# Network Influence Analysis

Tao Zou, Ronghua Luo, Wei Lan and Chih-Ling Tsai

*The Australian National University, Southwestern University of Finance and Economics and University of California, Davis*

## Supplementary Material

This supplementary material consists of four parts. Section S.1 introduces five technical lemmas and their proofs. It is worth noting that the detailed expression of $\partial\ell(\theta)/\partial\theta$ used in the quasi-score test is given in equation (S1.6) in the proof of Lemma 5. Section S.2 discusses the technical conditions in Appendix. Section S.3 presents the proofs of theorems and corollaries. Section S.4 provides simulation studies to assess the robustness of our proposed parameter estimate, and additional empirical results.

## S1 Five Technical Lemmas

Before providing the technical lemmas, let $\|\cdot\|_s$ denote the vector $s$-norm or the matrix $s$-norm for $1 \le s \le \infty$. In other words, for any generic vector $x = (x_1, \cdots, x_q)^\top \in \mathbb{R}^q$, $\|x\|_s = (\sum_{i=1}^q |x_i|^s)^{1/s}$, and, for any generic matrix $G \in \mathbb{R}^{m\times q}$,

$$\|G\|_s = \sup\left\{\frac{\|Gx\|_s}{\|x\|_s} : x \in \mathbb{R}^{q\times 1} \text{ and } x \neq 0\right\}.$$

Moreover, define the element-wise $\ell_\infty$ norm for any generic matrix $G$ as $|G|_\infty = \|\mathrm{vec}(G)\|_\infty$, where $\mathrm{vec}(G)$ denotes the vectorization for any generic matrix $G$. In addition, we denote the Frobenius norm of any generic matrix $G$ as $\|G\|_F = \|\text{vec}(G)\|_2$. Since Lemma 3 is directly modified from Theorem 1 of Kelejian and Prucha (2001), we only present the proofs of the rest of four lemmas.

**Lemma 1.** *For any generic vector $x = (x_1, \cdots, x_q)^\top \in \mathbb{R}^q$ and for $1 \leq s_1 \leq s_2$, we have that*

$$\|x\|_{s_1} \leq \|x\|_{s_2} q^{1/s_1 - 1/s_2} \text{ and } \left|\sum_{j=1}^{q} x_j\right|^{s_1} \leq q^{s_1-1} \sum_{j=1}^{q} |x_j|^{s_1}.$$

**Proof.** By Hölder's inequality, we obtain

$$|x_1|^{s_1} + \cdots + |x_q|^{s_1} \leq (|x_1|^{s_2} + \cdots + |x_q|^{s_2})^{s_1/s_2} q^{1-s_1/s_2}.$$

Hence,

$$\|x\|_{s_1} \leq \|x\|_{s_2} q^{1/s_1 - 1/s_2}.$$

In addition, by Jensen's inequality (Durrett, 2010, p. 23, Theorem 1.5.1), we have

$$\left\{\frac{1}{q}(|x_1| + \cdots |x_q|)\right\}^{s_1} \leq \frac{1}{q}(|x_1|^{s_1} + \cdots |x_q|^{s_1}).$$

The above results, together with triangular inequality, imply

$$|x_1 + \cdots + x_q|^{s_1} \leq (|x_1| + \cdots + |x_q|)^{s_1} \leq q^{s_1-1}(|x_1|^{s_1} + \cdots + |x_q|^{s_q}),$$

which completes the proof.

**Lemma 2.** *For any generic vector $x = (x_1, \cdots, x_q)^\top \in \mathbb{R}^q$ and generic matrices $G \in \mathbb{R}^{m\times q}$ and $U \in \mathbb{R}^{m\times m}$ and for any $s \geq 1$, we have that $\|UGx\|_s \leq m^{1/s}|G|_\infty \|U\|_\infty \|x\|_1$.*

**Proof.** Let $U = (u_{ij})_{m\times m} \in \mathbb{R}^{m\times m}$ and $G = (g_{ij})_{m\times q} \in \mathbb{R}^{m\times q}$. For any $q$-dimensional real vector $x = (x_1, \cdots, x_q)^\top$, we have

$$\begin{aligned}
\left|\sum_{j=1}^{m}\sum_{k=1}^{q} u_{ij}g_{jk}x_k\right| &\leq \sum_{j=1}^{m}\sum_{k=1}^{q} |u_{ij}|\,|g_{jk}|\,|x_k| = \sum_{k=1}^{q}|x_k|\sum_{j=1}^{m}|u_{ij}|\,|g_{jk}| \\
\leq \quad & |G|_\infty \sum_{k=1}^{q}|x_k| \max_{1\leq i\leq m}\sum_{j=1}^{m}|u_{ij}| = |G|_\infty \|U\|_\infty \|x\|_1.
\end{aligned}$$

Then

$$\|UGx\|_s = \left\{\sum_{i=1}^{m}\left|\sum_{j=1}^{m}\sum_{k=1}^{q} u_{ij}g_{jk}x_k\right|^s\right\}^{1/s} \leq m^{1/s}|G|_\infty\|U\|_\infty \|x\|_1 ,$$

which completes the proof.

