Supplementary Materials for “Model Selection of Generalized Estimating Equation with Divergent Model Size”

Shicheng Wu\textsuperscript{1}, Xin Gao\textsuperscript{1} and \textsuperscript{2}Raymond J. Carroll

\textsuperscript{1} York University and \textsuperscript{2} Texas A\&M University

Supplementary Material

This online supplementary file contains the proofs of Lemmas 1 - 8 in the main paper and some technical lemmas S2.1 - S2.8 and their proofs. In the following proofs, we assume $m_i = m$ for simplicity.

S1 Proofs of Lemmas 1 - 7

Proof of Lemma 1. This lemma is similar to Lemma A6 of Gao and Carroll (2017). Because $Q(\beta_s)$ satisfies the cumulant boundedness condition, its first and second moments are uniformly bounded. Given a model $s$, by Lemma A5 of Gao and Carroll (2017),

\[
\Pr(\sum_{i=1}^n [Q_i(\beta_s) - \mathbb{E}(Q_i(\beta_s))] / \text{Var}(Q_i(\beta_s)) > (2np_n \log p_n)^{1/2}) = o(p_n^{-p_n}).
\]

According to Bonferroni in-
equality,

\[ \Pr(\max_{s \in S} \sum_{i=1}^{n} [Q_i(\beta_s) - E(Q_i(\beta_s))] > b_{var}(2np_n \log p_n)^{1/2}) \leq o(p_n^{-p_n})2^{p_n} = 0, \]

as there are \(2^{p_n}\) models in the model space, and \(b_{var}\) is the upper bound for \(\text{Var}\{Q_i(\beta_s)\}\). Similar arguments apply to the result for each element of the score function, and its first and second derivatives.

\[ \square \]

**Proof of Lemma 2.** We know for true model \(V_i(\beta_s^*) = V_i(\beta_F^*)\) for \(s \in S_+\).

Therefore we know that

\[ ||\hat{V}_i^{-1} - V_i(\beta_s^*)^{-1}||_{\text{max}} = ||V_i(\hat{\beta}_F)^{-1} - V_i(\beta_s^*)^{-1}||_{\text{max}} = ||V_i(\hat{\beta}_F)^{-1} - V_i(\beta_s^*)^{-1}||_{\text{max}} = O_p\{(p_n^3 \log p_n/n)^{1/2}\}, \]

according to Lemma S2.5. In addition

\[ ||\hat{V}_i^{-1} - V_i(\beta_s^*)^{-1}||_{\text{max}} \leq ||\hat{V}_i^{-1} - V_i(\beta_s^*)^{-1}||_{\text{max}} + ||V_i(\beta_s^*)^{-1} - V_i(\hat{\beta}_s)^{-1}||_{\text{max}} = O_p\{(p_n^3 \log p_n/n)^{1/2}\}. \]

This implies:

\[ \max |\lambda_{\text{max}}(\hat{V}_i^{-1} - V_i(\beta_s^*)^{-1})|, \lambda_{\text{min}}(\hat{V}_i^{-1} - V_i(\beta_s^*)^{-1})| = O_p\{(p_n^3 \log p_n/n)^{1/2}\}, \]

\[ \max |\lambda_{\text{max}}(\hat{V}_i^{-1} - V_i(\beta_s^*)^{-1})|, \lambda_{\text{min}}(\hat{V}_i^{-1} - V_i(\beta_s^*)^{-1})| = O_p\{(p_n^3 \log p_n/n)^{1/2}\}. \]

\[ \square \]

**Proof of Lemma 3.** From Taylor expansion around \(\beta_s^*\), there exists \(\tilde{\beta}_s\) between \(\beta_s^*\) and \(\hat{\beta}_s\) such that \((1/n)U(\tilde{\beta}_s)_{[r]} = 0\). Therefore we know

\[ \frac{1}{n} U(\beta_s^*)_{[r]} + \sum_k \frac{1}{n} U(\beta_s^*)_{[rk]}(\hat{\beta}_s - \beta_s^*)_{[k]} + \sum_{k,l} \frac{1}{n} U(\hat{\beta}_s)_{[rkl]}(\hat{\beta}_s - \beta_s^*)_{[k]}(\hat{\beta}_s - \beta_s^*)_{[l]} = 0. \]
According to Lemma 1, \( \max_{s \in S} |(1/n)U(\beta^*_s)_{[rk]} + \Omega(\beta^*_s)_{[rk]}| = \max_{s \in S} |(1/n)U(\beta^*_s)_{[rk]} - E[U(\beta^*_s)_{[rk]}]| = O_p\{p_n \log p_n/n^{1/2}\} \), then we have

\[
\sum_k \frac{1}{n} U(\beta^*_s)_{[rk]}(\hat{\beta}_s - \beta^*_s)_{[k]}
= \sum_k [-\Omega(\beta^*_s)_{[rk]} + \{\Omega(\beta^*_s)_{[rk]} + \frac{1}{n} U(\beta^*_s)_{[rk]}\}](\hat{\beta}_s - \beta^*_s)_{[k]}
= \sum_k [-\Omega(\beta^*_s)_{[rk]} + O_p\{(p_n \log p_n/n)^{1/2}\}](\hat{\beta}_s - \beta^*_s)_{[k]}.
\]

Similarly from Lemma 1, \( n^{-1}U(\tilde{\beta}_s)_{[rkl]} = n^{-1}E[U(\tilde{\beta}_s)_{[rkl]}]\{1 + o_p(1)\} \).

From Assumption 4, \( n^{-1}E[U(\tilde{\beta}_s)_{[rkl]}] \) is bounded. So \( n^{-1}U(\tilde{\beta}_s)_{[rkl]} = O_p(1) \).

According to Theorem 1, \( ||\hat{\beta}_s - \beta^*_s|| = O_p\{(p_n^2 \log p_n/n)^{1/2}\} \). Then

\[
\left| \sum_l n^{-1}U(\beta^*_s)_{[rk]}(\hat{\beta}_s - \beta^*_s)_{[l]} \right|
\leq \max_l \left| n^{-1}U(\beta^*_s)_{[rk]} \right| \sum_l |(\hat{\beta}_s - \beta^*_s)_{[l]}|
= O_p(1) \times \sum_l |(\hat{\beta}_s - \beta^*_s)_{[l]}|
\leq O_p(1) \times d_s^{1/2} \times ||\hat{\beta}_s - \beta^*_s||
= O_p\{(p_n^3 \log p_n/n)^{1/2}\}.
\]

Combining the second and the third order terms of Taylor expansion,
we have
\[
0 = \frac{1}{n} U(\tilde{\beta}_s)[r]
\]
\[
= \frac{1}{n} U(\beta^*_s)[r] - \sum_k \{\Omega(\beta^*_s)[r_k] + O_p\{p_n \log p_n / n\}^{1/2}\}(\tilde{\beta}_s - \beta^*_s)[k]
\]
\[
+ \sum_k O_p\{(p_n^3 \log p_n / n)^{1/2}\}(\tilde{\beta}_s - \beta^*_s)[k].
\]

We can reformat it as
\[
\frac{1}{n} U(\beta^*_s) - \{\Omega(\beta^*_s) + Res_d\}(\tilde{\beta}_s - \beta^*_s) = 0,
\]
where $Res_d$ is a matrix that all elements are at order of $O_p\{(p_n^3 \log p_n / n)^{1/2}\}$ uniformly. Let $v_{\min}$ be the corresponding eigenvector of smallest eigenvalue $\lambda_{\min}\{\Omega(\beta^*_s)\}$. According to matrix perturbation theory ([Stewart](1990)), we have
\[
\lambda_{\min}\{\Omega(\beta^*_s) + Res_d\}
\]
\[
= \lambda_{\min}\{\Omega(\beta^*_s)\} + v_{\min}^T Res_d v_{\min} + o(||Res_d||^2)
\]
\[
\geq \lambda_{\min}\{\Omega(\beta^*_s)\} + d_s \times ||Res_d||_{\max} + o(1)
\]
\[
= \lambda_{\min}\{\Omega(\beta^*_s)\} + O_p\{(p_n^5 \log p_n / n)^{1/2}\} + o(1)
\]
Since $\lambda_{\min}\{\Omega(\beta^*_s)\} > 0$ and $p_n^5 \log p_n / n \to 0$, we have $\lambda_{\min}\{\Omega(\beta^*_s) + Res_d\} > 0$ and therefore $\Omega(\beta^*_s) + Res_d$ is invertible. This entails
\[
\tilde{\beta}_s - \beta^*_s = \frac{1}{n}\{\Omega(\beta^*_s) + Res_d\}^{-1}U(\beta^*_s).
\]
Proof of Lemma 4. Considering a competing model $s$.

$$2Q(\beta_s^*) = \sum_{i=1}^{n} \{Y_i - \mu_i(\beta_s^*)\}^T \hat{V}_i^{-1} \{Y_i - \mu_i(\beta_s^*)\}$$

$$= \sum_{i=1}^{n} \{Y_i - \mu_i(\hat{\beta}_s) + \mu_i(\beta_s^*)\}^T \hat{V}_i^{-1} \{Y_i - \mu_i(\hat{\beta}_s) + \mu_i(\beta_s^*)\}$$

$$= 2Q(\hat{\beta}_s) + \sum_{i=1}^{n} \{\mu_i(\beta_s^*) - \mu_i(\hat{\beta}_s)\}^T \hat{V}_i^{-1} \{\mu_i(\beta_s^*) - \mu_i(\hat{\beta}_s)\}$$

$$+ 2 \sum_{i=1}^{n} \{\mu_i(\hat{\beta}_s) - \mu_i(\beta_s^*)\}^T \hat{V}_i^{-1} \{Y_i - \mu_i(\hat{\beta}_s)\}.$$ 

Lemma S2.7 shows that the last term $\sum_{i=1}^{n} \{\mu_i(\hat{\beta}_s) - \mu_i(\beta_s^*)\}^T \hat{V}_i^{-1} \{Y_i - \mu_i(\hat{\beta}_s)\} = n\|\beta_s^* - \hat{\beta}_s\|^2 o_p(1)$. We consider the second term. Applying Equation (S2.1) from Lemma S2.3 to the second term, we have

$$\sum_{i=1}^{n} \{\mu_i(\beta_s^*) - \mu_i(\hat{\beta}_s)\}^T \hat{V}_i^{-1} \{\mu_i(\beta_s^*) - \mu_i(\hat{\beta}_s)\}$$

$$= \sum_{i=1}^{n} (\beta_s^* - \hat{\beta}_s)^T \{D_i(\beta_s^*) + \frac{1}{2} D_i^{(1)}(\hat{\beta}_s, \beta_s^*, B_s^\mu)\}^T \hat{V}_i^{-1} \{D_i(\beta_s^*) + \frac{1}{2} D_i^{(1)}(\hat{\beta}_s, \beta_s^*, B_s^\mu)\}(\beta_s^* - \hat{\beta}_s)$$

$$= (\beta_s^* - \hat{\beta}_s)^T \{\sum_{i=1}^{n} D_i(\beta_s^*)^T \hat{V}_i^{-1} D_i(\beta_s^*) + \frac{1}{2} D_i^{(1)}(\hat{\beta}_s, \beta_s^*, B_s^\mu) \hat{V}_i^{-1} D_i(\beta_s^*) + \frac{1}{4} D_i^{(1)}(\hat{\beta}_s, \beta_s^*, B_s^\mu) \hat{V}_i^{-1} D_i^{(1)}(\hat{\beta}_s, \beta_s^*, B_s^\mu)\}(\beta_s^* - \hat{\beta}_s)$$

$$= n(\beta_s^* - \hat{\beta}_s)^T \{\Omega(\beta_s^*) + \frac{1}{n} \sum_{i=1}^{n} D_i(\beta_s^*)^T \{\hat{V}_i^{-1} - V_i(\beta_s^*)\} D_i(\beta_s^*) + D_i^{(1)}(\hat{\beta}_s, \beta_s^*, B_s^\mu) \hat{V}_i^{-1} D_i(\beta_s^*)$$

$$+ \frac{1}{4} D_i^{(1)}(\hat{\beta}_s, \beta_s^*, B_s^\mu) \hat{V}_i^{-1} D_i^{(1)}(\hat{\beta}_s, \beta_s^*, B_s^\mu)\}(\beta_s^* - \hat{\beta}_s)$$

$$= n(\beta_s^* - \hat{\beta}_s)^T \{\Omega(\beta_s^*) + \text{Res}_3\}(\beta_s^* - \hat{\beta}_s).$$

Let $\text{Res}_3 = \text{Res}_{31} + \text{Res}_{32} + \text{Res}_{33}$, with $\text{Res}_{31} = \sum_{i=1}^{n} D_i(\beta_s^*)^T \{\hat{V}_i^{-1} - V_i(\beta_s^*)\} D_i(\beta_s^*)/n$, $\text{Res}_{32} = \sum_{i=1}^{n} D_i^{(1)}(\hat{\beta}_s, \beta_s^*, B_s^\mu) \hat{V}_i^{-1} D_i(\beta_s^*)/n$, and $\text{Res}_{33} = 5$. 


\[
\sum_{i=1}^{n} D_i(1) (\hat{\beta}_s, \beta_s^*, B_{s}^{\mu_i})^T \hat{V}_{i}^{-1} D_i(1) (\hat{\beta}_s, \beta_s^*, B_{s}^{\mu_i}) / 4n.
\]
Let \(v\) be a \(d_s\) dimensional unit vector with \(||v||^2 = 1\). Then

\[
|v^T \text{Res}_{33} v| = |v^T \frac{1}{n} \sum_{i=1}^{n} D_i(\beta_s^*)^T \{\hat{V}_{i}^{-1} - V_i(\beta_s^*)\} D_i(\beta_s^*) v|
\]

\[
\leq \max \{ |\lambda_{\max} \{\hat{V}_{i}^{-1} - V_i(\beta_s^*)\} |, |\lambda_{\min} \{\hat{V}_{i}^{-1} - V_i(\beta_s^*)\} | \} \max \{ v^T \frac{1}{n} \sum_{i=1}^{n} D_i(\beta_s^*)^T D_i(\beta_s^*) v \}
\]

\[
\leq O_p \{(p_n^3 \log p_n / n)^{1/2} \}.
\]

From Lemma \([\text{S2.4}]\) we have

\[
|v^T \text{Res}_{32} v| = |v^T (1/n) \sum_{i=1}^{n} D_i(1) (\hat{\beta}_s, \beta_s^*, B_{s}^{\mu_i})^T V_i^{-1} D_i(\beta_s^*) v| = O_p \{(p_n^3 \log p_n / n)^{1/2} \}.
\]

For \(\text{Res}_{33}\),

\[
|v^T \text{Res}_{33} v| = |v^T \frac{1}{n} \sum_{i=1}^{n} D_i(1) (\hat{\beta}_s, \beta_s^*, B_{s}^{\mu_i})^T V_i^{-1} D_i(1) (\hat{\beta}_s, \beta_s^*, B_{s}^{\mu_i})^T v|
\]

\[
\leq \lambda_{\max} (V_i^{-1}) \max \{ v^T \frac{1}{n} \sum_{i=1}^{n} D_i(1) (\hat{\beta}_s, \beta_s^*, B_{s}^{\mu_i})^T D_i(1) (\hat{\beta}_s, \beta_s^*, B_{s}^{\mu_i}) v \}
\]

\[
= O_p \{p_n^3 \log p_n / n \}.
\]

From Assumption \([2]\) the eigenvalues of \(\Omega(\beta_s^*)\) are bounded from zero and infinity. We have

\[
\sup_{||v||=1} |v^T (\Omega(\beta_s^*) + \text{Res}_3) v| = \sup_{||v||=1} |v^T \Omega(\beta_s^*) v| (1 + O_p \{(p_n^3 \log p_n / n)^{1/2} \}).
\]

