde{h}_l^{sos}}-\frac{h_i^{sos}}{\sum_{l=1}^n h_l^{sos}}\right|}{\rho r n^{-1}}\right| \tag{1.57}$$

$$\le \frac{\frac{\sum_{l=1}^n h_l^{sos}}{\sum_{l=1}^n \tilde{h}_l^{sos}}\left|\tilde{h}_i^{sos}-h_i^{sos}\right|+\left|\frac{\sum_{l=1}^n h_l^{sos}}{\sum_{l=1}^n \tilde{h}_l^{sos}}-1\right|(h_i^{sos})}{\rho r n^{-1}\sum_{l=1}^n h_l^{sos}}$$

To well exam the distance between $\boldsymbol{V}_0^{\tilde{\boldsymbol{\theta}}_0}$ and $\boldsymbol{V}_0$, we will show the following equalities hold:

$$\left|\frac{\sum_{l=1}^n h_l^{sos}}{\sum_{l=1}^n \tilde{h}_l^{sos}}-1\right|\frac{1}{\rho r n^{-1}\sum_{l=1}^n h_l^{sos}}\sum_{i=1}^n \frac{h_i^{sos}[y_i-p_i(\hat{\boldsymbol{\beta}}_{MLE})]^2||\boldsymbol{x}_i||^2}{n}=o_{P|\mathcal{F}_n}(1) \tag{1.58}$$

$$\frac{\sum_{l=1}^n h_l^{sos}}{\sum_{l=1}^n \tilde{h}_l^{sos}}\frac{1}{\rho r n^{-1}\sum_{l=1}^n h_l^{sos}}\sum_{i=1}^n \frac{|\tilde{h}_i^{sos}-h_i^{sos}|[y_i-p_i(\hat{\boldsymbol{\beta}}_{MLE})]^2||\boldsymbol{x}_i||^2}{n}=o_{P|\mathcal{F}_n}(1) \tag{1.59}$$

Now we begin with showing 1.58. It suffices to show that the following 3 equalities hold.

$$\frac{1}{n}\sum_{i=1}^n h_i^{sos}=O_P(1), \tag{1.60}$$

$$\frac{1}{n}\sum_{i=1}^n h_i^{sos}-\frac{1}{n}\sum_{i=1}^n \tilde{h}_i^{sos}=o_{P|\mathcal{F}_n}(1), \tag{1.61}$$

and

$$\sum_{i=1}^n \frac{h_i^{sos}[y_i-p_i(\hat{\boldsymbol{\beta}}_{MLE})]^2||\boldsymbol{x}_i||^2}{n}=O_P(1) \tag{1.62}$$

To show 1.60, it suffices to show that

$$\frac{1}{n}\sum_{i=1}^n ||\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||=O_P(1) \tag{1.63}$$

and

$$\frac{1}{n}\sum_{i=1}^n ||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||=O_P(1). \tag{1.64}$$

1.63 follows directly from the fact that

$$\frac{1}{n}\sum_{i=1}^n |y_i-p_i(\hat{\boldsymbol{\beta}}_{MLE})|||\boldsymbol{x}_i||\le\frac{1}{n}\sum_{i=1}^n ||\boldsymbol{x}_i||=O_P(1) \tag{1.65}$$

and

$$\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{1\le j\le n\\ j\neq i}}|a_{ij}-\pi_{ij}^{y_iy_j}(\hat{\gamma}_{MLE})|||\boldsymbol{z}_{ij}||\le\sqrt{5}=O_P(1). \tag{1.66}$$

With 1.63 and condition 3.3, 1.64 follows directly from the fact that

$$\frac{1}{n}\sum_{i=1}^{n}||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||\le\lambda_{max}\{\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\}\frac{1}{n}\sum_{i=1}^{n}||\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||=O_P(1) \tag{1.67}$$

and therefore 1.60 holds. To prove 1.61, it suffices that

$$\frac{1}{n}\sum_{i=1}^{n}||\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})-\boldsymbol{s}_i(\tilde{\boldsymbol{\theta}}_0)||=o_{P|\mathcal{F}_n}(1) \tag{1.68}$$

and

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left[||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||-||\tilde{\boldsymbol{M}}(\tilde{\boldsymbol{\theta}}_0)^{-1}s_i(\tilde{\boldsymbol{\theta}}_0)||\right]=o_{P|\mathcal{F}_n}(1)\right| \tag{1.69}$$

1.68 follows directly from the fact that

$$\frac{1}{n}\sum_{i=1}^{n}|p_i(\tilde{\boldsymbol{\beta}}_0)-p_i(\hat{\boldsymbol{\beta}}_{MLE})|||\boldsymbol{x}_i||\le\frac{1}{n}\sum_{i=1}^{n}||\hat{\boldsymbol{\beta}}_{MLE}-\tilde{\boldsymbol{\beta}}_0||||\boldsymbol{x}_i||=O_{P|\mathcal{F}_n}(r_0^{-1/2}) \tag{1.70}$$

and

$$\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{1\le j\le n\\ j\neq i}}|\pi_{ij}^{y_iy_j}(\hat{\gamma}_{MLE})-\pi_{ij}^{y_iy_j}(\tilde{\gamma}_0)|||\boldsymbol{z}_{ij}||\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{1\le j\le n\\ j\neq i}}\le||\hat{\gamma}_{MLE}-\tilde{\gamma}_0||\sqrt{5}=O_{P|\mathcal{F}_n}(r_0^{-1/2}). \tag{1.71}$$