**Lemma 3.** *Let $\mathcal{E} = (\varepsilon_1, \cdots, \varepsilon_n)^\top$, where $\varepsilon_1, \cdots, \varepsilon_n$ are independent and identically distributed random variables with mean 0 and finite variance $\sigma^2$. Define*

$$Q_n = \mathcal{E}^\top A\mathcal{E} + b^\top\mathcal{E} - \sigma^2\mathrm{tr}(A),$$

*where $A = (a_{ij})_{n\times n} \in \mathbb{R}^{n\times n}$ and $b = (b_1, \cdots, b_n)^\top \in \mathbb{R}^{n\times 1}$. Suppose the following assumptions are satisfied:*

*(1) for $i, j = 1, \cdots, n$, $a_{ij} = a_{ji}$;*

*(2) $\sup_{n\geq 1}\|A\|_1 < \infty$;*

*(3) for some $\eta_1 > 0$, $\sup_{n\geq 1} n^{-1}\|b\|_{2+\eta_1}^{2+\eta_1} < \infty$;*

*(4) for some $\eta_2 > 0$, $\mathrm{E}|\varepsilon_i|^{4+\eta_2} < \infty$.*

*Then, we have $\mathrm{E}(Q_n) = 0$ and*

$$\sigma_{Q_n}^2 := \mathrm{var}(Q_n) = 4\sigma^4\sum_{i=1}^{n}\sum_{j=1}^{i-1}a_{ij}^2 + \sigma^2\sum_{i=1}^{n}b_i^2 + \sum_{i=1}^{n}\left[\left\{\mu^{(4)} - \sigma^4\right\}a_{ii}^2 + 2\mu^{(3)}b_ia_{ii}\right],$$

*where* $\mu^{(s)} = \mathrm{E}(\varepsilon_i^s)$ *for* $s = 3, 4$. *Furthermore, suppose*

(5) $n^{-1}\sigma_{Q_n}^2 \geq c$ *for some* $c > 0$.

*Then, we obtain*

$$\sigma_{Q_n}^{-1} Q_n \xrightarrow{d} N(0,1).$$

**Lemma 4.** *Let* $\mathcal{E} = (\varepsilon_1, \cdots, \varepsilon_n)^\top$, *where* $\varepsilon_1, \cdots, \varepsilon_n$ *are independent and identically distributed with mean 0 and finite variance* $\sigma^2$. *Define*

$$\mathcal{Q}_n = \begin{pmatrix} \mathrm{vec}^\top(A_1) \\ \vdots \\ \mathrm{vec}^\top(A_L) \end{pmatrix} \mathrm{vec}\left(\mathcal{E}\mathcal{E}^\top - \sigma^2 I_n\right) + B^\top \mathcal{E}$$

*where* $A_l = (a_{ij}^{(l)})_{n\times n} \in \mathbb{R}^{n\times n}$ *for* $l = 1, \cdots, L$ *with* $L < \infty$ *and* $B = (b_{il})_{n\times L} \in \mathbb{R}^{n\times L}$. *Suppose that*

(1) *for all* $i, j = 1, \cdots, n$ *and* $l = 1, \cdots, L$, $a_{ij}^{(l)} = a_{ji}^{(l)}$;

(2) $\sup_{n\geq 1} \|A_l\|_1 < \infty$ *for any* $l = 1, \cdots, L$;

(3) *for some* $\eta_1 > 0$, $\sup_{n\geq 1} n^{-1}\|\mathrm{vec}(B)\|_{2+\eta_1}^{2+\eta_1} < \infty$;

(4) *for some* $\eta_2 > 0$, $\mathrm{E}|\varepsilon_i|^{4+\eta_2} < \infty$.

*Then, we have* $\mathrm{E}\mathcal{Q}_n = 0$ *and*

$$\mathrm{Cov}\left(\mathcal{Q}_n\right) = 2\sigma^4\big(\mathrm{tr}(A_{l_1}A_{l_2})\big)_{L\times L} + \sigma^2 B^\top B + \left\{\mu^{(4)} - 3\sigma^4\right\}\Psi^\top\Psi + \mu^{(3)}\left\{\Psi^\top B + B^\top \Psi\right\},$$

*where* $\Psi = (\psi_1, \cdots, \psi_L) \in \mathbb{R}^{n\times L}$ *with* $\psi_l = (a_{11}^{(l)}, \cdots, a_{nn}^{(l)})^\top \in \mathbb{R}^n$ *for* $l = 1, \cdots, L$, *and* $\mu^{(s)} = \mathrm{E}(\varepsilon_i^s)$ *for* $s = 3, 4$. *Moreover,* $n^{-1/2-\epsilon}\mathcal{Q}_n \xrightarrow{L_2} 0$ *for any* $\epsilon > 0$. *In addition, assume that*