Combining the above equation and Lemma \([\text{S2.7}]\) we have

\[
2[Q(\hat{\beta}_s) - Q(\beta_s^*)] = -n(\beta_s^* - \hat{\beta}_s)^T (\Omega(\beta_s^*) + \text{Res}_3) (\beta_s^* - \hat{\beta}_s) + n||\beta_s^* - \hat{\beta}_s||^2 O_p \{(p_n^3 \log p_n / n)^{1/2} \} = -n(\beta_s^* - \hat{\beta}_s)^T \Omega(\beta_s^*) (\beta_s^* - \hat{\beta}_s) (1 + o_p(1)).
\]

According to Lemma \([3]\), \(\hat{\beta}_s - \beta_s^* = \)
\[
\{\Omega(\beta_s^*) + \text{Res}_d\}^{-1}U(\beta_s^*). \text{ We rewrite the equation as }
\]
\[
2\{Q(\hat{\beta}_s) - Q(\beta_s^*)\} = -n(\beta_s^* - \hat{\beta}_s)^T \Omega(\beta_s^*) \{\Omega(\beta_s^*) + \text{Res}_d\}^{-1}U(\beta_s^*) \{1 + o_p(1)\}
\]
\[
= -\frac{1}{n}U(\beta_s^*)^T \Omega(\beta_s^*) + \text{Res}_d^T \Omega(\beta_s^*)^{-1} \{\Omega(\beta_s^*) + \text{Res}_d\}^{-1}U(\beta_s^*) \{1 + o_p(1)\}.
\]
\[
= -\frac{1}{n}U(\beta_s^*)^T \{\{\Omega(\beta_s^*) + \text{Res}_d^T\} \Omega(\beta_s^*)^{-1} \{\Omega(\beta_s^*) + \text{Res}_d\}^{-1}U(\beta_s^*) \{1 + o_p(1)\}.
\]
\[
= -\frac{1}{n}U(\beta_s^*)^T \{\Omega(\beta_s^*) + \text{Res}_d + \text{Res}_d^T \Omega(\beta_s^*)^{-1} \text{Res}_d\}^{-1}U(\beta_s^*) \{1 + o_p(1)\}.
\]

Let \(\text{Res}_s = \text{Res}_d + \text{Res}_d^T \Omega(\beta_s^*)^{-1} \text{Res}_d\) and we have
\[
2\{Q(\hat{\beta}_s) - Q(\beta_s^*)\} = -1/nU(\beta_s^*)^T \{\Omega(\beta_s^*) + \text{Res}_s\}^{-1}U(\beta_s^*) \{1 + o_p(1)\}.
\]

We estimate the order of the matrix \(\text{Res}_s\) as follows:
\[
\sup_{||v||=1} v^T (\text{Res}_d + \text{Res}_d^T) v \leq 2 \sup_{||v||=1} v^T \text{Res}_d v
\]
\[
\leq \max_{k,r} \|[\text{Res}_d]_{kr}\| \times d_s \times ||v||^2
\]
\[
= O_p \{(p_n^5 \log p_n)^{1/2}\};
\]
\[
\inf_{||v||=1} v^T (\text{Res}_d + \text{Res}_d^T) v \geq 2 \inf_{||v||=1} v^T \text{Res}_d v
\]
\[
\geq -\max_{k,r} \|[\text{Res}_d]_{kr}\| \times d_s \times ||v||^2
\]
\[
= -O_p \{(p_n^5 \log p_n)^{1/2}\};
\]
\[
\sup_{||v||=1} v^T (\text{Res}_d^T \Omega(\beta_s^*)^{-1} \text{Res}_d) v
\]
\[
\leq \lambda_{\max} \{\Omega(\beta_s^*)^{-1}\} d_s \times ||\text{Res}_d||^2_{\max} \times d_s \times ||v||^2
\]
\[
= O_p \{p_n^5 \log p_n / n\};
\]
7
Thus we have $\inf_{||v||=1} v^T (Res_d^T \Omega(\beta_s^*)^{-1} Res_d) v$

\[ \geq -\lambda_{\text{min}} \{ \Omega(\beta_s^*)^{-1} \} d_s \times ||Res_d||_{\text{max}}^2 \times d_s \times ||v||^2 \]

\[ = -O_p \{ p_n^5 \log p_n / n \}. \]

Thus we have $\sup_{||v||=1} |v^T Res_s v| = O_p \{ (p_n^5 \log p_n / n)^{1/2} \} = o_p(1)$. This implies that the eigenvalues of $Res_s$ are of the order of $o_p(1)$. It follows that

\[ \sup_{||v||^2=1} v^T [\Omega(\beta_s^*)^{-1} - \{ \Omega(\beta_s^*) + Res_s \}^{-1}] v \]

\[ = \sup_{||v||^2=1} v^T [\Omega(\beta_s^*)^{-1/2} I - \{ I + \Omega(\beta_s^*)^{-1/2} Res_s \Omega(\beta_s^*)^{-1/2} \}^{-1}] \Omega(\beta_s^*)^{-1/2} v \]

\[ \leq \lambda_{\text{max}} \{ \Omega(\beta_s^*)^{-1} \} \lambda_{\text{max}} (I - \{ I + \Omega(\beta_s^*)^{-1/2} Res_s \Omega(\beta_s^*)^{-1/2} \}^{-1}) ||v||^2 \]

\[ = \lambda_{\text{max}} \{ \Omega(\beta_s^*)^{-1} \} (1 - \lambda_{\text{min}}[\{ I + \Omega(\beta_s^*)^{-1/2} Res_s \Omega(\beta_s^*)^{-1/2} \}^{-1}]) \]

\[ = \lambda_{\text{max}} \{ \Omega(\beta_s^*)^{-1} \} \frac{\lambda_{\text{max}} \{ \Omega(\beta_s^*)^{-1/2} Res_s \Omega(\beta_s^*)^{-1/2} \}}{1 + \lambda_{\text{max}} \{ \Omega(\beta_s^*)^{-1/2} Res_s \Omega(\beta_s^*)^{-1/2} \} }. \]

Furthermore,

\[ \lambda_{\text{max}} \{ \Omega(\beta_s^*)^{-1/2} Res_s \Omega(\beta_s^*)^{-1/2} \} \]

\[ = \sup_{||v||=1} v^T \{ \Omega(\beta_s^*)^{-1/2} Res_s \Omega(\beta_s^*)^{-1/2} \} v \]

\[ \leq \lambda_{\text{max}} \{ \Omega(\beta_s^*)^{-1} \} \lambda_{\text{max}} \{ Res_s \} ||v||^2 = o_p(1) \]

Thus $\sup_{||v||^2=1} v^T [\Omega(\beta_s^*)^{-1} - \{ \Omega(\beta_s^*) + Res_s \}^{-1}] v = o_p(1)$. Therefore,

\[ 2 \{ Q(\beta_s) - Q(\beta_s^*) \} = -1/n U(\beta_s^*)^T \{ \Omega(\beta_s^*) + Res_s \}^{-1} U(\beta_s^*) \{ 1 + o_p(1) \} \]

\[ = - \frac{1}{n} U(\beta_s^*)^T \Omega(\beta_s^*)^{-1} U(\beta_s^*) \{ 1 + o_p(1) \}. \]
Proof of Lemma 4. We first consider the true and overfitting situation. By Lemma 4 shows that \(|Q(\hat{\beta}_s) - Q(\beta_s^*)| = (n/2)(\beta_s^* - \hat{\beta}_s)^T \Omega(\beta_s^*)(\beta_s^* - \hat{\beta}_s)\{1 + o_p(1)\}\). Theorem 1 shows that \(||\beta_s^* - \hat{\beta}_s|| = O_p\{(p_n^2 \log p_n/n)^{1/2}\}\). And Assumption 2 indicates that all eigenvalue of \(\Omega(\beta_s^*)\) is bounded.

\[
|Q(\hat{\beta}_s) - Q(\beta_s^*)| = \frac{n}{2}(\beta_s^* - \hat{\beta}_s)^T \Omega(\beta_s^*)(\beta_s^* - \hat{\beta}_s)\{1 + o_p(1)\} \\
\leq \frac{n}{2}\lambda_{\max}\{\Omega(\beta_s^*)\}||\beta_s^* - \hat{\beta}_s||^2\{1 + o_p(1)\} \\
= O_p(p_n^2 \log p_n).
\]

Then we consider the underfitting situation.

\[
|2Q(\hat{\beta}_s) - 2Q(\beta_s^*)| = |\sum_{i=1}^{n}\{Y_i - \mu_i(\hat{\beta}_s)\}^T \hat{V}_i^{-1}\{Y_i - \mu_i(\hat{\beta}_s)\} - 2Q(\beta_s^*)| \\
= |\sum_{i=1}^{n}\{Y_i - \mu_i(\beta_s^*) + \mu_i(\beta_s^*) - \mu_i(\hat{\beta}_s)\}^T \hat{V}_i^{-1}\{Y_i - \mu_i(\beta_s^*) + \mu_i(\beta_s^*) - \mu_i(\hat{\beta}_s)\} - 2Q(\beta_s^*)| \\
\leq |\sum_{i=1}^{n}\{\mu_i(\beta_s^*) - \mu_i(\hat{\beta}_s)\}^T \hat{V}_i^{-1}\{\mu_i(\beta_s^*) - \mu_i(\hat{\beta}_s)\}| + 2|\sum_{i=1}^{n}\{\mu_i(\beta_s^*) - \mu_i(\hat{\beta}_s)\}^T \hat{V}_i^{-1}\{Y_i - \mu_i(\beta_s^*)\}|.
\]

We consider the first term from the above formula. According to Taylor expansion, there exists a \(\hat{\beta}_s\) between \(\beta_s^*\) and \(\hat{\beta}_s\) such that \(\mu_i(\beta_s^*) - \mu_i(\hat{\beta}_s) = D_i(\hat{\beta})(\beta_s^* - \hat{\beta}_s)\). Then we have

\[
|\sum_{i=1}^{n}\{\mu_i(\beta_s^*) - \mu_i(\hat{\beta}_s)\}^T \hat{V}_i^{-1}\{\mu_i(\beta_s^*) - \mu_i(\hat{\beta}_s)\}| \\
= \sum_{i=1}^{n}(\beta_s^* - \hat{\beta}_s)^TD_i(\hat{\beta})^T \hat{V}_i^{-1}D_i(\hat{\beta})(\beta_s^* - \hat{\beta}_s) \\
\leq n\lambda_{\max}\{\hat{V}_i^{-1}\} \times ||\beta_s^* - \hat{\beta}_s||^2 \times \max_{||v||^2 = 1}\{v^T \frac{1}{n}\sum_{i=1}^{n}D_i(\hat{\beta})^T D_i(\hat{\beta})v\} \\
= O_p(p_n^2 \log p_n).
\]
Next we consider the second term. Lemma S2.6 implies that \( \max_j \{(1/n) \sum_{i=1}^n |Y_{ij} - \mu_{ij}(\beta^*_s)|\} = O_p(1) \). Assumption 3 implies that \( \max_i \{||D_i(\beta)\||_{\max}||\hat{V}_i^{-1}||_{\max}\} = O_p(1) \). Combining these results, we have

\[
\left| \sum_{i=1}^n \{\mu_i(\beta^*_s) - \mu_i(\hat{\beta}_s)\}^T \hat{V}_i^{-1} \{Y_i - \mu_i(\beta^*_s)\} \right| \\
= \left| \sum_{i=1}^n (\beta^*_s - \hat{\beta}_s)^T D_i(\hat{\beta})^T \hat{V}_i^{-1} \{Y_i - \mu_i(\beta^*_s)\} \right| \\
\leq ||\beta^*_s - \hat{\beta}_s|| \times \left| \sum_{i=1}^n D_i(\hat{\beta})^T \hat{V}_i^{-1} \{Y_i - \mu_i(\beta^*_s)\} \right| \\
\leq ||\beta^*_s - \hat{\beta}_s|| \times p_n^{1/2} \max_k \left| \sum_{i=1}^n [D_i(\hat{\beta})^T]_{[k,i]} \hat{V}_i^{-1} \{Y_i - \mu_i(\beta^*_s)\} \right| \\
\leq n p_n^{1/2} ||\beta^*_s - \hat{\beta}_s|| \times \frac{1}{n} \sum_{i=1}^n m^2 ||D_i(\hat{\beta})||_{\max} ||\hat{V}_i^{-1}||_{\max} \times \max_j |Y_{ij} - \mu_{ij}(\beta^*_s)| \\
\leq n p_n^{1/2} m^2 ||\beta^*_s - \hat{\beta}_s|| \times \max \{||D_i(\hat{\beta})||_{\max} ||\hat{V}_i^{-1}||_{\max}\} \times \max_j \left\{ \frac{1}{n} \sum_{i=1}^n |Y_{ij} - \mu_{ij}(\beta^*_s)| \right\} \\
= O_p\{ (p_n^3 \log p_n/n)^{1/2} \},
\]

\( \square \)

**Proof of Lemma**. From Assumption 2, both \( \Omega(\beta^*_T)^{-1} \) and \( \Omega(\beta^*_s)^{-1} \) are positive definite. The \( \Omega(\beta^*_T) \) is a sub-block of \( \Omega(\beta^*_s) \). We define \( \Omega = \Omega(\beta^*_T) \) and define block matrix \( \Omega(\beta^*_s) = \\
\begin{bmatrix} \Omega & \hat{\Omega} \\ \hat{\Omega}^T & \tilde{\Omega} \end{bmatrix} \), where \( \tilde{\Omega} \) is a positive definite \( d_s \times d_s \) matrix and \( \tilde{\Omega} \) a \( d_s \times (d_s - d_T) \) matrix. For any \( d_T \times 1 \) vector \( \nu_1 \) and \( (d_s - d_T) \times 1 \) vector \( \nu_2 \), we can show that the matrix \( M_{s/T} \) is non-negative.
definite by the formula below:

$$
\begin{bmatrix}
\eta_1 \\
\eta_2
\end{bmatrix}^T 
\begin{bmatrix}
\Omega & \tilde{\Omega}
\end{bmatrix}^{-1} 
\begin{bmatrix}
\Omega^{-1} 0 \\
0 0
\end{bmatrix} 
\begin{bmatrix}
\eta_1 \\
\eta_2
\end{bmatrix} 
= 
\begin{bmatrix}
\eta_1 \\
\eta_2
\end{bmatrix}^T 
\begin{bmatrix}
\Omega^{-1}\tilde{\Omega}(\tilde{\Omega} - \tilde{\Omega}^T\Omega^{-1}\tilde{\Omega})^{-1}\tilde{\Omega}^T\Omega^{-1} - \Omega^{-1}\tilde{\Omega}(\tilde{\Omega} - \tilde{\Omega}^T\Omega^{-1}\tilde{\Omega})^{-1} \\
-(\tilde{\Omega} - \tilde{\Omega}^T\Omega^{-1}\tilde{\Omega})^{-1}\tilde{\Omega}^T\Omega^{-1} \\
(\tilde{\Omega} - \tilde{\Omega}^T\Omega^{-1}\tilde{\Omega})^{-1}
\end{bmatrix} 
\begin{bmatrix}
\eta_1 \\
\eta_2
\end{bmatrix} 
= (\tilde{\Omega}^T\Omega^{-1}\eta_1 - \eta_2)^T (\tilde{\Omega} - \tilde{\Omega}^T\Omega^{-1}\tilde{\Omega})^{-1}(\tilde{\Omega}^T\Omega^{-1}\eta_1 - \eta_2) 
\geq 0.
$$