With 1.68 and condition 3.3, 1.64 follows directly from the fact that

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}\left[||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})||-||\tilde{\boldsymbol{M}}(\tilde{\boldsymbol{\theta}}_0)^{-1}s_i(\tilde{\boldsymbol{\theta}}_0)||\right]\right|\\
&\le\frac{1}{n}\sum_{i=1}^{n}||\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}[\boldsymbol{s}_i(\hat{\boldsymbol{\theta}})-\boldsymbol{s}_i(\tilde{\boldsymbol{\theta}}_0)]||+\frac{1}{n}\sum_{i=1}^{n}||\tilde{\boldsymbol{M}}(\tilde{\boldsymbol{\theta}}_0)^{-1}[\boldsymbol{s}_i(\hat{\boldsymbol{\theta}})-\boldsymbol{s}_i(\tilde{\boldsymbol{\theta}}_0)]||\\
&\le\lambda_{max}\{\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\}\frac{1}{n}\sum_{i=1}^{n}||\boldsymbol{s}_i(\hat{\boldsymbol{\theta}})-\boldsymbol{s}_i(\tilde{\boldsymbol{\theta}}_0)||+\lambda_{max}\{\tilde{\boldsymbol{M}}(\tilde{\boldsymbol{\theta}}_0)^{-1}\}\frac{1}{n}\sum_{i=1}^{n}||\boldsymbol{s}_i(\hat{\boldsymbol{\theta}})-\boldsymbol{s}_i(\tilde{\boldsymbol{\theta}}_0)||=o_{P|\mathcal{F}_n}(1)
\end{aligned} \tag{1.72}$$

and therefore 1.61 holds. We then prove 1.62. By Cauchy-Schwarz inequality,

$$\sum_{i=1}^{n}\frac{h_i^{sos}[y_i-p_i(\hat{\boldsymbol{\beta}}_{MLE})]^2||\boldsymbol{x}_i||^2}{n}\le\frac{1}{n}\sum_{i=1}^{n}(h_i^{sos})^2\cdot\frac{1}{n}\sum_{i=1}^{n}[y_i-p_i(\hat{\boldsymbol{\beta}}_{MLE})]^4||\boldsymbol{x}_i||^4=O_P(1), \tag{1.73}$$

where the last equality follows from condition 2.2 and the fact that $\frac{1}{n}\sum_{i=1}^{n}(h_i^{sos})^2=O_P(1)$, whose proof follows the same procedure as the proof of 1.60 so we omit it. Combining 1.60, 1.61 and 1.62 proves 1.58. To prove 1.59, with 1.60 and 1.61, it suffices to show that

$$\sum_{i=1}^{n}\frac{|\tilde{h}_i^{sos}-h_i^{sos}|[y_i-p_i(\hat{\boldsymbol{\beta}}_{MLE})]^2||\boldsymbol{x}_i||^2}{n}=o_{P|\mathcal{F}_n}. \tag{1.74}$$

Note that $\frac{1}{n}\sum_{i=1}^{n}||\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{MLE})-\boldsymbol{s}_i(\tilde{\boldsymbol{\theta}}_0)||^2=o_{P|\mathcal{F}_n}(1)$, the proof of 1.74 follows the same procedure as the proof of 1.62. Combining 1.60, 1.61 and 1.74 proves 1.59. Combining 1.58 and 1.59, we have that $||\boldsymbol{V}_0^{\tilde{\boldsymbol{\theta}}_0}-\boldsymbol{V}_0||=o_{P|\mathcal{F}_n}$. Therefore, the desired result follows from 1.54 and the fact that

$$\begin{aligned}
&\boldsymbol{V}^{-1/2}\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}(\boldsymbol{V}_0^{\tilde{\boldsymbol{\theta}}_0})^{1/2}\{\boldsymbol{V}^{-1/2}\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}(\boldsymbol{V}_0^{\tilde{\boldsymbol{\theta}}_0})^{1/2}\}^T\\
&=\boldsymbol{V}^{-1/2}\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}(\boldsymbol{V}_0^{\tilde{\boldsymbol{\theta}}_0})M(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{V}^{-1/2}\\
&=\boldsymbol{V}^{-1/2}\boldsymbol{M}(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{V}_0M(\hat{\boldsymbol{\theta}}_{MLE})^{-1}\boldsymbol{V}^{-1/2}+o_{P|\mathcal{F}_n}(r^{-1/2})\\
&=\boldsymbol{I}+o_{P|\mathcal{F}_n}(r^{-1/2}).
\end{aligned}\tag{1.75}$$

References

[1] Ferguson, T. S. *A Course in Large Sample Theory*, Chapman and Hall, 1996.
[2] van der Vaart, A. *Asymptotic Statistics*, Canbridge University Press, London.
[3] Tzavelas, G. *A note on the uniqueness of the quasi-likelihood estimator*, Statistics and Probability Letters, 38, pp. 125–130.