*(5) there exists a positive definite matrix $\mathscr{Q} \in \mathbb{R}^{L\times L}$ such that $n^{-1}\mathrm{Cov}(\mathcal{Q}_n) \to \mathscr{Q}$. Then we obtain*

$$n^{-1/2}\mathcal{Q}_n \xrightarrow{d} N(0, \mathscr{Q}).$$

**Proof.** For any generic vector $t = (t_1, \cdots, t_L)^\top \neq 0 \in \mathbb{R}^L$, let $D_n = t^\top \mathcal{Q}_n = \mathcal{E}^\top A(t)\mathcal{E} + t^\top B^\top \mathcal{E} - \sigma^2 \mathrm{tr}\{A(t)\}$, where $A(t) = \sum_{l=1}^L t_l A_l = \left(\sum_{l=1}^L t_l a_{ij}^{(l)}\right)_{n\times n}$. For the sake of simplicity, we denote $a_{ij}(t) = \sum_{l=1}^L t_l a_{ij}^{(l)}$ so that $A(t) = \left(a_{ij}(t)\right)_{n\times n}$. By Conditions (1) and (2) in this lemma, we have that $A(t)$ is a symmetric matrix and $\sup_{n\geq 1}\|A(t)\|_1 < \infty$, respectively. Note that, by Lemma 1, we have that, for $\eta_1 > 0$,

$$\|Bt\|_{2+\eta_1}^{2+\eta_1} = \sum_{i=1}^n \left|\sum_{l=1}^L b_{il}t_l\right|^{2+\eta_1} \leq L^{1+\eta_1} \max_{1\leq l\leq L} |t_l|^{2+\eta_1} \|\mathrm{vec}(B)\|_{2+\eta_1}^{2+\eta_1}.$$

This, together with Condition (3) in this lemma, implies that

$$\sup_{n\geq 1} n^{-1}\|Bt\|_{2+\eta_1}^{2+\eta_1} \leq L^{1+\eta_1} \max_{1\leq l\leq L} |t_l|^{2+\eta_1} \sup_{n\geq 1} n^{-1} \|\mathrm{vec}(B)\|_{2+\eta_1}^{2+\eta_1} < \infty.$$

As a consequence, $D_n$ satisfies Conditions (1)-(4) in Lemma 3, which leads to $\mathrm{E}(D_n) = 0$ and $\mathrm{E}(\mathcal{Q}_n) = 0$. Re-express $B^\top = (b_1, \cdots, b_n)$, where $b_i$ is a $L\times 1$ vector for $i = 1, \cdots, n$. By Lemma 3 we then obtain that

$$\begin{aligned}
\sigma_{D_n}^2 &:= \mathrm{var}(D_n) = 4\sigma^4 \sum_{i=1}^n \sum_{j=1}^{i-1} a_{ij}^2(t) + \sum_{i=1}^n \left[\left\{\mu^{(4)} - \sigma^4\right\} a_{ii}^2(t) + 2\mu^{(3)} t^\top b_i a_{ii}(t)\right] \\
&\quad + \sigma^2 \sum_{i=1}^n (t^\top b_i)^2 \\
&= 2\sigma^4 \left\{\sum_{i,j=1}^n a_{ij}^2(t) - \sum_{i=1}^n a_{ii}^2(t)\right\} + \sigma^2 t^\top \left(\sum_{i=1}^n b_i b_i^\top\right) t \\
&\quad + \left\{\mu^{(4)} - \sigma^4\right\} \sum_{i=1}^n a_{ii}^2(t) + 2\mu^{(3)} t^\top \sum_{i=1}^n b_i a_{ii}(t),
\end{aligned}$$

where

$$2\sigma^4 \sum_{i,j=1}^{n} a_{ij}^2(t) = 2\sigma^4 \sum_{l_1,l_2=1}^{L} t_{l_1} t_{l_2} \sum_{i,j=1}^{n} a_{ij}^{(l_1)} a_{ij}^{(l_2)} = 2\sigma^4 t^\top \left(\mathrm{tr}(A_{l_1} A_{l_2})\right)_{L\times L} t,$$

$$\left\{\mu^{(4)} - 3\sigma^4\right\} \sum_{i=1}^{n} a_{ii}^2(t) = \left\{\mu^{(4)} - 3\sigma^4\right\} \sum_{l_1,l_2=1}^{L} t_{l_1} t_{l_2} \sum_{i=1}^{n} a_{ii}^{(l_1)} a_{ii}^{(l_2)} = \left\{\mu^{(4)} - 3\sigma^4\right\} t^\top \Psi^\top \Psi t, \text{ and}$$

$$2\mu^{(3)} t^\top \sum_{i=1}^{n} b_i a_{ii}(t) = 2\mu^{(3)} \sum_{l_1,l_2=1}^{L} t_{l_1} t_{l_2} \sum_{i=1}^{n} b_{il_1} a_{ii}^{(l_2)} = 2\mu^{(3)} t^\top B^\top \Psi t.$$

Accordingly, we have

$$\sigma_{D_n}^2 = t^\top \left\{2\sigma^4 \left(\mathrm{tr}(A_{l_1} A_{l_2})\right)_{L\times L} + \sigma^2 B^\top B + \left\{\mu^{(4)} - 3\sigma^4\right\} \Psi^\top \Psi + \mu^{(3)} B^\top \Psi + \mu^{(3)} \Psi^\top B\right\} t. \tag{S1.1}$$

This implies that

$$\mathrm{Cov}(\mathcal{Q}_n) = 2\sigma^4 \left(\mathrm{tr}(A_{l_1} A_{l_2})\right)_{L\times L} + \sigma^2 B^\top B + \left\{\mu^{(4)} - 3\sigma^4\right\} \Psi^\top \Psi + \mu^{(3)} B^\top \Psi + \mu^{(3)} \Psi^\top B.$$