Proof of Lemma 7. Let \( \eta_s = n^{-1/2}W^{-1/2}(\beta^*_s)\) and \( U(\beta^*_s) \). According to Lemma S2.8, it satisfies the exponential moment condition, 

$$
\log[\mathbb{E}\{\exp(t^T\eta_s)\}] \leq a^2\|t\|^2/2,
$$

with \( t \in \mathbb{R}^{d_s}, \|t\|^2 \leq p_n^2 \log p_n \) and some constant \( a^2 > 1 \). We scale the vector \( \eta \) as \( \eta^* = \eta/a \), so that \( \log[\mathbb{E}\{\exp(t^T\eta^*)\}] \leq \|t\|^2/2 \) with \( \|t\| \leq \{a^2p_n^2(\log p_n)\}^{1/2} = a\times \rho \). Given matrix \( B_{s/T} = W^{1/2}(\beta^*_s)M_{s/T}(\beta^*_s)W^{1/2}(\beta^*_s) \)
and \( \text{Tr}(B_{s/T}) = d^*_s - d^{*T} \), we define \( B^*_{s/T} = B_{s/T}/\tau \) where \( \tau = \lambda_{\max}(B_{s/T}) \). Therefore \( \lambda_{\max}(B^*_{s/T}) = 1 \). We scale the quadratic form \( \Delta^*_s = \Delta_s/a^2\tau = (\eta^*)^TB^*_s\eta^* \). Let \( \Delta_{s/T} = \eta^TB\eta \) and \( \Delta^*_{s/T} = \Delta/a^2\tau = (\eta^*)B^*\eta^* \). Define \( P_G = \text{Tr}[B^*] = (d^*_s - d^{*T})/\tau \) and \( V^2_G = \text{Tr}[(B^*)^2] \). Using the inequality for the trace of matrix product, we obtain \( V_G \leq (2P_G)^{1/2} = O(p_n) \). We
apply the large deviation result from Corollary 4.2 of Spokoiny and Zhilova (2013). For \(3/2 \rho^2 > K > V_G/3\),

\[
\Pr(\Delta_{s/T}^* > P_G + K) \leq 10.4 \exp(-K/6).
\]

Choosing \(K = \{(d_s^* - d_T^*)/\tau\}\{\gamma_n/a^2 - 1\}\) leads to \(K = O(p_n \log p_n)\). Given \(\rho^2 = O(p_n^2 \log p_n)\) and \(V_G = O(p_n)\). We have \(3/2 \rho^2 > K > V_G/3\). Let \(\tilde{\tau} = (d_s^* - d_T^*)/(d_s - d_T)\). We have

\[
\Pr\{\max_{s \in S_+, s \neq T} \Delta_{s/T} > (d_s^* - d_T^*)\gamma_n\} \\
\leq \sum_{s \in S_+, s \neq T} \Pr\{\Delta_{s/T}^* > [(d_s^* - d_T^*)\gamma_n/(a^2 \tilde{\tau})]\} \\
\leq \sum_{s \in S_+, s \neq T} \Pr\{\Delta_{s/T}^* > P_G + P_G(\frac{\gamma_n}{a^2} - 1)\} \\
\leq \sum_{s \in S_+, s \neq T} \Pr\{\Delta_{s/T}^* > P_G + K\} \\
\leq \sum_{s \in S_+, s \neq T} 10.4 \exp\{-\frac{(d_s - d_T)\tilde{\tau}}{6\tilde{\tau}}(\frac{\gamma_n}{a^2} - 1)\} \\
\leq \sum_{d_s = d_T + 1}^{p_n} C_{p_n-d_T}^{d_s-d_T} 10.4 \exp\{-\frac{(d_s - d_T)\tilde{\tau}}{6\tilde{\tau}}(\frac{\gamma_n}{a^2} - 1)\} \\
\leq \sum_{m' = 1}^{p_n - d_T} C_{p_n-d_T}^{m'} 10.4 \exp\{-\frac{m'}{6\omega}(\frac{\gamma_n}{a^2} - 1)\} \\
\leq \{1 + 10.4 \exp\{-\frac{\gamma_n}{a^2} - 1\}\}^{p_n-d_T} - 1.
\]

As \(a^2\) can be chosen as close to 1 as possible with increasing sample size \(n\), the choices of \(\gamma_n = 6\omega(1 + \gamma) \log p_n\) for some \(\gamma > 0\) or \(\gamma_n = 6\omega(1 + \log \log p_n) \log p_n\) lead to \(\lim_{n \to \infty} (1 + 10.4 \exp\{-\frac{\gamma_n}{a^2} - 1\}/(6\omega))^{p_n-d_T} = 1\).
This entails \( \Pr \{ \max_{s \in S, s \neq T} \Delta_{s/T} > (d_s^* - d_T^*)/\gamma_n \} \to 0. \)

**Proof of Lemma 8.** First we decompose the difference between \( d_s^* \) and \( \hat{d}_s \) as follows:

\[
d_s^* - \hat{d}_s = \text{Tr} \{ W_s(\beta_s^*)\Omega_s^{-1}(\beta_s^*) - W_s(\beta_s)\Omega_s^{-1}(\beta_s) \}
\]

\[
= \text{Tr} \{ W_s(\beta_s^*) - W_s(\beta_s) \{ \Omega_s^{-1}(\beta_s^*) - \Omega_s^{-1}(\beta_s) \} \}
\]

\[
+ \text{Tr} \{ W_s(\beta_s) \{ \Omega_s^{-1}(\beta_s^*) - \Omega_s^{-1}(\beta_s) \} \}
\]

\[
+ \text{Tr} \{ W_s(\beta_s^*) - W_s(\beta_s) \} \Omega_s^{-1}(\beta_s)\]

Next we will prove that \( |\text{Tr} \{ W_s(\beta_s^*) - W_s(\beta_s)\} \Omega_s^{-1}(\beta_s)\} = O_p \{ (p^5 \log p_n/n)^{1/2} \}. \)

Let the subscript \([j, k]\) denote the \((j, k)\)th element of a matrix. The covariance matrix of the score vector can be expressed as

\[
W_s(\beta)_{jk} = n^{-1} \text{Cov} \{ U(\beta)_{[j]}, U(\beta)_{[k]} \}
\]

\[
= n^{-1} \sum_{i=1}^{n} \text{Cov} \{ U_i(\beta)_{[j]}, U_i(\beta)_{[k]} \}
\]

\[
= n^{-1} \sum_{i=1}^{n} [ E \{ U_i(\beta)_{[j]} U_i(\beta)_{[k]} \} - E \{ U_i(\beta)_{[j]} \} E \{ U_i(\beta)_{[k]} \} ]
\]

Then we have

\[
\{ W_s(\beta_s^*) - W_s(\beta_s) \}_{jk} = n^{-1} \sum_{i=1}^{n} [ E \{ U_i(\beta_s^*)_{[j]} U_i(\beta_s^*)_{[k]} - U_i(\beta_s)_{[j]} U_i(\beta_s)_{[k]} \}
\]

\[
- n^{-1} \sum_{i=1}^{n} [ E \{ U_i(\beta_s^*)_{[j]} \} E \{ U_i(\beta_s^*)_{[k]} \} - E \{ U_i(\beta_s)_{[j]} \} E \{ U_i(\beta_s)_{[k]} \} ]
\]
For the first component, we have

\[
U_i(\beta_s^*)_{[j]}U_i(\beta_s^*)_{[k]} - U_i(\tilde{\beta}_s)_{[j]}U_i(\tilde{\beta}_s)_{[k]}
= U_i(\tilde{\beta}_s)_{[j]} \{U_i(\beta_s^*)_{[k]} - U_i(\tilde{\beta}_s)_{[k]}\} + \{U_i(\beta_s^*)_{[j]} - U_i(\tilde{\beta}_s)_{[j]}\}U_i(\tilde{\beta}_s)_{[k]}
+ \{U_i(\beta_s^*)_{[j]} - U_i(\tilde{\beta}_s)_{[j]}\}\{U_i(\beta_s^*)_{[k]} - U_i(\tilde{\beta}_s)_{[k]}\}.
\]

From Taylor expansion, there exist a \(\tilde{\beta}_s\) between \(\beta_s^*\) and \(\tilde{\beta}_s\) such that

\[
|U_i(\beta_s^*)_{[j]} - U_i(\tilde{\beta}_s)_{[j]}| \times |U_i(\beta_s^*)_{[k]}| = |(\beta_s^* - \tilde{\beta}_s)^T U_i(\tilde{\beta}_s)_{[j]}| \times |U_i(\beta_s^*)_{[k]}| \\
\leq ||\beta_s^* - \tilde{\beta}_s|| \times |U_i(\tilde{\beta}_s)_{[j]}| \times |U_i(\beta_s^*)_{[k]}| \\
= O_p\{(p_n^3 \log p_n/n)^{1/2}\},
\]

where \(||\beta_s^* - \tilde{\beta}_s|| = O_p\{(p_n^3 \log p_n/n)^{1/2}\}\) according to Theorem 1 and \(|U_i(\beta_s^*)_{[k]}|\) is bounded and each element of \(p_n\) vector \(U_i(\tilde{\beta}_s)_{[j]}\) is bounded according to Lemma S2.2. Similarly we get \(\{U_i(\beta_s^*)_{[j]} - U_i(\tilde{\beta}_s)_{[j]}\}U_i(\tilde{\beta}_s)_{[k]} = O_p\{(p_n^3 \log p_n/n)^{1/2}\}\) and \(\{U_i(\beta_s^*)_{[j]} - U_i(\tilde{\beta}_s)_{[j]}\}\{U_i(\beta_s^*)_{[k]} - U_i(\tilde{\beta}_s)_{[k]}\} = O_p\{(p_n^3 \log p_n/n)^{1/2}\}\). Thus \(|U_i(\beta_s^*)_{[j]}U_i(\beta_s^*)_{[k]} - U_i(\tilde{\beta}_s)_{[j]}U_i(\tilde{\beta}_s)_{[k]}| = O_p\{(p_n^3 \log p_n/n)^{1/2}\}\), and \(1/n\sum_{i=1}^n|E\{U_i(\beta_s^*)_{[j]}\}E\{U_i(\beta_s^*)_{[k]}\} - E\{U_i(\tilde{\beta}_s)_{[j]}\}E\{U_i(\tilde{\beta}_s)_{[k]}\}| = O_p\{(p_n^3 \log p_n/n)^{1/2}\}\). Combining the terms above, we have \(\max_{j,k}\{W_s(\beta_s^*) - W_s(\tilde{\beta}_s)\}_{[j,k]} = O_p\{(p_n^3 \log p_n/n)^{1/2}\}\). Based on Assumption 2, all eigenvalues of matrices \(W_s(\beta_s^*), W_s(\tilde{\beta}_s)\) and \(\Omega_s^{-1}(\tilde{\beta}_s)\) are bounded. According to Von
Neumann’s Trace Inequality, we have

$$
| \text{Tr}\{ W_s(\beta_s^*) - W_s(\hat{\beta}_s) \Omega_s^{-1}(\hat{\beta}_s) \}| \\
\leq | \sum_{r=1} \lambda_r \{ W_s(\beta_s^*) - W_s(\hat{\beta}_s) \} \lambda_r \{ \Omega_s^{-1}(\hat{\beta}_s) \} | \\
= O_p\{ (p_n^5 \log p_n/n)^{1/2} \},
$$

where $\lambda_r$ denotes the $r$th eigenvalues ordered from the least to the greatest.

Next we have

$$
\text{Tr}\{ W_s(\hat{\beta}_s) \{ \Omega_s^{-1}(\beta_s^*) - \Omega_s^{-1}(\hat{\beta}_s) \} \} = \text{Tr}\{ W_s(\hat{\beta}_s) \Omega_s^{-1}(\hat{\beta}_s) \{ \Omega_s(\hat{\beta}_s) - \Omega_s(\beta_s^*) \} \Omega_s^{-1}(\beta_s^*) \} \\
= \text{Tr}\{ [\Omega_s(\hat{\beta}_s) - \Omega_s(\beta_s^*)] \Omega_s^{-1}(\beta_s^*) W_s(\beta_s^*) \Omega_s^{-1}(\hat{\beta}_s) \}.
$$

Similarly we obtain $\max_{j,k} | [\Omega_s(\beta_s^*) - \Omega_s(\hat{\beta}_s) ]_{jk} | = O_p\{ (p_n^3 \log p_n/n)^{1/2} \}$. According to Assumption 2, all eigenvalues of $\Omega_s^{-1}(\beta_s^*) W_s(\beta_s^*) \Omega_s^{-1}(\hat{\beta}_s)$ are positive and bounded. This entails $| \text{Tr}\{ W_s(\hat{\beta}_s) \{ \Omega_s^{-1}(\beta_s^*) - \Omega_s^{-1}(\hat{\beta}_s) \} \}| = O_p\{ (p_n^5 \log p_n/n)^{1/2} \}$. Following similar argument, we have $| \text{Tr}\{ W_s(\beta_s^*) - W_s(\hat{\beta}_s) \} \{ \Omega_s^{-1}(\beta_s^*) - \Omega_s^{-1}(\hat{\beta}_s) \} | = O_p\{ (p_n^5 \log p_n/n)^{1/2} \}$. Combining all the results above, we have $| d_s^* - \hat{d}_s | = O_p\{ (p_n^5 \log p_n/n)^{1/2} \}$. As all the asymptotic orders are established uniformly for all model $s$ in the model space, the consistency result of $\hat{d}_s$ is uniform over the space of all models. \qed
**Lemma S2.1.** Let the joint distribution of the observations from the $i$th cluster follow a canonical multivariate exponential family with the likelihood

$$L_i(Y_i, \theta_i) = \exp\{\sum_{j=1}^{m_i} Y_{ij} \theta_{ij} - b(\theta_i) + c(Y_i)\},$$

where $\theta_i = (\theta_{i1}, \theta_{i2} \ldots \theta_{im})^T$. Assume the parameter space for $\theta_i$ is a compact subspace of $\mathbb{R}^{m_i}$ with the true value $\theta_i^*$ being an interior point of the parameter space, the function $b(\theta_i)$ is three times differentiable and bounded on the parameter space and $Y_{ij}$s are sub-Gaussian random variables. Under Assumption 3, $Q_i(\beta)$, $U_i(\beta)_{[k]}$, $U_i(\beta)_{[kl]}^{(1)}$, and $U_i(\beta)_{[klr]}^{(2)}$ satisfy the cumulant boundedness condition uniformly for all model $s \in S$ and all $\beta$ in the neighborhood $||\beta - \beta_s^*|| \leq (p^2 n \log p n / n)^{1/2}$.