Using the fact that $A_l$ is symmetric, we have $\|A_l\|_\infty = \|A_l\|_1$ for any $l \in \{1, \cdots, L\}$. By 4.67 (e) in Seber (2008, p. 69), we obtain

$$\|A_l\|_2 \le \sqrt{\|A_l\|_1 \|A_l\|_\infty}.$$

This, together with Condition (2) in this lemma, implies that $\sup_{n\ge 1} \|A_l\|_2 < \infty$. In addition, applying Lemmas 1 and 2 of Zou et al. (2017), we have

$$n^{-1}\mathrm{tr}(A_{l_1} A_{l_2}) = O(1) \text{ for any } l_1, l_2 \in \{1, \cdots, L\}. \tag{S1.2}$$

Moreover, by Lemma 1 and Condition (3) in this lemma, we have

$$\sup_{n\geq 1} n^{-1} \|\mathrm{vec}(B)\|_2^2 \leq L^{1-2/(2+\eta_1)} \left(\sup_{n\geq 1} n^{-1} \|\mathrm{vec}(B)\|_{2+\eta_1}^{2+\eta_1}\right)^{2/(2+\eta_1)} < \infty.$$

This result, together with 4.67 (a) in Seber (2008, p. 68), implies

$$\sup_{n\geq 1} n^{-1} \left\|B^\top B\right\|_2 \leq \sup_{n\geq 1} n^{-1} \|B\|_2^2 \leq \sup_{n\geq 1} n^{-1} \|\mathrm{vec}(B)\|_2^2 < \infty. \tag{S1.3}$$

By 4.67 (a) in Seber (2008, p. 68), we also have

$$\sup_{n\geq 1} n^{-1} \|\Psi\|_2^2 \leq \sup_{n\geq 1} n^{-1} \|\mathrm{vec}(\Psi)\|_2^2 \leq \sup_{n\geq 1} n^{-1} \sum_{l=1}^{L} \|A_l\|_F^2 \leq \sum_{l=1}^{L} \left(\sup_{n\geq 1} \|A_l\|_2\right)^2 < \infty,$$

which leads to

$$\sup_{n\geq 1} n^{-1} \left\|\Psi^\top \Psi\right\|_2 \leq \sup_{n\geq 1} n^{-1} \|\Psi\|_2^2 < \infty \text{ and} \tag{S1.4}$$

$$\sup_{n\geq 1} n^{-1} \left\|B^\top \Psi\right\|_2 \leq \sup_{n\geq 1} n^{-1/2} \|B\|_2 \sup_{n\geq 1} n^{-1/2} \|\Psi\|_2 < \infty. \tag{S1.5}$$

By (S1.2)-(S1.5), we obtain $\mathrm{Cov}(n^{-1/2-\epsilon}\mathcal{Q}_n) \to 0$. This implies

$$n^{-1/2-\epsilon}\mathcal{Q}_n \xrightarrow{L_2} 0,$$

which completes the proof of the first part of Lemma 4.

We next show the second part of this lemma. By Condition (5) in this lemma and equation (S1.1), we obtain

$$\frac{1}{n}\sigma_{D_n}^2 \longrightarrow t^\top \mathscr{Q} t > 0.$$

Furthermore, employing the results in Lemma 3, we have

$$n^{-1/2}\frac{t^\top \mathcal{Q}_n}{\sqrt{t^\top \mathscr{Q} t}} = \frac{D_n}{\sqrt{\sigma^2_{D_n}}}\sqrt{\frac{\frac{1}{n}\sigma^2_{D_n}}{t^\top \mathscr{Q} t}} \xrightarrow{d} N(0,1),$$

which leads to

$$n^{-1/2}t^\top \mathcal{Q}_n \xrightarrow{d} N(0, t^\top \mathscr{Q} t).$$

This, together with the Cramér-Wold device, implies that

$$n^{-1/2}\mathcal{Q}_n \xrightarrow{d} N(0, \mathscr{Q}),$$

which completes the entire proof of this lemma.

**Lemma 5.** *Under Conditions (C1)-(C5) in Appendix, we have that, as $n \to \infty$,*

$$\text{(i) } n^{-1/2}\frac{\partial \ell(\theta)}{\partial \theta} \xrightarrow{d} N\left(0, \mathcal{I}(\theta) + \mathcal{J}(\theta, \mu^{(3)}, \mu^{(4)})\right) \text{ and (ii) } -n^{-1}\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta^\top} \xrightarrow{\mathrm{P}} \mathcal{I}(\theta).$$

**Proof.** We firstly prove the part (i) of Lemma 5. After tedious calculation, we have

$$\frac{\partial \ell(\theta)}{\partial \theta} = \begin{pmatrix} \mathrm{vec}^\top(A_1) \\ \vdots \\ \mathrm{vec}^\top(A_L) \end{pmatrix} \mathrm{vec}\left\{\mathcal{E}(\alpha,\beta)\mathcal{E}(\alpha,\beta)^\top - \sigma^2 I_n\right\} + B^\top \mathcal{E}(\alpha,\beta), \tag{S1.6}$$

where $L = p + d + 1$, $A_l = 0_{n\times n}$ for $l = 1, \cdots p$, $A_{p+k} = 2^{-1}\sigma^{-2}\{W\Lambda_{\beta_k}(\beta)S^{-1}(\beta) + S^{-1}(\beta)^\top \Lambda_{\beta_k}(\beta)W^\top\}$ with $\Lambda_{\beta_k}(\beta) := \partial\Lambda(\beta)/\partial\beta_k = \mathrm{diag}\{z_{1k}F'(Z_1^\top\beta), \cdots, z_{nk}F'(Z_n^\top\beta)\}$

for $k = 1, \cdots, d$, $A_{p+d+1} = 2^{-1}\sigma^{-4} I_n$,

$$B^\top = \begin{pmatrix} \frac{1}{\sigma^2}\mathbb{X}^\top \\ \frac{1}{\sigma^2}\alpha^\top \mathbb{X}^\top S^{-1}(\beta)^\top \Lambda_{\beta_1}(\beta) W^\top \\ \vdots \\ \frac{1}{\sigma^2}\alpha^\top \mathbb{X}^\top S^{-1}(\beta)^\top \Lambda_{\beta_d}(\beta) W^\top \\ 0_{1\times n} \end{pmatrix} \text{ and}$$

$\mathcal{E}(\alpha, \beta) = S(\beta)\mathbb{Y} - \mathbb{X}\alpha$. It is obvious that Condition (1) in Lemma 4 is satisfied.

Note that, for $l = 1, \cdots, p$, we have $\|A_l\|_1 = 0$. In addition, for $k = 1, \cdots, d$,

$$\|\Lambda_{\beta_k}(\beta)\|_1 = \max_{1\le i\le n} \left| z_{ik} F'(Z_i^\top \beta) \right|,$$

$$\|A_{p+k}\|_1 \le \frac{1}{2\sigma^2}\left\{ \|W\|_1 \|\Lambda_{\beta_k}(\beta)\|_1 \left\|S^{-1}(\beta)\right\|_1 + \left\|S^{-1}(\beta)\right\|_\infty \|\Lambda_{\beta_k}(\beta)\|_1 \|W\|_\infty \right\}, \text{ and}$$

$$\|A_{p+d+1}\| = \frac{1}{2\sigma^4}\|I_n\|_1 = \frac{1}{2\sigma^4}.$$

By Conditions (C2)-(C4), we obtain

$$\sup_{n\ge 1} \|A_l\|_1 < \infty, \text{ for } l = 1, \cdots, (p+d+1). \tag{S1.7}$$

Hence, Condition (2) in Lemma 4 is satisfied.

By Lemma 1 and Lemma 2, we have that, for any $\eta_1 > 0$,

$$\begin{aligned} n^{-1}\|\text{vec}(B)\|_{2+\eta_1}^{2+\eta_1} &= n^{-1}\frac{1}{\sigma^{2(2+\eta_1)}}\left\{ \|\text{vec}(\mathbb{X})\|_{2+\eta_1}^{2+\eta_1} + \sum_{k=1}^d \left\| W\Lambda_{\beta_k}(\beta) S^{-1}(\beta)\mathbb{X}\alpha \right\|_{2+\eta_1}^{2+\eta_1} \right\} \\ \le\ & \frac{1}{\sigma^{2(2+\eta_1)}}\left\{ p\,|\mathbb{X}|_\infty^{2+\eta_1} + \|\alpha\|_1^{2+\eta_1} |\mathbb{X}|_\infty^{2+\eta_1} \|W\|_\infty^{2+\eta_1} \left\|S^{-1}(\beta)\right\|_\infty^{2+\eta_1} \sum_{k=1}^d \|\Lambda_{\beta_k}(\beta)\|_\infty^{2+\eta_1} \right\}. \end{aligned}$$

By Conditions (C2)-(C4) and employing similar techniques to those used for verifying (S1.7), we obtain $\sup_{n\geq 1} n^{-1}\|\mathrm{vec}(B)\|_{2+\eta_1}^{2+\eta_1} < \infty$. Thus, Condition (3) in Lemma 4 is satisfied. By Condition (C1), Condition (4) in Lemma 4 is also satisfied.

Lastly, by the first part of Lemma 4 and Condition (C5), we obtain $\mathrm{E}\{\partial\ell(\theta)/\partial\theta\} = 0$ and

$$\frac{1}{n}\mathrm{Cov}\left\{\frac{\partial\ell(\theta)}{\partial\theta}\right\} = \mathcal{I}_n(\theta) + \mathcal{J}_n(\theta,\mu^{(3)},\mu^{(4)}) \to \mathcal{I}(\theta) + \mathcal{J}(\theta,\mu^{(3)},\mu^{(4)}).$$

which implies that Condition (5) in Lemma 4 is satisfied. In sum, we have verified that $\partial\ell(\theta)/\partial\theta$ satisfies Conditions (1)-(5) in Lemma 4, which lead to the first part of Lemma 5.