**Proof of Lemma S2.1.** The cumulant generating function of $Y_i$ can be formulated as

$$C_{Y_i}(t_i) = \log E\{\exp(t_i^T \cdot Y_i)\}$$

$$= \log \int \exp\{\sum_{j=1}^{m} t_{ij} Y_{ij}\} \exp\{\sum_{j=1}^{m} \theta_{ij} Y_{ij} - b(\theta_i) + c(Y_i)\} dY_i$$

$$= \log \int \exp\{\sum_{j=1}^{m} t_{ij} + \theta_{ij} Y_{ij} - b(\theta_i + t_i) + c(Y_i)\} \exp\{b(\theta_i + t_i) - b(\theta_i)\} dY_i$$

$$= b(\theta_i + t_i) - b(\theta_i),$$

with $b(\theta_i + t_i) = b(\theta_{i1} + t_{i1}, \theta_{i2} + t_{i2} \ldots \theta_{im} + t_{im})$. Next we consider the cumulant boundedness condition of $U_i(\beta_s^*)_{[k]}$. Let $[R^{-1}]_{[jh]}$ be the $(j, h)$ element of...
matrix $R^{-1}$. We define $f_{ji}^k(\beta) = \{\partial \mu_{ij}(\beta)^T / \partial \beta_{[k]}\} A_{ij}(\beta)^{-1/2}[R^{-1}]_{[j]h} A_{ih}(\beta)^{-1/2}$
and $\tilde{f}_{ji}^k(\beta) = \{\partial \mu_{ij}(\beta)^T / \partial \beta_{[k]}\} A_{ij}(\beta)^{-1/2}[R^{-1}]_{[j]h} A_{ih}(\beta)^{-1/2} \mu_{ih}(\beta)$. We rewrite $U_i(\beta)_{[k]}$ as follows:

$$U_i(\beta)_{[k]} = \frac{\partial \mu_i(\beta)^T}{\partial \beta_{[k]}} V_i(\beta)^{-1} \{Y_i - \mu_i(\beta)\}$$

$$= \sum_{j=1}^{m} \sum_{h=1}^{m} \frac{\partial \mu_{ij}(\beta)^T}{\partial \beta_{[k]}} A_{ij}(\beta)^{-1/2}[R^{-1}]_{[j]h} A_{ih}(\beta)^{-1/2} \{Y_{ih} - \mu_{ih}(\beta)\}$$

$$= \sum_{j=1}^{m} \sum_{h=1}^{m} f_{ji}^k(\beta) Y_{ih} - \tilde{f}_{ji}^k(\beta).$$

This entails the following expressions:

$$U_i(\beta)^{(1)}_{[kl]} = \sum_{j=1}^{m} \sum_{h=1}^{m} \frac{\partial f_{ji}^k(\beta)}{\partial \beta_{[l]}} Y_{ih} - \frac{\partial \tilde{f}_{ji}^k(\beta)}{\partial \beta_{[l]}},$$

$$U_i(\beta)^{(2)}_{[klr]} = \sum_{j=1}^{m} \sum_{h=1}^{m} \frac{\partial^2 f_{ji}^k(\beta)}{\partial \beta_{[l]} \partial \beta_{[r]}} Y_{ih} - \frac{\partial^2 \tilde{f}_{ji}^k(\beta)}{\partial \beta_{[l]} \partial \beta_{[r]}},$$

We first consider the cumulant generating function of $U_i(\beta)_{[k]}$.

$$C_{U_i(\beta)_{[k]}}(t) = \log E[\exp\{t U_i(\beta)_{[k]}\}]$$

$$= \log \int \exp\{t \sum_{j=1}^{m} \sum_{h=1}^{m} f_{ji}^k(\beta) Y_{ih} - \tilde{f}_{ji}^k(\beta)\} \exp\{\sum_{j=1}^{m} \theta_{ij} Y_{ij} - b(\theta_i) + c(Y_i)\} dY_i$$

$$= \log \int \exp\{\sum_{h=1}^{m} \{\theta_{ih} + t \sum_{j=1}^{m} f_{ji}^k(\beta)\} Y_{ih} - b(\theta_i + t(0)) + c(Y_i)\}$$

$$\exp\{t \sum_{j=1}^{m} \sum_{h=1}^{m} \tilde{f}_{ji}^k(\beta) + b(\theta_i + t(0)) - b(\theta_i)\} dY_i$$

$$= b(\theta_i + t(0)) - b(\theta_i) - t \sum_{j=1}^{m} \sum_{h=1}^{m} \tilde{f}_{ji}^k(\beta),$$

17
with $t^{(0)} = (t \sum_{j=1}^{m} f_{j1}^{k}(\beta), t \sum_{j=1}^{m} f_{j2}^{k}(\beta) \ldots t \sum_{j=1}^{m} f_{jm}^{k}(\beta))^T$. Similarly we can calculate the cumulant generating function of $U_i(\beta)_{[kl]}^{(1)}$ and $U_i(\beta)_{[kl]}^{(2)}$ below:

$$C_{U_i(\beta_s)_{[kl]}}^{(1)}(t) = b(\theta_i + t^{(1)}) - b(\theta_i) - t \sum_{j=1}^{m} \sum_{h=1}^{m} \partial \tilde{f}_{jh}^{ik}(\beta) / \partial \beta_{[l]}$$

$$C_{U_i(\beta_s)_{[klr]}}^{(2)}(t) = b(\theta_i + t^{(2)}) - b(\theta_i) - t \sum_{j=1}^{m} \sum_{h=1}^{m} \partial^2 \tilde{f}_{jh}^{ik}(\beta) / \partial \beta_{[l]} \partial \beta_{[r]},$$

with

$$t^{(1)} = (t \sum_{j=1}^{m} \partial f_{j1}^{ik}(\beta) / \partial \beta_{[l]}, t \sum_{j=1}^{m} \partial f_{j2}^{ik}(\beta) / \partial \beta_{[l]}, \ldots, t \sum_{j=1}^{m} \partial f_{jm}^{ik}(\beta) / \partial \beta_{[l]}))^T,$$

$$t^{(2)} = (t \sum_{j=1}^{m} \partial^2 f_{j1}^{ik}(\beta) / \partial \beta_{[l]} \partial \beta_{[r]}, t \sum_{j=1}^{m} \partial^2 f_{j2}^{ik}(\beta) / \partial \beta_{[l]} \partial \beta_{[r]}, \ldots, t \sum_{j=1}^{m} \partial^2 f_{jm}^{ik}(\beta) / \partial \beta_{[l]} \partial \beta_{[r]}))^T.$$
Let $V_{ijh}$ be the $(j, h)$ element of matrix $\hat{V}_i^{-1}$. We can rewrite the $Q_i(\beta)$ as follows:

$$2Q_i(\beta) = \{Y_i - \mu_i(\beta)\}^T \hat{V}_i^{-1} \{Y_i - \mu_i(\beta)\}$$

$$= \sum_{j=1}^{m} \sum_{h=1}^{m} \{Y_{ij} - \mu_{ij}(\beta)\}V_{ijh}\{Y_{ih} - \mu_{ih}(\beta)\}$$

$$= \sum_{j=1}^{m} \sum_{h=1}^{m} V_{ijh}Y_{ij} + V_{ijh}\mu_{ij}(\beta)\mu_{ih}(\beta) - (V_{ijh} + V_{ihj})Y_{ij}\mu_{ih}(\beta).$$

Next we consider the cumulant generating function of $Q_i(\beta)$:

$$C_{Q_i(\beta)}(t) = \log E[\exp\{Q_i(\beta)t\}]$$

$$= \log \int \exp\{t \sum_{j=1}^{m} \sum_{h=1}^{m} V_{ijh}Y_{ij} + V_{ijh}\mu_{ij}(\beta)\mu_{ih}(\beta) - (V_{ijh} + V_{ihj})Y_{ij}\mu_{ih}(\beta)\}$$

$$\exp\{\sum_{j=1}^{m} \theta_{ij} Y_{ij} - b(\theta_i) + c(Y_i)\}dY_i$$

$$= t \sum_{j=1}^{m} \sum_{h=1}^{m} V_{ijh}\mu_{ij}(\beta)\mu_{ih}(\beta) + b(\theta_i + \bar{t}) - b(\theta_i) + \log E\{\exp(t \sum_{j=1}^{m} \sum_{h=1}^{m} V_{ijh}Y_{ij}Y_{ih})\},$$

with $\bar{t} = (-t \sum_{h=1}^{m} (V_{ih1} + V_{i1h})\mu_{i1}(\beta), -t \sum_{h=1}^{m} (V_{ih2} + V_{i2h})\mu_{i2}(\beta), \ldots, -t \sum_{h=1}^{m} (V_{ihm} + V_{imh})\mu_{im}(\beta))^T$. For the last term, we have $\log E\{\exp(t \sum_{j=1}^{m} \sum_{h=1}^{m} V_{ijh}Y_{ij}Y_{ih})\} \leq \max_{i,j,h} \log E[\exp\{tm^2(\max_{i,j,h} |V_{ijh}|)Y_{ij}^2\}]$. If $Y_{ij}$s are sub-Gaussian random variables with uniformly bounded mean and variances, there exist constants $b_s$ and $b_g$ such that for $\forall i, j$,

$$\max_{i,j}[E\{\exp(b_sY_{ij}^2)\}] < b_g$$

([Rudelson and Vershynin, 2013]). Let $b_6 = b_s/(m^2 \max_{i,j} |V_{ijh}|)$, then if
Lemma S2.2. Under Assumptions [1] - [4] for all $\beta$ in the neighborhood of $||\beta - \beta^*|| \leq (p_n^2 \log p_n/n)^{1/2}$, $A_{ij}(\beta)$ are uniformly bounded away from zero and infinity as $n \to \infty$. Furthermore, $|\partial A_{ij}(\beta)/\partial \beta_k|$, $|\partial^2 A_{ij}(\beta)/\partial \beta_k \partial \beta_l|$, $|A_{ij}^{-1/2}(\beta)|$, $|\partial A_{ij}^{-1/2}(\beta)/\partial \beta_k|$, $|\partial^2 A_{ij}^{-1/2}(\beta)/\partial \beta_k \partial \beta_l|$, $|\mu_{ij}(\beta)|$, $|\partial \mu_{ij}(\beta)/\partial \beta_k|$, $|\partial^2 \mu_{ij}(\beta)/\partial \beta_k \partial \beta_l|$, $|\partial^3 \mu_{ij}(\beta)/\partial \beta_k \partial \beta_l \partial \beta_r|$ are all uniformly bounded for all $i \in \{1, 2 \ldots n\}$, $j \in \{1, 2 \ldots m\}$, $k, l, r \in \{1, 2 \ldots p_n\}$.

Proof of Lemma S2.2. We first consider the boundedness of the third derivative of $\mu_{ij}(\beta)$. As $|\partial^3 \mu_{ij}(\beta)/\partial \beta_k \partial \beta_l \partial \beta_r| = X_{ijk}X_{ijr}X_{ijr}g^{-1}\{\zeta_{ij}(\beta)\}/\{\zeta_{ij}(\beta^*_n)\}^3$ and $\partial^3 g^{-1}\{\zeta_{ij}(\beta^*_n)\}/\{\zeta_{ij}(\beta^*_n)\}^3$ is bounded and continuous, therefore $|\partial^3 \mu_{ij}(\beta)/\partial \beta_k \partial \beta_l \partial \beta_r|$ is bounded for $||\beta - \beta^*|| \leq (p_n^2 \log p_n/n)^{1/2}$.

Similarly we can also prove the boundedness of $|\mu_{ij}(\beta)|$, $|\partial \mu_{ij}(\beta)/\partial \beta_k|$, $|\partial^2 \mu_{ij}(\beta)/\partial \beta_k \partial \beta_l|$. Furthermore, $\partial^2 A_{ij}(\beta)/\partial \beta_k \partial \beta_l = [\partial \nu\{\mu_{ij}(\beta)\}/\partial \mu_{ij}(\beta)]$ $[\partial^2 \mu_{ij}(\beta)/\partial \beta_k \partial \beta_l] + [\partial \nu\{\mu_{ij}(\beta)\}/\partial \mu_{ij}(\beta)] [\partial \mu_{ij}(\beta)/\partial \beta_k] [\partial \mu_{ij}(\beta)/\partial \beta_l]$
and
\[
\frac{\partial^2 A_{ij}}{\partial \beta_k \partial \beta_l}(\beta) = (3/4)A_{ij}^{-5/2}(\beta)[\partial A_{ij}(\beta)/\partial \beta_k][\partial A_{ij}(\beta)/\partial \beta_l]
\]

are also bounded. Similarly we can prove the boundedness of the other terms.

Let \( B \) and \( \tilde{B} \) denote \( d_s \times m \) matrices. Let \( D_i^{(1)}(\beta, \tilde{\beta}, B) \) be an \( m \times d_s \) matrix and its \( j \)th row and \( k \)th column entry is
\[
D_i^{(1)}(\beta, \tilde{\beta}, B)_{jk} = (\beta - \beta)T \{\partial^2 \mu_{ij}(B_{[j]})/\partial \beta_k \partial \beta\}.
\]

Let \( D_i^{(2)}(\beta, \tilde{\beta}, B) \) be an \( m \times d_s \) matrix with the \( j \)th row and \( k \)th column entry as
\[
D_i^{(2)}(\beta, \tilde{\beta}, B)_{jk} = (\beta - \beta)T \{\partial^3 \mu_{ij}(B_{[j]})/\partial \beta_k \partial \beta \partial \beta T\}(B_{[j]} - \beta).
\]

**Lemma S2.3.** Let \( B_s^* = (\beta_s^*, \beta_s^* \ldots \beta_s^*) \). Under Assumptions [1 - 4] for all model \( s \in S, i \in \{1, 2 \ldots n\} \), there exist \( d_s \times m \) matrices \( B_s^{i1}, B_s^{i2}, B_s^{i3}, \) and \( B_s^{i4} \) such that each column of these four matrices is between \( \hat{\beta}_s \) and \( \beta_s^* \) and they satisfy:

\[
\mu_i(\hat{\beta}_s) - \mu_i(\beta_s^*) = D_i(\beta_s^*)[\hat{\beta}_s - \beta_s^*] + \frac{1}{2} D_i^{(1)}(\hat{\beta}_s, \beta_s^*, B_s^{i4})(\hat{\beta}_s - \beta_s^*); \quad (S2.1)
\]

\[
\mu_i(\beta_s^*) - \mu_i(\tilde{\beta}_s) = D_i(\tilde{\beta}_s)[\beta_s^* - \tilde{\beta}_s] + \frac{1}{2} D_i^{(1)}(\beta_s^*, \tilde{\beta}_s, B_s^{i4})(\beta_s^* - \tilde{\beta}_s); \quad (S2.2)
\]
\[ D_i(\tilde{\beta}_s) = D_i(\beta^*_s) + D_i^{(1)}(\tilde{\beta}_s, \beta^*_s, B^D_s); \]  
(S2.3)

\[ D_i^{(2)}(\beta^*_s, \tilde{\beta}_s, B^\tilde{\mu}_s, B^\mu_s) = D_i^{(1)}(\beta^*_s, \tilde{\beta}_s, B^\tilde{\mu}_s) - D_i^{(1)}(\beta^*_s, \tilde{\beta}_s, B^*). \]  
(S2.4)

The max norms of the matrices have the following uniform bounds for all model \( s \in S, i \in \{1, 2 \ldots n\} \):

\[ ||D_i^{(1)}(\tilde{\beta}_s, \beta^*_s, B^\mu_i)||_{\text{max}} = O_p\{(p^3_n \log p_n/n)^{1/2}\}, \]

\[ ||D_i^{(1)}(\beta^*_s, \tilde{\beta}_s, B^\tilde{\mu}_i)||_{\text{max}} = O_p\{(p^3_n \log p_n/n)^{1/2}\}, \]

\[ ||D_i^{(1)}(\tilde{\beta}_s, \beta^*_s, B^D_i)||_{\text{max}} = O_p\{(p^3_n \log p_n/n)^{1/2}\}, \]

\[ ||D_i^{(2)}(\beta^*_s, \tilde{\beta}_s, B^\tilde{\mu}_i, B^\mu_i)||_{\text{max}} = O_p(p^3_n \log p_n/n). \]

**Proof of Lemma S2.3** From Taylor expansion, there exists a \( \beta^\mu_{ij} \) between \( \beta^*_s \) and \( \tilde{\beta}_s \) such that

\[ \mu_{ij}(\tilde{\beta}_s) - \mu_{ij}(\beta^*_s) = \frac{\partial \mu_{ij}(\beta^*_s)}{\partial \beta^T}(\tilde{\beta}_s - \beta^*_s) + \frac{1}{2}(\tilde{\beta}_s - \beta^*_s)^T \frac{\partial^2 \mu_{ij}(\beta^\mu_{ij})}{\partial \beta \partial \beta^T}(\tilde{\beta}_s - \beta^*_s). \]  
(S2.5)