We next demonstrate the second part of Lemma 5. After tedious calculations, we have that

$$\mathrm{E}\frac{\partial^2\ell(\theta)}{\partial\alpha\partial\alpha^\top} = -\frac{1}{\sigma^2}\mathbb{X}^\top\mathbb{X}, \quad \mathrm{E}\frac{\partial^2\ell(\theta)}{\partial\alpha\partial\beta_k} = -\frac{1}{\sigma^2}\mathbb{X}^\top W\Lambda_{\beta_k}(\beta)S^{-1}(\beta)\mathbb{X}\alpha, \quad \mathrm{E}\frac{\partial^2\ell(\theta)}{\partial\alpha\partial\sigma^2} = 0_{p\times 1},$$

$$\begin{aligned}\mathrm{E}\frac{\partial^2\ell(\theta)}{\partial\beta_{k_1}\partial\beta_{k_2}} \quad = \quad & -\mathrm{tr}\left\{W\Lambda_{\beta_{k_1}}(\beta)S^{-1}(\beta)W\Lambda_{\beta_{k_2}}(\beta)S^{-1}(\beta)\right\} \\ & -\mathrm{tr}\left\{W\Lambda_{\beta_{k_1}}(\beta)S^{-1}(\beta)S^{-1}(\beta)^\top\Lambda_{\beta_{k_2}}(\beta)W^\top\right\} \\ & -\frac{1}{\sigma^2}\alpha^\top\mathbb{X}^\top S^{-1}(\beta)^\top\Lambda_{\beta_{k_1}}(\beta)W^\top W\Lambda_{\beta_{k_2}}(\beta)S^{-1}(\beta)\mathbb{X}\alpha,\end{aligned}$$

$$\mathrm{E}\frac{\partial^2\ell(\theta)}{\partial\beta_k\partial\sigma^2} = -\frac{1}{\sigma^2}\mathrm{tr}\left\{W\Lambda_{\beta_k}(\beta)S^{-1}(\beta)\right\} \text{ and } \mathrm{E}\frac{\partial^2\ell(\theta)}{\partial\sigma^2\partial\sigma^2} = -\frac{n}{2\sigma^4}$$

for $k, k_1, k_2 \in \{1, \cdots, d\}$. Accordingly, we have

$$-n^{-1}\mathrm{E}\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta^\top} = \mathcal{I}_n(\theta).$$

We next prove

$$-n^{-1}\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta^\top} = \mathcal{I}_n(\theta) + o_{\mathrm{P}}(1)$$

via the following six steps, which correspond to their components of $\theta$ and $\theta^\top$ in the Fisher information matrix.

STEP I. It can be seen straightforwardly that

$$-n^{-1}\frac{\partial^2 \ell(\theta)}{\partial\alpha\partial\alpha^\top} = -n^{-1}\mathrm{E}\frac{\partial^2 \ell(\theta)}{\partial\alpha\partial\alpha^\top}.$$

In the next five steps, we will apply Lemma 4 to show the desired results. Note that Condition (C1) in this lemma indicates that Condition (4) in Lemma 4 holds. Hence, we only check Conditions (1)-(3) of Lemma 4.

STEP II. For any $k = 1, \cdots d$, we have

$$-n^{-1}\frac{\partial^2 \ell(\theta)}{\partial\alpha\partial\beta_k} = n^{-1}\sigma^{-2}\mathbb{X}^\top W\Lambda_{\beta_k}(\beta)S^{-1}(\beta)\mathcal{E}(\alpha,\beta) - n^{-1}\mathrm{E}\frac{\partial^2 \ell(\theta)}{\partial\alpha\partial\beta_k}. \tag{S1.8}$$

We employ Lemma 4 by defining $A_l = 0_{n\times n}$ for $l = 1, \cdots, p$, and $B = \sigma^{-2}S^{-1}(\beta)^\top \times\Lambda_{\beta_k}(\beta)W^\top\mathbb{X}$. Accordingly, Conditions (1)-(2) in Lemma 4 are satisfied. Furthermore, by Lemma 2, we have, for any $\eta_1 > 0$,

$$n^{-1}\left\|\mathrm{vec}\left\{S^{-1}(\beta)^\top \Lambda_{\beta_k}(\beta) W^\top \mathbb{X}\right\}\right\|_{2+\eta_1}^{2+\eta_1} = n^{-1}\sum_{j=1}^{p}\left\|S^{-1}(\beta)^\top \Lambda_{\beta_k}(\beta) W^\top \mathbb{X}_j\right\|_{2+\eta_1}^{2+\eta_1}$$
$$\leq \ p\left\|S^{-1}(\beta)^\top\right\|_\infty^{2+\eta_1}\left\|\Lambda_{\beta_k}(\beta)\right\|_\infty^{2+\eta_1}\left\|W^\top\right\|_\infty^{2+\eta_1}|\mathbb{X}|_\infty^{2+\eta_1}.$$

Subsequently, using Conditions (C2)-(C4) and employing similar techniques to those used for verifying (S1.7), we obtain

$$\sup_{n\geq 1} n^{-1}\left\|\mathrm{vec}\left\{S^{-1}(\beta)^\top \Lambda_{\beta_k}(\beta) W^\top \mathbb{X}\right\}\right\|_{2+\eta_1}^{2+\eta_1} < \infty. \tag{S1.9}$$

Hence, Condition (3) in Lemma 4 holds. This, together with the validity of Conditions (1)-(2), allows us to apply Lemma 4 and obtain the order of the first term on the right-hand side of (S1.8) as $o_{\mathrm{P}}(1)$, which leads to

$$-n^{-1}\frac{\partial^2 \ell(\theta)}{\partial\alpha\partial\beta_k} = -n^{-1}\mathrm{E}\frac{\partial^2 \ell(\theta)}{\partial\alpha\partial\beta_k} + o_{\mathrm{P}}(1).$$