Let \( B^\mu_s = (\beta^\mu_{11}, \beta^\mu_{12}, \ldots, \beta^\mu_{m}) \) and each column of \( B^\mu_s \) is between \( \beta^*_s \) and
\( \hat{\beta}_s \). Define

\[
D_i^{(1)}(\hat{\beta}_s, \beta^*_s, \beta_{s}^{\mu_i}) = \begin{bmatrix}
(\hat{\beta}_s - \beta^*_s)^T \{ \partial^2 \mu_{i1}(\beta_{s}^{\mu_1}) / \partial \beta \partial \beta^T \} \\
(\hat{\beta}_s - \beta^*_s)^T \{ \partial^2 \mu_{i2}(\beta_{s}^{\mu_2}) / \partial \beta \partial \beta^T \} \\
\ldots \\
(\hat{\beta}_s - \beta^*_s)^T \{ \partial^2 \mu_{im}(\beta_{s}^{\mu_m}) / \partial \beta \partial \beta^T \}
\end{bmatrix}.
\]

Then Equation (S2.5) can be reformulated as

\[
\mu_i(\hat{\beta}_s) - \mu_i(\beta^*_s) = D_i(\beta^*_s)(\hat{\beta}_s - \beta^*_s) + \frac{1}{2} D_i^{(1)}(\hat{\beta}_s, \beta^*_s, B_{s}^{\mu_i})(\hat{\beta}_s - \beta^*_s).
\]

Similarly if we perform Taylor Expansion at \( \mu_i(\hat{\beta}_s) \), we obtain

\[
\mu_i(\beta^*_s) - \mu_i(\hat{\beta}_s) = D_i(\hat{\beta}_s)(\beta^*_s - \hat{\beta}_s) + \frac{1}{2} D_i^{(1)}(\beta^*_s, \hat{\beta}_s, B_{s}^{\mu_i})(\beta^*_s - \hat{\beta}_s).
\]

By similar argument, there exists a \( \beta_{s}^{D_{ij}} \) between \( \beta^*_s \) and \( \hat{\beta}_s \) such that

\[
\partial \mu_{ij}(\hat{\beta}_s) / \partial \beta_{[k]} = \partial \mu_{ij}(\beta^*_s) / \partial \beta_{[k]} + (\hat{\beta}_s - \beta^*_s)^T \{ \partial^2 \mu_{ij}^{D_{ij}}(\beta_{s}^{D_{ij}}) / \partial \beta_{[k]} \partial \beta \}.
\]

Define \( B_{s}^{D_i} = (\beta_{s}^{D_{i1}}, \beta_{s}^{D_{i2}}, \ldots, \beta_{s}^{D_{im}}) \) and we can reformulate the equation above as

\[
D_i(\hat{\beta}_s) = D_i(\beta^*_s) + D_i^{(1)}(\hat{\beta}_s, \beta^*_s, \beta_{s}^{D_{i}}).
\]

According to Taylor Expansion, there exists a \( \beta_{s}^{\mu_{ij}} \) between \( \beta^*_s \) and \( \beta_{s}^{\mu_{ij}} \) such that
Define $B_s^\beta = (\beta_s^{\mu_1}, \beta_s^{\mu_2}, \ldots, \beta_s^{\mu_m})$. Then the equation above can be simplified as

$$D_i^{(2)}(\beta_s^*, \tilde{\beta}_s, B_s^{\mu_1}, B_s^{\mu_2}) = D_i^{(1)}(\beta_s^*, \tilde{\beta}_s, B_s^{\mu_1}) - D_i^{(1)}(\beta_s^*, \tilde{\beta}_s, B_s^*)$$

Next we estimate the orders of $D_i^{(1)}(\tilde{\beta}_s, \beta_s^*, B_s^{\mu_k})$ and $D_i^{(2)}(\beta_s^*, \tilde{\beta}_s, B_s^{\mu_1}, B_s^{\mu_2})$.

According to Cauchy-Schwarz inequality, we have

$$|D_i^{(1)}(\tilde{\beta}_s, \beta_s^*, B_s^{\mu_k})| = |(\tilde{\beta}_s - \beta_s^*)^T \frac{\partial^2 \mu_{ij}(\beta_s^{\mu_k})}{\partial \beta_{[kl]} \partial \beta}| \leq ||\tilde{\beta}_s - \beta_s^*|| \times \left| \frac{\partial^2 \mu_{ij}(\beta_s^{\mu_k})}{\partial \beta_{[kl]} \partial \beta} \right|.$$
Lemma S2.4. Let $\beta_s, \hat{\beta}_s, \tilde{\beta}_s, \check{\beta}_s$ and every column of $B_i$ be a $d_s \times 1$ vector that falls within a $(p_n^2 \log p_n/n)^{1/2}$ neighborhood of $\beta_T^*$, $i = 1, 2 \ldots n$. Under Assumptions 1 - 4, for any unit vector $||v||^2 = 1$ and any model $s \in S_+$ we have the following bounds:

$$\max_{||v||^2 = 1} v^T \left\{ \frac{1}{n} \sum_{i=1}^n D_i(\beta_s)^T D_i(\beta_s) \right\} v = O(1),$$

$$\max_{||v||^2 = 1} v^T \left\{ \frac{1}{n} \sum_{i=1}^n D_i^{(1)}(\beta_s, \hat{\beta}_s, B_i)^T D_i^{(1)}(\beta_s, \check{\beta}_s, B_i) \right\} v = O_p\left(\frac{p_n^3 \log p_n}{n}\right),$$

$$\max_{||v||^2 = 1} v^T \left\{ \frac{1}{n} \sum_{i=1}^n D_i(\hat{\beta}_s)^T V_i(\tilde{\beta}_s)^{-1} D_i^{(1)}(\beta_s, \check{\beta}_s, B_i) \right\} v = O_p\left(\frac{p_n^3 \log p_n}{n}\right)^{1/2}. $$

Proof of Lemma S2.4. First we have the following bound:

$$\max_{||v||^2 = 1} |v^T \frac{1}{n} \sum_{i=1}^n D_i(\beta_s)^T D_i(\beta_s) v| = \max_{||v||^2 = 1} |v^T \frac{1}{n} \sum_{i=1}^n X_i^T \Lambda_i(\beta_s)^2 X_i v|$$

$$\leq \max_i \lambda_{\max}\{\Lambda_i(\beta_s)^2\} \lambda_{\max}\left\{ \frac{1}{n} \sum_{i=1}^n X_i^T X_i \right\} ||v||^2$$

$$\leq \max_{i,j} \lambda_{\max}\{\Lambda_{ij}(\beta_s)^2\} \lambda_{\max}\left\{ \frac{1}{n} \sum_{i=1}^n X_i^T X_i \right\} ||v||^2$$

$$= O(1).$$

As $\mu_{ij}(\beta_s)$ is differentiable to the third order, we rewrite $\partial^2 \mu_{ij}(\beta_s)/\partial \beta_s \partial \beta_s^T = \{\partial \Lambda_{ij}(\beta_s)/\partial \beta_s\} X_i^T$. Here $\{\partial \Lambda_{ij}(\beta_s)/\partial \beta_s\}$ is a $d_s \times 1$ column vector and $X_i^T$ is a $1 \times d_s$ row vector. We have $D_i^{(1)}(\beta_s, \check{\beta}_s, B)_{[ij]} = (\beta_s - \check{\beta}_s)^T \{\partial^2 \mu_{ij}(B_{[ij]})/\partial \beta_s \partial \beta_s^T\} = 25$
\[(\beta_s - \tilde{\beta}_s)^T \{\partial \Lambda_{ij}(B_{ij})/\partial \beta_s\} X_{ij}^T\]. 

Therefore we have

\[D_i^{(1)}(\beta_s, \tilde{\beta}_s, B) = \begin{bmatrix} (\beta_s - \tilde{\beta}_s)^T \{\partial \Lambda_{i1}(B_{i1})/\partial \beta_s\} X_{i1}^T \\ (\beta_s - \tilde{\beta}_s)^T \{\partial \Lambda_{i2}(B_{i2})/\partial \beta_s\} X_{i2}^T \\ \vdots \\ (\beta_s - \tilde{\beta}_s)^T \{\partial \Lambda_{im}(B_{im})/\partial \beta_s\} X_{im}^T \end{bmatrix}\].

Let \(\text{diag}_{j=1}^n \{(\beta_s - \tilde{\beta}_s)^T [\partial \{\Lambda_{ij}(\beta_s)\}/\partial \beta_s]\}\) represent a diagonal matrix with the \(j\)th diagonal entry equal to \((\beta_s - \tilde{\beta}_s)^T [\partial \{\Lambda_{ij}(\beta_s)\}/\partial \beta_s]\). Then we can reformat \(D_i^{(1)}(\beta_s, \tilde{\beta}_s, B) = \text{diag}_{j=1}^n \{(\beta_s - \tilde{\beta}_s)^T [\partial \{\Lambda_{ij}(\beta_s)\}/\partial \beta_s]\} X_i\). From Assumption 2 we have the boundedness of \(\lambda_{\text{max}} \{n^{-1} \sum_{i=1}^n X_i^T X_i\}\). This entails

\[
\begin{align*}
\max_{||v||^2 = 1} |v^T \frac{1}{n} \sum_{i=1}^n D_i^{(1)}(\beta_s, \tilde{\beta}_s, B_i) D_i^{(1)}(\beta_s, \tilde{\beta}_s, B_i) v| & = \max_{||v||^2 = 1} |v^T \frac{1}{n} \sum_{i=1}^n X_i^T \text{diag}_{j=1}^n \{(\beta_s - \tilde{\beta}_s)^T [\partial \Lambda_{ij}(B_{ij})/\partial \beta_s]\}^2 X_i v| \\
& \leq \max_i \lambda_{\text{max}} \{\text{diag}_{j=1}^n \{(\beta_s - \tilde{\beta}_s)^T [\partial \Lambda_{ij}(B_{ij})/\partial \beta_s]\}^2\} \lambda_{\text{max}} \left(\frac{1}{n} \sum_{i=1}^n X_i^T X_i\right) ||v||^2 \\
& \leq \max_{i,j} \{(\beta_s - \tilde{\beta}_s)^T [\partial \Lambda_{ij}(B_{ij})/\partial \beta_s]\}^2 \lambda_{\text{max}} \left(\frac{1}{n} \sum_{i=1}^n X_i^T X_i\right) \\
& \leq ||\beta_s - \tilde{\beta}_s||^2 \max_{i,j} \left(\frac{||\partial \Lambda_{ij}(B_{ij})/\partial \beta_s||^2}{\partial \beta_s}\right) \lambda_{\text{max}} \left(\frac{1}{n} \sum_{i=1}^n X_i^T X_i\right) \\
& \leq ||\beta_s - \tilde{\beta}_s||^2 \times p_n \times \max_{i,j,k} \left(\frac{||\partial \Lambda_{ij}(B_{ij})/\partial \beta_s||^2}{\partial \beta_s}\right) \lambda_{\text{max}} \left(\frac{1}{n} \sum_{i=1}^n X_i^T X_i\right) \\
& = O_p(p_n^3 \log p_n/n);
\end{align*}
\]
\[
\max_{||v||^2 = 1} \left| v^T \frac{1}{n} \sum_{i=1}^{n} D_i (\hat{\beta}_s)^T V_i (\tilde{\beta}_s)^{-1} D_i^{(1)} (\beta_s, \tilde{\beta}_s, B_i) v \right|
\]

\[
= \max_{||v||^2 = 1} \left| v^T \frac{1}{n} \sum_{i=1}^{n} X_i^T A_i (\hat{\beta}_s) V_i (\tilde{\beta}_s)^{-1} \text{diag}_{j=1}^{m} \left\{ (\beta_s - \tilde{\beta}_s)^T \frac{\partial \Lambda_{ij}(B_{i[j]})}{\partial \beta_s} \right\} X_i v \right|
\]

\[
\leq \max_{i,j} \{ A_{ij}(\hat{\beta}_s) \} \times \max_{i} \lambda_{\text{max}} \{ V_i (\tilde{\beta}_s)^{-1} \} \times \max_{i,j} \left\{ (\beta_s - \tilde{\beta}_s)^T \frac{\partial \Lambda_{ij}(B_{i[j]})}{\partial \beta_s} \right\}
\]

\[
\times \lambda_{\max} \left( \frac{1}{n} \sum_{i=1}^{n} X_i^T X_i \right) \times ||v||^2
\]

\[
\leq ||\beta_s - \tilde{\beta}_s|| \times \max_{i,j} \left\| \frac{\partial \Lambda_{ij}(B_{i[j]})}{\partial \beta_s} \right\| \times O(1)
\]

\[
\leq ||\beta_s - \tilde{\beta}_s|| \times (p_n^{1/2}) \times \max_{i,j,k} \left\| \frac{\partial \Lambda_{ij}(B_{i[j]})}{\partial \beta_{s[k]}} \right\| \times O(1)
\]

\[
= O_p \{ (p_n^3 \log p_n/n)^{1/2} \}.
\]

\[\square\]

**Lemma S2.5.** Under Assumption [4] - [7], the estimated inverse working covariance matrices can be decomposed into the sum of several matrices of the same dimension \( V_i (\tilde{\beta}_s)^{-1} = V_i^{-1}(\beta^*_s) + V_i^{(1)}(\hat{\beta}_s, \beta^*_s) + V_i^{(2)}(\tilde{\beta}_s, \beta^*_s, B_{sA}) \), where \( B_{sA} = (\beta_{sA_{1}}, \ldots, \beta_{sA_{m}}) \), and each \( \beta_{sA_{ij}} \), \( j = 1, \ldots, m \), is a vector between \( \hat{\beta}_s \) and \( \beta^*_s \). Let \( \tilde{V}_i^{(1)}(\hat{\beta}_s, \beta^*_s, B_{sA}) = V_i^{(1)}(\hat{\beta}_s, \beta^*_s) + V_i^{(2)}(\tilde{\beta}_s, \beta^*_s, B_{sA}) \).

The bounds

\[
||V_i^{(1)}(\hat{\beta}_s, \beta^*_s)||_{\text{max}} = O_p \{ (p_n^3 \log p_n/n)^{1/2} \},
\]

\[
||\tilde{V}_i^{(1)}(\hat{\beta}_s, \beta^*_s, B_{sA})||_{\text{max}} = O_p \{ (p_n^3 \log p_n/n)^{1/2} \},
\]

\[
||V_i^{(2)}(\tilde{\beta}_s, \beta^*_s, B_{sA})||_{\text{max}} = O_p \{ p_n^3 \log p_n/n \}
\]

27
are uniformly held for all model \( s \in S \), and \( i = 1, 2 \ldots n \).

**Proof of Lemma S2.5.** According to Taylor expansion, there exists a \( \beta^A_{s} \) between \( \hat{\beta}_s \) and \( \beta^*_s \) such that

\[
A_{ij}^{-1/2} (\hat{\beta}_s) = A_{ij}^{-1/2} (\beta^*_s) + (\hat{\beta}_s - \beta^*_s)^T \frac{\partial A_{ij}^{-1/2} (\beta^*_s)}{\partial \beta} + \frac{1}{2} (\hat{\beta}_s - \beta^*_s)^T \frac{\partial^2 A_{ij}^{-1/2} (\beta^A_{ij})}{\partial \beta \partial \beta^T} (\hat{\beta}_s - \beta^*_s).
\]

For the \( j \)th row and \( h \)th column of matrix \( V_i^{-1} (\hat{\beta}_s) - V_i^{-1} (\beta^*_s) \), we apply the formula above and obtain

\[
[V_i^{-1} (\hat{\beta}_s) - V_i^{-1} (\beta^*_s)]_{jh} = A_{ij}^{-1/2} (\hat{\beta}_s)[R^{-1}]_{jh} A_{ih}^{-1/2} (\hat{\beta}_s) - A_{ij}^{-1/2} (\beta^*_s)[R^{-1}]_{jh} A_{ih}^{-1/2} (\beta^*_s)
\]

\[
= (\hat{\beta}_s - \beta^*_s)^T [R^{-1}]_{jh} \{ A_{ih}^{-1/2} (\beta^*_s) \frac{\partial A_{ij}^{-1/2} (\beta^*_s)}{\partial \beta} + A_{ij}^{-1/2} (\beta^*_s) \frac{\partial A_{ih}^{-1/2} (\beta^*_s)}{\partial \beta} \}
\]

\[
+ (\hat{\beta}_s - \beta^*_s)^T [R^{-1}]_{jh} \{ \frac{1}{2} A_{ih}^{-1/2} (\beta^*_s) \frac{\partial^2 A_{ij}^{-1/2} (\beta^A_{ij})}{\partial \beta \partial \beta^T} + \frac{1}{2} A_{ij}^{-1/2} (\beta^*_s) \frac{\partial^2 A_{ih}^{-1/2} (\beta^A_{ih})}{\partial \beta \partial \beta^T} \}
\]

\[
+ \frac{\partial A_{ij}^{-1/2} (\beta^*_s)}{\partial \beta} \frac{\partial A_{ih}^{-1/2} (\beta^*_s)}{\partial \beta^T} (\hat{\beta}_s - \beta^*_s) + O_p (||\hat{\beta}_s - \beta^*_s||^3).
\]

Denote the first term in the expansion as \( V_i^{(1)} (\hat{\beta}_s, \beta^*_s)_{[jh]} \) and the remaining three terms as \( V_i^{(2)} (\hat{\beta}_s, \beta^*_s, B^A_{s})_{[jh]} \). Based on the Cauchy-Schwarz inequality, the bounds determined in Lemma S2.2 and Assumption 2 we have

\[
|V_i^{(1)} (\hat{\beta}_s, \beta^*_s)_{[jh]}|
\]

\[
\leq ||\hat{\beta}_s - \beta^*_s|| \times |[R^{-1}]_{jh} \{ A_{ij}^{-1/2} (\beta^*_s) \frac{\partial A_{ij}^{-1/2} (\beta^*_s)}{\partial \beta} + A_{ij}^{-1/2} (\beta^*_s) \frac{\partial A_{ih}^{-1/2} (\beta^*_s)}{\partial \beta} \} |
\]

\[
\leq p_n^{1/2} ||\hat{\beta}_s - \beta^*_s|| \times \max_i |[R^{-1}]_{jh} \{ A_{ij}^{-1/2} (\beta^*_s) \frac{\partial A_{ij}^{-1/2} (\beta^*_s)}{\partial \beta_{[i]} } + A_{ij}^{-1/2} (\beta^*_s) \frac{\partial A_{ih}^{-1/2} (\beta^*_s)}{\partial \beta_{[i]} } \} |
\]

\[
= O_p (p_n^{1/2} ||\hat{\beta}_s - \beta^*_s||)
\]

\[
= O_p (p_n^3 \log p_n / n)^{1/2};
\]

28
and
\[
|V_i^{(2)}(\hat{\beta}_s, \beta^*_s, B_s^A)| \leq \|\hat{\beta}_s - \beta^*_s\| \times p_n^{1/2} \max_l ||[R^{-1}]_{jl}\left\{ \frac{1}{2} A_{ij}^{-1/2}(\beta^*_s) \frac{\partial^2 A_{ih}^{-1/2}(\beta^*_s)}{\partial \beta_{[jl]} \partial \beta_{[r]}} \right\} \|
\]
\[
+ \frac{1}{2} A_{ij}^{-1/2}(\beta^*_s) \frac{\partial^2 A_{ij}^{-1/2}(\beta^*_s)}{\partial \beta_{[jl]} \partial \beta_{[r]}} \}
\]
\[
\leq \|\hat{\beta}_s - \beta^*_s\|^2 \times p_n \max_l ||[R^{-1}]_{jl}\left\{ \frac{1}{2} A_{ij}^{-1/2}(\beta^*_s) \frac{\partial^2 A_{ij}^{-1/2}(\beta^*_s)}{\partial \beta_{[jl]} \partial \beta_{[r]}} \right\} \|
\]
\[
+ \frac{1}{2} A_{ij}^{-1/2}(\beta^*_s) \frac{\partial^2 A_{ij}^{-1/2}(\beta^*_s)}{\partial \beta_{[jl]} \partial \beta_{[r]}} \}
\]
\[
= O_p(||\hat{\beta}_s - \beta^*_s||^2 \times p_n) = O_p(p^3 \log p/n).
\]

Lemma S2.6. Under Assumptions 1 - 4, for all model s ∈ S,
\[
\max_{j=1}^m \frac{1}{n} \sum_{i=1}^n |Y_{ij} - \mu_{ij}(\beta^*_s)| = O_p(1).
\]

Proof of Lemma S2.6. First we consider the true and overfitting models s ∈ S+. As Y_{ij}'s are independent and their variances are uniformly bounded, we have
\[
\text{var}\{|Y_{ij} - \mu_{ij}(\beta^*_s)|\} = \mathbb{E}\{|Y_{ij} - \mu_{ij}(\beta^*_s)|^2\} - \left[\mathbb{E}\{|Y_{ij} - \mu_{ij}(\beta^*_s)|\}\right]^2
\]
\[
\leq \mathbb{E}\{|Y_{ij} - \mu_{ij}(\beta^*_s)|^2\} = \text{var}(Y_{ij})
\]
\[
\leq b_2.
\]

By the Law of Large Numbers,
\[
\frac{1}{n} \sum_{i=1}^n |Y_{ij} - \mu_{ij}(\beta^*_s)| \overset{p}{\rightarrow} \frac{1}{n} \sum_{i=1}^n \mathbb{E}\{|Y_{ij} - \mu_{ij}(\beta^*_s)|\}.
\]
Furthermore,
\[
\frac{1}{n} \sum_{i=1}^{n} E\{|Y_{ij} - \mu_{ij}(\beta_s^*)|\} \leq \frac{1}{2n} \sum_{i=1}^{n} E\{|Y_{ij} - \mu_{ij}(\beta_s^*)|^2 + 1\}
\]
\[
\leq \frac{1}{2n} \left\{ \sum_{i=1}^{n} \text{var}(Y_{ij}) + 1 \right\}
\]
\[
\leq (b_2 + 1)/2.
\]

For all \(j = 1, 2, \ldots, m\), we have \(n^{-1} \sum_{i=1}^{n} |Y_{ij} - \mu_{ij}(\beta_s^*)| = O_p(1)\).

For underfitting models \(s \in S_-\), Lemma [S2.2] implies that both \(\mu_{ij}(\beta_T^*)\) and \(\mu_{ij}(\beta_s^*)\) are bounded. Thus \(n^{-1} \sum_{i=1}^{n} |\mu_{ij}(\beta_T^*) - \mu_{ij}(\beta_s^*)| = O(1)\). For \(j = 1, 2, \ldots, m\), we have \(n^{-1} \sum_{i=1}^{n} |Y_{ij} - \mu_{ij}(\beta_T^*)| \leq n^{-1} \sum_{i=1}^{n} |Y_{ij} - \mu_{ij}(\beta_s^*)| + n^{-1} \sum_{i=1}^{n} |\mu_{ij}(\beta_T^*) - \mu_{ij}(\beta_s^*)| = O_p(1)\).

**Lemma S2.7.** Under Assumptions [7 - 4] for the true and overfitting models,
\[
\sum_{i=1}^{n} \{\mu_i(\beta_s^*) - \mu_i(\tilde{\beta}_s)\}^T \hat{V}_i^{-1} \{Y_i - \mu_i(\tilde{\beta}_s)\} = n||\beta_s^* - \tilde{\beta}_s||^2 o_p(1),
\]
where the \(o_p(1)\) term holds for all models \(s \in S_+\).

**Proof of Lemma S2.7.** From Equation (S2.2) of Lemma S2.3, we have
\[
\sum_{i=1}^{n} \{\mu_i(\beta_s^*) - \mu_i(\tilde{\beta}_s)\}^T \hat{V}_i^{-1} \{Y_i - \mu_i(\tilde{\beta}_s)\}
\]
\[
= \sum_{i=1}^{n} \{D_i(\tilde{\beta}_s)(\beta_s^* - \tilde{\beta}_s) + \frac{1}{2} D_i^{(1)}(\tilde{\beta}_s, \beta_s^*, B_s^\mu)(\beta_s^* - \tilde{\beta}_s)\}^T \hat{V}_i^{-1} \{Y_i - \mu_i(\tilde{\beta}_s)\}
\]
\[
= (\beta_s^* - \tilde{\beta}_s)^T \sum_{i=1}^{n} D_i(\tilde{\beta}_s)^T \{\hat{V}_i^{-1} - V_i(\tilde{\beta}_s)^{-1} + V_i(\tilde{\beta}_s)^{-1}\} \{Y_i - \mu_i(\tilde{\beta}_s)\}
\]
\[ + \frac{1}{2}(\beta_s^* - \tilde{\beta}_s)^T \sum_{i=1}^n D_i^{(1)}(\beta_s^*, \tilde{\beta}_s, B_s^\mu) T \tilde{V}_i^{-1} \{Y_i - \mu_i(\tilde{\beta}_s)\} \]
\[ = (\beta_s^* - \tilde{\beta}_s)^T \sum_{i=1}^n D_i(\tilde{\beta}_s) T \{\tilde{V}_i^{-1} - V_i(\tilde{\beta}_s)^{-1}\} \{Y_i - \mu_i(\tilde{\beta}_s)\} \]
\[ + (\beta_s^* - \tilde{\beta}_s)^T U(\tilde{\beta}_s) \]
\[ + \frac{1}{2}(\beta_s^* - \tilde{\beta}_s)^T \sum_{i=1}^n D_i^{(1)}(\beta_s^*, \tilde{\beta}_s, B_s^\mu) T \tilde{V}_i^{-1} \{Y_i - \mu_i(\tilde{\beta}_s)\} \]
\[ = (\beta_s^* - \tilde{\beta}_s)^T \sum_{i=1}^n D_i(\tilde{\beta}_s) T \{\tilde{V}_i^{-1} - V_i(\tilde{\beta}_s)^{-1}\} \{Y_i - \mu_i(\tilde{\beta}_s)\} \]
\[ + \frac{1}{2}(\beta_s^* - \tilde{\beta}_s)^T \sum_{i=1}^n D_i^{(1)}(\beta_s^*, \tilde{\beta}_s, B_s^\mu) T \tilde{V}_i^{-1} \{Y_i - \mu_i(\tilde{\beta}_s)\} \]
\[ = Res_1 + Res_2. \]

We expand the residual terms as follows.

\[ Res_1 = (\beta_s^* - \tilde{\beta}_s)^T \sum_{i=1}^n D_i(\tilde{\beta}_s) T \{\tilde{V}_i^{-1} - V_i(\tilde{\beta}_s)^{-1}\} \{Y_i - \mu_i(\tilde{\beta}_s)\} \]
\[ = (\beta_s^* - \tilde{\beta}_s)^T \sum_{i=1}^n D_i(\tilde{\beta}_s) T \{\tilde{V}_i^{-1} - V_i(\tilde{\beta}_s)^{-1}\} \{Y_i - \mu_i(\beta_s^*)\} \]
\[ + (\beta_s^* - \tilde{\beta}_s)^T \sum_{i=1}^n D_i(\tilde{\beta}_s) T \{\tilde{V}_i^{-1} - V_i^{-1}(\beta_s)\} \{\mu_i(\beta_s^*) - \mu_i(\tilde{\beta}_s)\} \]
\[ = (\beta_s^* - \tilde{\beta}_s)^T \sum_{i=1}^n D_i(\tilde{\beta}_s) T \{\tilde{V}_i^{-1} - (V_i^*)^{-1} + (V_i^*)^{-1} - V_i(\tilde{\beta}_s)^{-1}\} \{Y_i - \mu_i(\beta_s^*)\} + Res_{11} \]
\[ = Res_{11} + (\beta_s^* - \tilde{\beta}_s)^T \sum_{i=1}^n D_i(\tilde{\beta}_s) T \{\tilde{V}_i^{-1} - (V_i^*)^{-1}\} \{Y_i - \mu_i(\beta_s^*)\} \]
\[ - (\beta_s^* - \tilde{\beta}_s)^T \sum_{i=1}^n D_i(\tilde{\beta}_s) T \{V_i(\tilde{\beta}_s)^{-1} - (V_i^*)^{-1}\} \{Y_i - \mu_i(\beta_s^*)\} \]
\[ = Res_{11} + Res_{12} - Res_{13}. \]
The first term can be further decomposed:

\[
Res_{11} = (\beta_\star - \hat{\beta}_s)^T \sum_{i=1}^n D_i(\hat{\beta}_s)^T \{\hat{V}_i^{-1} - V_i(\hat{\beta}_s)^{-1}\}\{\mu_i(\beta_\star) - \mu_i(\hat{\beta}_s)\}
\]

\[
= (\beta_\star - \hat{\beta}_s)^T \sum_{i=1}^n D_i(\hat{\beta}_s)^T \{\hat{V}_i^{-1} - V_i(\hat{\beta}_s)^{-1}\} \{D_i(\hat{\beta}_s) + \frac{1}{2} D_i^{(1)}(\beta_\star, \hat{\beta}_s, B_{\beta_s}^\beta)\}(\beta_\star - \hat{\beta}_s)
\]

\[
= (\beta_\star - \hat{\beta}_s)^T \sum_{i=1}^n D_i(\hat{\beta}_s)^T \{\hat{V}_i^{-1} - V_i(\hat{\beta}_s)^{-1}\} D_i(\hat{\beta}_s)(\beta_\star - \hat{\beta}_s)
\]

\[
+ \frac{1}{2} (\beta_\star - \hat{\beta}_s)^T \sum_{i=1}^n D_i(\hat{\beta}_s)^T \hat{V}_i^{-1} D_i^{(1)}(\beta_\star, \hat{\beta}_s, B_{\beta_s}^\beta)(\beta_\star - \hat{\beta}_s)
\]

\[
- \frac{1}{2} (\beta_\star - \hat{\beta}_s)^T \sum_{i=1}^n D_i(\hat{\beta}_s)^T V_i(\hat{\beta}_s)^{-1} D_i^{(1)}(\beta_\star, \hat{\beta}_s, B_{\beta_s}^\beta)(\beta_\star - \hat{\beta}_s)
\]

\[
= Res_{111} + Res_{112} + Res_{113}.
\]