STEP III. After algebraic calculation, we have

$$-n^{-1}\frac{\partial^2 \ell(\theta)}{\partial\alpha\partial\sigma^2} = n^{-1}\sigma^{-4}\mathbb{X}^\top\mathcal{E}(\alpha,\beta) - n^{-1}\mathrm{E}\frac{\partial^2 \ell(\theta)}{\partial\alpha\partial\sigma^2}.$$

We employ Lemma 4 by defining $A_l = 0_{n\times n}$ for $l = 1,\cdots,p$, and $B = \sigma^{-4}\mathbb{X}$. Accordingly, Conditions (1)-(2) in Lemma 4 are naturally satisfied. In addition, by Lemma 2, for any $\eta_1 > 0$,

$$n^{-1}\left\|\mathrm{vec}(\mathbb{X})\right\|_{2+\eta_1}^{2+\eta_1} = n^{-1}\sum_{j=1}^{p}\left\|\mathbb{X}_j\right\|_{2+\eta_1}^{2+\eta_1} \leq p\,|\mathbb{X}|_\infty^{2+\eta_1}.$$

Then using Condition (C2), we obtain

$$\sup_{n\geq 1} n^{-1} \left\| \mathrm{vec}(\mathbb{X}) \right\|_{2+\eta_1}^{2+\eta_1} < \infty,$$

which verifies Condition (3) in Lemma 4. This, in conjunction with the validity of Conditions (1)-(2), allows us to employ Lemma 4 and demonstrate that $n^{-1}\sigma^{-4}\mathbb{X}^\top \mathcal{E}(\alpha,\beta) = o_{\mathrm{P}}(1)$. Consequently,

$$-n^{-1}\frac{\partial^2 \ell(\theta)}{\partial \alpha \partial \sigma^2} = -n^{-1}\mathrm{E}\frac{\partial^2 \ell(\theta)}{\partial \alpha \partial \sigma^2} + o_{\mathrm{P}}(1).$$

STEP IV. After tedious calculation, for any $k_1, k_2 \in \{1, \cdots, d\}$, we have

$$\begin{aligned} -n^{-1}\frac{\partial^2 \ell(\theta)}{\partial \beta_{k_1}\partial \beta_{k_2}} \ = \ & -n^{-1}2^{-1}\sigma^{-2}\mathrm{vec}^\top \Big\{ W\Lambda_{\beta_{k_1}\beta_{k_2}}(\beta)S^{-1}(\beta) + S^{-1}(\beta)^\top \Lambda_{\beta_{k_1}\beta_{k_2}}(\beta)W^\top \\ & -S^{-1}(\beta)^\top \Lambda_{\beta_{k_1}}(\beta)W^\top W \Lambda_{\beta_{k_2}}(\beta)S^{-1}(\beta) \\ & -S^{-1}(\beta)^\top \Lambda_{\beta_{k_2}}(\beta)W^\top W \Lambda_{\beta_{k_1}}(\beta)S^{-1}(\beta)\Big\} \mathrm{vec}\left\{\mathcal{E}(\alpha,\beta)\mathcal{E}(\alpha,\beta)^\top - \sigma^2 I_n\right\} \\ & +n^{-1}\sigma^{-2}\Big\{ \alpha^\top \mathbb{X}^\top S^{-1}(\beta)^\top \Lambda_{\beta_{k_1}}(\beta)W^\top W\Lambda_{\beta_{k_2}}(\beta)S^{-1}(\beta) \\ & +\alpha^\top \mathbb{X}^\top S^{-1}(\beta)^\top \Lambda_{\beta_{k_2}}(\beta)W^\top W\Lambda_{\beta_{k_1}}(\beta)S^{-1}(\beta) \\ & -\alpha^\top \mathbb{X}^\top S^{-1}(\beta)^\top \Lambda_{\beta_{k_1}\beta_{k_2}}(\beta)W^\top \Big\} \mathcal{E}(\alpha,\beta) - n^{-1}\mathrm{E}\frac{\partial^2 \ell(\theta)}{\partial \beta_{k_1}\partial \beta_{k_2}}, \end{aligned} \tag{S1.10}$$

where $\Lambda_{\beta_{k_1}\beta_{k_2}}(\beta) := \frac{\partial^2 \Lambda(\beta)}{\partial \beta_{k_1}\partial \beta_{k_2}} = \mathrm{diag}\left\{z_{1k_1}z_{1k_2}F''(Z_1^\top \beta), \cdots, z_{nk_1}z_{nk_2}F''(Z_n^\top \beta)\right\}$.

For given $k_1, k_2 \in \{1, \cdots, d\}$, we employ Lemma 4 by defining $A_1 = -2^{-1}\sigma^{-2}\big\{W \times\Lambda_{\beta_{k_1}\beta_{k_2}}(\beta)S^{-1}(\beta) + S^{-1}(\beta)^\top \Lambda_{\beta_{k_1}\beta_{k_2}}(\beta)W^\top - S^{-1}(\beta)^\top \Lambda_{\beta_{k_1}}(\beta)W^\top W\Lambda_{\beta_{k_2}}(\beta)S^{-1}(\beta) - S^{-1}(\beta)^\top \Lambda_{\beta_{k_2}}(\beta)W^\top W\Lambda_{\beta_{k_1}}(\beta)S^{-1}(\beta)\big\}$, $L = 1$, and $B^\top = \sigma^{-2}\alpha^\top \mathbb{X}^\top S^{-1}(\beta)^\top \Lambda_{\beta_{k_1}}(\beta)W^\top \times W\Lambda_{\beta_{k_2}}(\beta)S^{-1}(\beta) + \sigma^{-2}\alpha^\top \mathbb{X}^\top S^{-1}(\beta)^\top \Lambda_{\beta_{k_2}}(\beta)W^\top W\Lambda_{\beta_{k_1}}(\beta)S^{-1}(\beta) - \sigma^{-2}\alpha^\top \mathbb{X}^\top S^{-1}(\beta)^\top$

$\Lambda_{\beta_{k_1}\beta_{k_2}}(\beta)W^\top$.