We obtain the bounds for each of the residual terms:

\[
|Res_{111}| = |(\beta_\star - \hat{\beta}_s)^T \sum_{i=1}^n D_i(\hat{\beta}_s)^T \{\hat{V}_i^{-1} - V_i(\hat{\beta}_s)^{-1}\} D_i(\hat{\beta}_s)(\beta_\star - \hat{\beta}_s)|
\]

\[
\leq n \max\{|\lambda_{\max}\{\hat{V}_i^{-1} - V_i(\hat{\beta}_s)^{-1}\}|, |\lambda_{\min}\{\hat{V}_i^{-1} - V_i(\hat{\beta}_s)^{-1}\}|\}
\]

\[
\times (\beta_\star - \hat{\beta}_s)^T \frac{1}{n} \sum_{i=1}^n D_i(\hat{\beta}_s)^T D_i(\hat{\beta}_s)(\beta_\star - \hat{\beta}_s)
\]

\[
\leq n ||\beta_\star - \hat{\beta}_s||^2 \times \max\{|\lambda_{\max}\{\hat{V}_i^{-1} - V_i(\hat{\beta}_s)^{-1}\}|, |\lambda_{\min}\{\hat{V}_i^{-1} - V_i(\hat{\beta}_s)^{-1}\}|\}
\]

\[
\times \lambda_{\max}\{\frac{1}{n} \sum_{i=1}^n D_i(\hat{\beta}_s)^T D_i(\hat{\beta}_s)\}
\]

\[
= n ||\beta_\star - \hat{\beta}_s||^2 O_p\{(p_n^3 \log p_n/n)^{1/2}\};
\]
$$|2\text{Res}_{12}| = |(\beta_s^* - \hat{\beta}_s)^T \sum_{i=1}^{n} D_i (\hat{\beta}_s)^T \hat{V}_i^{-1} D_i^{(1)} (\beta_s^*, \hat{\beta}_s, B_s^{\mu*})(\beta_s^* - \hat{\beta}_s)|$$

$$= |n(\beta_s^* - \hat{\beta}_s)^T \frac{1}{n} \sum_{i=1}^{n} D_i^{(1)} (\beta_s^*, \hat{\beta}_s, B_s^{\mu*})^T \hat{V}_i^{-1} D_i (\hat{\beta}_s)(\beta_s^* - \hat{\beta}_s)|$$

$$\leq n||\beta_s^* - \hat{\beta}_s||^2 \max \left\{ v^T \frac{1}{n} \sum_{i=1}^{n} D_i^{(1)} (\beta_s^*, \hat{\beta}_s, B_s^{\mu*})^T \hat{V}_i^{-1} D_i (\hat{\beta}_s)v \right\}$$

$$= n||\beta_s^* - \hat{\beta}_s||^2 O_p\{ (p_n^3 \log p_n / n)^{1/2} \}.$$ 

Following the same technique on Res_{12}, we obtain $|\text{Res}_{11}| = n||\beta_s^* - \hat{\beta}_s||^2 O_p\{ (p_n^3 \log p_n / n)^{1/2} \}$. Applying Lemma 2.3 and 2.5 to Res_{12}, we have

$$\text{Res}_{12} = (\beta_s^* - \hat{\beta}_s)^T \sum_{i=1}^{n} D_i (\hat{\beta}_s)^T (\hat{V}_i^{-1} - (V_i^*)^{-1}) \{ Y_i - \mu_i(\beta_s^*) \}$$

$$= (\beta_s^* - \hat{\beta}_s)^T \sum_{i=1}^{n} \{ D_i (\beta_s^*)^T + D_i^{(1)} (\hat{\beta}_s, \beta_s^*, B_s^{\mu*}) \} \{ V_i^{(1)}(\hat{\beta}_F, \beta_F^*)$$

$$+ V_i^{(2)}(\hat{\beta}_F, \beta_F^*, B_F^{A_i}) \} \{ Y_i - \mu_i(\beta_s^*) \}$$

$$= (\beta_s^* - \hat{\beta}_s)^T \sum_{i=1}^{n} \{ D_i (\beta_s^*)^T V_i^{(1)}(\hat{\beta}_F, \beta_F^*) + D_i (\beta_s^*)^T V_i^{(2)}(\hat{\beta}_F, \beta_F^*, B_F^{A_i})$$

$$+ D_i^{(1)} (\beta_s, \beta_s^*, B_s^{D_i}) V_i^{(1)}(\hat{\beta}_F, \beta_F^*) + D_i^{(1)} (\beta_s, \beta_s^*, B_s^{D_i}) V_i^{(2)}(\hat{\beta}_F, \beta_F^*, B_F^{A_i}) \} \{ Y_i - \mu_i(\beta_s^*) \}$$

$$= \text{Res}_{121} + \text{Res}_{122} + \text{Res}_{123} + \text{Res}_{124}.$$ 

For Res_{121}, define the $d_s \times 1$ vector $\Gamma = \sum_{i=1}^{n} D_i (\beta_s^*)^T V_i^{(1)}(\hat{\beta}_F, \beta_F^*) \{ Y_i - \mu_i(\beta_s^*) \}$. Res_{121} can be reformulated as $(\hat{\beta}_s - \beta_s^*)^T \Gamma$. The $k$th element of $\Gamma$ is denoted as $\Gamma_k$. The $k$th row of $D_i (\beta_s^*)^T$ is denoted as $[D_i (\beta_s^*)^T]_{[k]}$. 

33
\[
\Gamma_k = \sum_{i=1}^{n} [D_i(\beta_s^*)^T]_{[k]} V_i^{(1)}(\hat{\beta}_F, \beta_F^*) \{Y_i - \mu_i(\beta_s^*)\}
\]
\[
= \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{j=1}^{m} [D_i(\beta_s^*)^T]_{[k]} V_i^{(1)}(\hat{\beta}_F, \beta_F^*)_{[ij]} \{Y_{ij} - \mu_{ij}(\beta_s^*)\}
\]
\[
= \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{j=1}^{m} D_i(\beta_s^*)_{[jk]}(\hat{\beta}_F - \beta_F^*)^T [R^{-1}]_{[ij]} \{A_{ij}^{-1/2}(\beta_F^*) \frac{\partial A_{ij}^{-1/2}(\beta_F^*)}{\partial \beta} + A_{ij}^{-1/2}(\beta_F^*) \frac{\partial A_{ij}^{-1/2}(\beta_F^*)}{\partial \beta} \} \{Y_{ij} - \mu_{ij}(\beta_s^*)\}
\]
\[
= (\hat{\beta}_F - \beta_F^*)^T \Pi_k(\beta_s^*).
\]

Note that for true and overfitting model, \( \mu_i(\beta_F^*) = \mu_i(\beta_s^*) \). Here \( \Pi_k(\beta_s^*) = \sum_{i=1}^{n} \sum_{j=1}^{m} D_i(\beta_s^*)_{[jk]} [R^{-1}]_{[ij]} \{A_{ij}^{-1/2}(\beta_s^*) \frac{\partial A_{ij}^{-1/2}(\beta_s^*)}{\partial \beta} + A_{ij}^{-1/2}(\beta_s^*) \frac{\partial A_{ij}^{-1/2}(\beta_s^*)}{\partial \beta} \} \{Y_{ij} - \mu_{ij}(\beta_s^*)\} \) represents a \( d_s \times 1 \) vector. The \( r \)th element of \( \Pi_k(\beta_s^*) \) is denoted as \( \Pi_{kr}(\beta_s^*) \). Then we have

\[
\Pi_{kr}(\beta_s^*) = \sum_{i=1}^{n} \sum_{j=1}^{m} D_i(\beta_s^*)_{[jk]} [R^{-1}]_{[ij]} \{A_{ij}^{-1/2}(\beta_s^*) \frac{\partial A_{ij}^{-1/2}(\beta_s^*)}{\partial \beta_{[r]}} + A_{ij}^{-1/2}(\beta_s^*) \frac{\partial A_{ij}^{-1/2}(\beta_s^*)}{\partial \beta_{[r]}} \} \{Y_{ij} - \mu_{ij}(\beta_s^*)\}.
\]

Given that \( \text{E}[Y_{ij}] = \mu_{ij}(\beta_s^*) \), then \( \text{E}[n^{-1}\Pi_{kr}(\beta_s^*)] = 0 \). According to Lemma S2.2, \( A_{ih}^{-1/2}(\beta_s^*) \) and its first derivative are uniformly bounded for all \( i, h \) and all model \( s \). Therefore there exists a \( b_H \) for all model \( s \) such
that:

\[
\text{Var}\left\{ \frac{1}{n} \Pi_{kr}(\beta^*_s) \right\} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{h=1}^{m} \sum_{h=1}^{m} D_i(\beta^*_s)_{ij}[R^{-1}]_{ij} \left[ D_i(\beta^*_s)_{hh}[R^{-1}]_{hh} \right.
\]

\[
\left\{ A^{-1/2}_{ij}(\beta^*_s) \frac{\partial A^{-1/2}_{ij}(\beta^*_s)}{\partial \beta_r} + A^{-1/2}_{ij}(\beta^*_s) \frac{\partial A^{-1/2}_{ij}(\beta^*_s)}{\partial \beta_r} \right\}
\]

\[
\{ A^{-1/2}_{ih}(\beta^*_s) \frac{\partial A^{-1/2}_{ih}(\beta^*_s)}{\partial \beta_r} + A^{-1/2}_{ih}(\beta^*_s) \frac{\partial A^{-1/2}_{ih}(\beta^*_s)}{\partial \beta_r} \} \text{Cov}(Y_{ij}, Y_{ih}) \leq \frac{b\Pi}{n}.
\]

According to Chebyshev’s inequality,

\[
\text{Pr}\left\{ \left| \frac{1}{n} \Pi_{kr}(\beta^*_s) \right| \geq \left( \frac{b\Pi}{n} p_n^2 \log p_n \right)^{1/2} \right\} \leq \frac{1}{p_n^2 \log p_n}.
\]

When \( p_n \to \infty \), according to Bonferroni inequality,

\[
\text{Pr}\left\{ \max_{k,r} \left| \frac{1}{n} \Pi_{kr}(\beta^*_s) \right| \geq \left( \frac{b\Pi}{n} p_n^2 \log p_n \right)^{1/2} \right\} \leq \frac{p_n^2}{p_n^2 \log p_n} = (\log p_n)^{-1} \to 0,
\]

Or equivalently we have

\[
\max_{k,r} \left| \Pi_{kr}(\beta^*_s) \right| = O_p\{ (p_n^2 \log p_n)^{1/2} \}.
\]
According to Cauchy-Schwarz inequality

$$|Res_{121}| \leq ||\hat{\beta}_s - \beta^*_s|| \times ||\Gamma|| \leq p_n^{1/2} \times ||\hat{\beta}_s - \beta^*_s|| \times \max_k |\Gamma_k|.$$ 

$$|\Gamma_k| \leq ||\hat{\beta}_s - \beta^*_s|| \times ||\Pi_k|| \leq p_n^{1/2} \times ||\hat{\beta}_s - \beta^*_s|| \times \max \{\Pi_r(\beta^*_s)\} = ||\hat{\beta}_s - \beta^*_s||O_p\{p_n^3 \log p_n\}^{1/2}\}.$$ 

Therefore we have

$$|Res_{121}| = n||\hat{\beta}_s - \beta^*_s||^2 \times O_p\{(p_n^4 \log p_n/n)^{1/2}\}.$$ 

For the term $Res_{122}$, Lemma [S2.5] implies that the largest elements of matrix $||V_i^{(2)}(\hat{\beta}_F, \beta^*_F, B^A_F)||_{\max}$ is $O_p(p_n^3 \log p_n/n)$. Lemma [S2.2] implies that all elements from $D_i(\beta^*_s)$ are bounded. Lemma [S2.6] demonstrates that $\sum_{i=1}^n |Y_{ij} - \mu_{ij}(\beta^*_s)|/n$ are bounded for all $j \in \{1, 2 \ldots m\}$. Therefore we have

$$|Res_{122}| = (\hat{\beta}_s - \beta^*_s)^T \sum_{i=1}^n D_i(\beta^*_s)^T V_i^{(2)}(\hat{\beta}_F, \beta^*_F, B^A_F) \{Y_i - \mu_i(\beta^*_s)\}$$

$$\leq n||\hat{\beta}_s - \beta^*_s|| \times \frac{1}{n} \sum_{i=1}^n D_i(\beta^*_s)^T V_i^{(2)}(\hat{\beta}_F, \beta^*_F, B^A_F) \{Y_i - \mu_i(\beta^*_s)\}$$

$$\leq np_n^{1/2}||\hat{\beta}_s - \beta^*_s|| \times \max_k \left\{\frac{1}{n} \sum_{i=1}^n \left[D_i(\beta^*_s)^T \left[V_i^{(2)}(\hat{\beta}_F, \beta^*_F, B^A_F) \{Y_i - \mu_i(\beta^*_s)\}\right] \right] \right\}$$

$$\leq np_n^{1/2}||\hat{\beta}_s - \beta^*_s|| \times n \max_{ij} \left\{\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m |Y_{ij} - \mu_{ij}(\beta^*_s)|\right\}$$

$$\leq np_n^{1/2}||\hat{\beta}_s - \beta^*_s|| \times n \max_{ij} \left\{\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m |Y_{ij} - \mu_{ij}(\beta^*_s)|\right\}$$

36
Similarly we can estimate the orders of \( \text{Res}_{123} \) and \( \text{Res}_{124} \).

\[
\text{Res}_{123} = (\tilde{\beta}_s - \beta_s^*)^T \sum_{i=1}^{n} D_i^{(1)} (\tilde{\beta}_s, \beta_s^*, B_s^{D_i})^T V_i^{(1)} (\tilde{\beta}_F, \beta_F^*) \{Y_i - \mu_i(\beta_s^*)\}
\leq ||\tilde{\beta}_s - \beta_s^*|| \times ||\sum_{i=1}^{n} D_i^{(1)} (\tilde{\beta}_s, \beta_s^*, B_s^{D_i})^T V_i^{(1)} (\tilde{\beta}_F, \beta_F^*) \{Y_i - \mu_i(\beta_s^*)\}||
\leq ||\tilde{\beta}_s - \beta_s^*|| \times p_n^{1/2} \times \max \left\{ \sum_{i=1}^{n} D_i^{(1)} (\tilde{\beta}_s, \beta_s^*, B_s^{D_i})^T V_i^{(1)} (\tilde{\beta}_F, \beta_F^*) \{Y_i - \mu_i(\beta_s^*)\} \right\}
\leq ||\tilde{\beta}_s - \beta_s^*|| \times p_n^{1/2} \times ||D_i^{(1)} (\tilde{\beta}_s, \beta_s^*, B_s^{D_i})||_{\text{max}} \times ||V_i^{(1)} (\tilde{\beta}_F, \beta_F^*)||_{\text{max}} \times m^2 \max_{j} \sum_{i=1}^{n} |Y_{ij} - \mu_{ij}(\beta_s^*)|
= O(np_n^{1/2}) \times ||\tilde{\beta}_s - \beta_s^*|| \times ||D_i^{(1)} (\tilde{\beta}_s, \beta_s^*, B_s^{D_i})||_{\text{max}} \times O_p\left( (p_n \log p_n / n)^{1/2} \right)
= n||\tilde{\beta}_s - \beta_s^*||^2 O_F \left( (p_n \log p_n / n)^{1/2} \right);
\]

\[
\text{Res}_{124} = (\tilde{\beta}_s - \beta_s^*)^T \sum_{i=1}^{n} D_i^{(1)} (\tilde{\beta}_s, \beta_s^*, B_s^{D_i})^T V_i^{(2)} (\tilde{\beta}_F, \beta_F^*, B_F^A) \{Y_i - \mu_i(\beta_s^*)\}
\leq ||\tilde{\beta}_s - \beta_s^*||p_n^{1/2} ||D_i^{(1)} (\tilde{\beta}_s, \beta_s^*, B_s^{D_i})||_{\text{max}} \times ||V_i^{(2)} (\tilde{\beta}_F, \beta_F^*, B_F^A)||_{\text{max}} \times m^2 \max_{j} \sum_{i=1}^{n} |Y_{ij} - \mu_{ij}(\beta_s^*)|
= O_p \{np_n^{1/2}\} ||\tilde{\beta}_s - \beta_s^*||^3 \times ||\tilde{\beta}_F - \beta_F^*||
= n||\tilde{\beta}_s - \beta_s^*||^2 O_p \{p_n^{3/2} \log p_n / n\}.
\]

According to Lemma S2.5 both \( ||V_i^* - \hat{V}_i||_{\text{max}} \) and \( ||V_i^* - V_i(\tilde{\beta}_s)||_{\text{max}} \) have the same order. Similar to \( |\text{Res}_{12}| \), we have \( |\text{Res}_{13}| = n||\beta_s^* - \tilde{\beta}_s||^2 o_p(1) \).
Next we analyze the other residual terms.

\[ 2 \text{Res}_2 = (\beta_s - \widetilde{\beta}_s)^T \sum_{i=1}^{n} D_i^{(1)}(\beta_s^*, \widetilde{\beta}_s, B_s^{\mu_i})^T \hat{V}_i^{-1} \{Y_i - \mu_i(\widetilde{\beta}_s)\} \]

\[ = (\beta_s^* - \widetilde{\beta}_s)^T \sum_{i=1}^{n} D_i^{(1)}(\beta_s^*, \widetilde{\beta}_s, B_s^{\mu_i})^T \hat{V}_i^{-1} \{Y_i - \mu_i(\beta_s^*) + \mu_i(\beta_s^*) - \mu_i(\widetilde{\beta}_s)\} \]

\[ = (\beta_s^* - \widetilde{\beta}_s)^T \sum_{i=1}^{n} D_i^{(1)}(\beta_s^*, \widetilde{\beta}_s, B_s^{\mu_i})^T \hat{V}_i^{-1} \{Y_i - \mu_i(\beta_s^*)\} \]

\[ + (\beta_s^* - \widetilde{\beta}_s)^T \sum_{i=1}^{n} D_i^{(1)}(\beta_s^*, \widetilde{\beta}_s, B_s^{\mu_i})^T \hat{V}_i^{-1} \{\mu_i(\beta_s^*) - \mu_i(\widetilde{\beta}_s)\} \]

\[ = \text{Res}_{21} + \text{Res}_{22}. \]

\[ \text{Res}_{21} = (\beta_s^* - \widetilde{\beta}_s)^T \sum_{i=1}^{n} D_i^{(1)}(\beta_s^*, \widetilde{\beta}_s, B_s^{\mu_i})^T \hat{V}_i^{-1} \{Y_i - \mu_i(\beta_s^*)\} \]

\[ = (\beta_s^* - \widetilde{\beta}_s)^T \sum_{i=1}^{n} \{D_i^{(1)}(\beta_s^*, \widetilde{\beta}_s, B_s^{\mu_i}) + D_i^{(2)}(\beta_s^*, \widetilde{\beta}_s, B_s^{\mu_i}, B_s^{\mu_i})\}^T \]

\[ \{V_i^{-1}(\beta_F^*) + \hat{V}_i^{-1}(\beta_F^*, B_s^{A_i})\} \{Y_i - \mu_i(\beta_s^*)\} \]

\[ = (\beta_s^* - \widetilde{\beta}_s)^T \sum_{i=1}^{n} \{D_i^{(1)}(\beta_s^*, \widetilde{\beta}_s, B_s^{\mu_i})V_i^{-1}(\beta_F^*) + D_i^{(2)}(\beta_s^*, \widetilde{\beta}_s, B_s^{\mu_i}, B_s^{\mu_i})V_i^{-1}(\beta_F^*) \]

\[ + D_i^{(1)}(\beta_s^*, \widetilde{\beta}_s, B_s^{\mu_i})\hat{V}_i^{-1}(\beta_F^*, B_s^{A_i}) + D_i^{(2)}(\beta_s^*, \widetilde{\beta}_s, B_s^{\mu_i}, B_s^{\mu_i})\hat{V}_i^{-1}(\beta_F^*, B_s^{A_i})\} \{Y_i - \mu_i(\beta_s^*)\} \]

\[ = \text{Res}_{211} + \text{Res}_{212} + \text{Res}_{213} + \text{Res}_{214}. \]

By similar arguments as above, we are able to show \text{Res}_{211}, \text{Res}_{212}, \text{Res}_{213}, and \text{Res}_{214} are all of the order \( n|\widetilde{\beta}_s - \beta_s^*|^2 o_p(1) \). For \text{Res}_{22}, there exists a \widetilde{\beta}_s between \beta_s^* and \beta_s such that \( \mu_i(\beta_s^*) - \mu_i(\widetilde{\beta}_s) = D_i(\widetilde{\beta}_s)(\beta_s^* - \widetilde{\beta}_s). \)
This entails
\[ |\text{Res}_{22}| = |(\beta_s^* - \beta_s)\sum_{i=1}^{n} D_i^{(1)}(\beta_s^*, \beta_s, B_s^\mu_i)\hat{\nu}_i^{-1}\{\mu_i(\beta_s^*) - \mu_i(\beta_s)\}| \]
\[ = |(\beta_s^* - \beta_s)\sum_{i=1}^{n} D_i^{(1)}(\beta_s^*, \beta_s, B_s^\mu_i)\hat{\nu}_i^{-1} D_i(\beta_s^*)(\beta_s^* - \beta_s)| \]
\[ = n||\beta_s^* - \beta_s||^2 \max_{||v||=1} \left\{ v^T \sum_{i=1}^{n} D_i^{(1)}(\beta_s^*, \beta_s, B_s^\mu_i)\hat{\nu}_i^{-1} D_i(\beta_s^*)v \right\} \]
\[ = n||\beta_s^* - \beta_s||^2 O_p\{(p_n^3 \log p_n^2)/n\}^{1/2}. \]

Combining all the orders for each of the terms, results in the lemma follows.

\[ \square \]

**Lemma S2.8.** For \(s \in S_+\), let \(\eta = n^{-1/2}W(\beta_s^*)^{-1/2}\sum_{i=1}^{n} U_i(\beta_s^*)\). The random vectors \(U_1(\beta_s^*), U_2(\beta_s^*), \ldots, U_n(\beta_s^*)\) are independently distributed random vectors of dimension \(d_s\) with zero mean and satisfy the cumulant boundedness condition. Under Assumptions 1 - 4, \(\log E[e^{\mathbf{t}^T \eta}] \leq a^2 t^T/2\) for \(||\mathbf{t}|| \leq p_n^2 \log p_n\) and some constant \(a^2 > 1\).

**Proof of Lemma S2.8.** For \(s \in S_+\), \(E\{U_i(\beta_s^*)\} = E[D_i(\beta_s^*)^T V_i^{-1}\{Y_i - \mu_i(\beta_s^*)\}] = 0\). Let \(W_i = \text{Cov}\{U_i(\beta_s^*)\}\) be the covariance matrix of \(U_i(\beta_s^*)\) and \(W = \sum_{i=1}^{n} W_i/n\). The cumulant generating function of \(U\) is

\[ C_{U_i(\beta_s^*)}(t) = \log E\{e^{\mathbf{t}^T U_i(\beta_s^*)}\} \]
\[ = C(0) + t^T C^{(1)}(0) + \frac{1}{2} t^T \text{Cov}\{U_i(\beta_s^*)\} t + \frac{1}{6} \sum_{l,r,k} t_l t_r t_k C^{(3)}_{l,r,k}(t^*), \]

39
with a $t^*$ such that $||t^*|| \leq ||t||$. Let $\eta_1 = n^{-1/2} \sum_{i=1}^{n} U_i(\beta_s^*)$, then the cumulant generating function of $\eta_1$ is

$$C_{\eta_1}(t) = \sum_{i=1}^{n} C_{U_i(\beta_s^*)}(\frac{t}{n^{1/2}}) = \sum_{i=1}^{n} \{ \frac{1}{2n} t^T \text{Cov}\{U_i(\beta_s^*)\} t + \sum_{lrk} \frac{1}{6n^{3/2}} t_l t_r t_k C_{(3)}^{lrk}(\frac{t^*}{n^{1/2}}) \} = C_1 + C_2.$$ 

First, $C_1$ can be simplified as $C_1 = \frac{1}{2} t^T \{ \frac{1}{n} \sum_{i=1}^{n} \text{Cov}(U_i) \} t = \frac{1}{2} t^T (\frac{1}{n} \sum_{i=1}^{n} W_i) t = \frac{1}{2} t^T W t$. Next, $C_2$ has the bound as follows:

$$C_2 \leq \frac{b_c}{6n^{1/2} p_n^{3/2}} ||t||^3 = \frac{b_c}{6} (\frac{p_n^3}{n})^{1/2} ||t||^2 = O_p(\{p_n^5 \log p_n/n\}^{1/2} ||t||^2) = o_p(1) ||t||^2.$$ 

This entails

$$C_{\eta_1}(t) \leq \frac{1}{2} a^2 t^T W t,$$

for some $a$ with $a^2 > 1$ and $||t|| \leq \{(p_n^2 \log p_n)^{1/2}\}$. Let $\eta = W^{-1/2} \eta_1$, then the cumulant generating function of $\eta$ is $\log E[e^{t^T \eta}] \leq a^2 t^T t/2$. \qed
SIMULATION RESULTS

We conduct additional simulations on clustered binary responses and clustered Gaussian responses. Table 1 and 2 illustrate the simulation results of QIC and GIC for binary and Gaussian responses under 100 simulated datasets. We let the multiplicative factor $c$ to vary from 1 to 4 and examine the PSR and FDR under different values of $c$. It is observed that when $c = 1$, or 2, the proposed GIC achieves high PSR and low FDR. In comparison, when $c$ increases to 3 or 4, the GIC has much lower PSR. Table 3 provides additional simulation results for binary and Gaussian responses with the cluster size $m = 20$. The performance of the proposed GIC improves with higher PSR and lower FDR with larger cluster size.

References


Table 1: The PSR and FDR of GIC and QIC for Binary Response under different values of the multiplicative factor $c$

<table>
<thead>
<tr>
<th></th>
<th>n 1000</th>
<th>p 1000</th>
<th>n 1000</th>
<th>p 500</th>
<th>n 500</th>
<th>p 500</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>std</td>
<td>mean</td>
<td>std</td>
<td>mean</td>
<td>std</td>
</tr>
<tr>
<td></td>
<td>PSR</td>
<td>PSR</td>
<td>FDR</td>
<td>FDR</td>
<td>PSR</td>
<td>PSR</td>
</tr>
<tr>
<td>QIC(I)</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.7093</td>
<td>0.0241</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>QIC(E)</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.7099</td>
<td>0.0241</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>QIC(A)</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.7093</td>
<td>0.0241</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>QIC(U)</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.7234</td>
<td>0.0179</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

The true parameters size $d_F$ is 50 and the cluster size $m$ is 10. For working correlations, "I" represents independent, "E" represents exchangeable, "A" represents AR1, and "U" represents unstructured.
Table 2: The PSR and FDR of GIC and QIC for Gaussian Response under different values of the multiplicative factor $c$

<table>
<thead>
<tr>
<th></th>
<th>$n$ 1000</th>
<th>$p$ 1000</th>
<th></th>
<th>$n$ 1000</th>
<th>$p$ 500</th>
<th></th>
<th>$n$ 500</th>
<th>$p$ 500</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>std</td>
<td>mean</td>
<td>std</td>
<td>mean</td>
<td>std</td>
<td>mean</td>
<td>std</td>
<td>mean</td>
</tr>
<tr>
<td>PSR</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.6937</td>
<td>0.0326</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.5308</td>
<td>0.0652</td>
<td>1.0000</td>
</tr>
<tr>
<td>FDR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QIC(E)</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.6937</td>
<td>0.0326</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.5381</td>
<td>0.0648</td>
<td>1.0000</td>
</tr>
<tr>
<td>QIC(A)</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.6937</td>
<td>0.0326</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.5308</td>
<td>0.0652</td>
<td>1.0000</td>
</tr>
<tr>
<td>QIC(U)</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.7281</td>
<td>0.0136</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.7077</td>
<td>0.0354</td>
<td>1.0000</td>
</tr>
<tr>
<td>GIC(I,c=1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.1077</td>
<td>0.0460</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.0784</td>
<td>0.0415</td>
<td>1.0000</td>
</tr>
<tr>
<td>GIC(E,c=1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIC(A,c=1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIC(U,c=1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIC(I,c=2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.0012</td>
<td>0.0047</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.0024</td>
<td>0.0065</td>
<td>1.0000</td>
</tr>
<tr>
<td>GIC(E,c=2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIC(A,c=2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIC(U,c=2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIC(I,c=3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.0004</td>
<td>0.0028</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.0004</td>
<td>0.0028</td>
<td>1.0000</td>
</tr>
<tr>
<td>GIC(E,c=3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIC(A,c=3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIC(U,c=3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIC(I,c=4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>GIC(E,c=4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIC(A,c=4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIC(U,c=4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIC(U,c=1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.0226</td>
<td>0.0295</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.0116</td>
<td>0.0220</td>
<td>1.0000</td>
</tr>
<tr>
<td>GIC(U,c=2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIC(U,c=3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIC(U,c=4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The true parameters size $d_p$ is 50 and the cluster size $m$ is 10. For working correlations, "I" represents independent, "E" represents exchangeable, "A" represents AR1, and "U" represents unstructured.
Table 3: The PSR and FDR of GIC and QIC for Binary and Gaussian Responses with
the cluster size $m = 20$

<table>
<thead>
<tr>
<th></th>
<th>Binary</th>
<th>Gaussian</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n 500 p 500</td>
<td>n 500 p 500</td>
</tr>
<tr>
<td></td>
<td>mean std</td>
<td>mean std</td>
</tr>
<tr>
<td></td>
<td>PSR FDR</td>
<td>PSR FDR</td>
</tr>
<tr>
<td>QIC(I)</td>
<td>1.0000 0.0000</td>
<td>0.5482 0.0629</td>
</tr>
<tr>
<td>QIC(E)</td>
<td>1.0000 0.0000</td>
<td>0.5490 0.0634</td>
</tr>
<tr>
<td>QIC(A)</td>
<td>1.0000 0.0000</td>
<td>0.5490 0.0634</td>
</tr>
<tr>
<td>QIC(U)</td>
<td>1.0000 0.0000</td>
<td>0.5978 0.0553</td>
</tr>
<tr>
<td></td>
<td>0.9986 0.0051</td>
<td>0.0452 0.0375</td>
</tr>
<tr>
<td></td>
<td>0.9708 0.0380</td>
<td>0.0019 0.0070</td>
</tr>
<tr>
<td></td>
<td>0.8952 0.0755</td>
<td>0.0000 0.0000</td>
</tr>
<tr>
<td></td>
<td>0.7966 0.1006</td>
<td>0.0000 0.0000</td>
</tr>
<tr>
<td>GIC(I,c=1)</td>
<td>1.0000 0.0000</td>
<td>0.0456 0.0375</td>
</tr>
<tr>
<td>GIC(E,c=2)</td>
<td>0.9986 0.0051</td>
<td>0.0019 0.0070</td>
</tr>
<tr>
<td>GIC(A,c=3)</td>
<td>0.8952 0.0755</td>
<td>0.0000 0.0000</td>
</tr>
<tr>
<td>GIC(U,c=4)</td>
<td>0.7966 0.1006</td>
<td>0.0000 0.0000</td>
</tr>
<tr>
<td>GIC(E,c=1)</td>
<td>1.0000 0.0000</td>
<td>0.0775 0.0409</td>
</tr>
<tr>
<td>GIC(E,c=2)</td>
<td>0.9986 0.0051</td>
<td>0.0036 0.0086</td>
</tr>
<tr>
<td>GIC(A,c=3)</td>
<td>0.8952 0.0755</td>
<td>0.0000 0.0000</td>
</tr>
<tr>
<td>GIC(U,c=4)</td>
<td>0.7966 0.1006</td>
<td>0.0000 0.0000</td>
</tr>
<tr>
<td>GIC(A,c=1)</td>
<td>0.9986 0.0051</td>
<td>0.0448 0.0375</td>
</tr>
<tr>
<td>GIC(A,c=2)</td>
<td>0.9706 0.0379</td>
<td>0.0019 0.0070</td>
</tr>
<tr>
<td>GIC(A,c=3)</td>
<td>0.8952 0.0755</td>
<td>0.0000 0.0000</td>
</tr>
<tr>
<td>GIC(A,c=4)</td>
<td>0.7966 0.1006</td>
<td>0.0000 0.0000</td>
</tr>
<tr>
<td>GIC(U,c=1)</td>
<td>0.9992 0.0039</td>
<td>0.0346 0.0370</td>
</tr>
<tr>
<td>GIC(U,c=2)</td>
<td>0.9916 0.0187</td>
<td>0.0020 0.0065</td>
</tr>
<tr>
<td>GIC(U,c=3)</td>
<td>0.9562 0.0525</td>
<td>0.0004 0.0028</td>
</tr>
<tr>
<td>GIC(U,c=4)</td>
<td>0.9012 0.0733</td>
<td>0.0000 0.0000</td>
</tr>
</tbody>
</table>

The true parameters size $d_p$ is 50 and the cluster size $m$ is 20. For working correlations, "I" represents independent, "E" represents exchangeable, "A" represents AR1, and "U" represents unstructured.