It is clear that $A_1$ is symmetric. Analogously to the proof of (S1.7), by Conditions (C2)-(C4), we can show that

$$\sup_{n\geq 1}\left\|W\Lambda_{\beta_{k_1}\beta_{k_2}}(\beta)S^{-1}(\beta)+S^{-1}(\beta)^\top\Lambda_{\beta_{k_1}\beta_{k_2}}(\beta)W^\top\right\|_1<\infty \text{ and}$$

$$\sup_{n\geq 1}\left\|S^{-1}(\beta)^\top\Lambda_{\beta_{k_1}}(\beta)W^\top W\Lambda_{\beta_{k_2}}(\beta)S^{-1}(\beta)+S^{-1}(\beta)^\top\Lambda_{\beta_{k_2}}(\beta)W^\top W\Lambda_{\beta_{k_1}}(\beta)S^{-1}(\beta)\right\|_1<\infty.$$

Accordingly, $A_1$ satisfies Conditions (1)-(2) in Lemma 4.

Employing similar techniques to those used in verifying (S1.9), under Conditions (C2)-(C4), we can demonstrate that for any $\eta_1>0$,

$$\begin{aligned}
&\sup_{n\geq 1} n^{-1}\left\|\left\{\alpha^\top\mathbb{X}^\top S^{-1}(\beta)^\top\Lambda_{\beta_{k_1}}(\beta)W^\top W\Lambda_{\beta_{k_2}}(\beta)S^{-1}(\beta)\right\}^\top\right\|_{2+\eta_1}^{2+\eta_1}<\infty,\\
&\sup_{n\geq 1} n^{-1}\left\|\left\{\alpha^\top\mathbb{X}^\top S^{-1}(\beta)^\top\Lambda_{\beta_{k_2}}(\beta)W^\top W\Lambda_{\beta_{k_1}}(\beta)S^{-1}(\beta)\right\}^\top\right\|_{2+\eta_1}^{2+\eta_1}<\infty, \text{ and}\\
&\sup_{n\geq 1} n^{-1}\left\|\left\{\alpha^\top\mathbb{X}^\top S^{-1}(\beta)^\top\Lambda_{\beta_{k_1}\beta_{k_2}}(\beta)W^\top\right\}^\top\right\|_{2+\eta_1}^{2+\eta_1}<\infty.
\end{aligned}$$

Hence, $B$ satisfies Condition (3) in Lemma 4. Finally, applying Lemma 4, we obtain that the orders of the first seven terms on the right-hand side of (S1.10) are $o_{\mathrm{P}}(1)$. Consequently,

$$-n^{-1}\frac{\partial^2\ell(\theta)}{\partial\beta_{k_1}\partial\beta_{k_2}}=-n^{-1}\mathrm{E}\frac{\partial^2\ell(\theta)}{\partial\beta_{k_1}\partial\beta_{k_2}}+o_{\mathrm{P}}(1).$$

STEP V. After algebraic simplification, we have, for $k = 1, \cdots, d$,

$$
\begin{aligned}
-n^{-1}\frac{\partial^2 \ell(\theta)}{\partial\beta_k\partial\sigma^2} &= n^{-1}2^{-1}\sigma^{-4}\mathrm{vec}\left\{W\Lambda_{\beta_k}(\beta)S^{-1}(\beta) + S^{-1}(\beta)^\top\Lambda_{\beta_k}(\beta)W^\top\right\} \\
&\quad \times \mathrm{vec}\left\{\mathcal{E}(\alpha,\beta)\mathcal{E}(\alpha,\beta)^\top - \sigma^2 I_n\right\} + n^{-1}\sigma^{-4}\alpha^\top\mathbb{X}^\top S^{-1}(\beta)^\top\Lambda_{\beta_k}(\beta)W^\top \\
&\quad \times \mathcal{E}(\alpha,\beta) - n^{-1}\mathrm{E}\frac{\partial^2 \ell(\theta)}{\partial\beta_k\partial\sigma^2}. \qquad \text{(S1.11)}
\end{aligned}
$$

We employ Lemma 4 by defining $A_1 = 2^{-1}\sigma^{-4}\{W\Lambda_{\beta_k}(\beta)S^{-1}(\beta)+S^{-1}(\beta)^\top\Lambda_{\beta_k}(\beta)W^\top\}$, $L = 1$, and $B^\top = \sigma^{-4}\alpha^\top\mathbb{X}^\top S^{-1}(\beta)^\top\Lambda_{\beta_k}(\beta)W^\top$.

It is clear that $A_1$ is symmetric. Based on the proof of (S1.7), by Conditions (C2)-(C4), we can obtain

$$
\sup_{n\geq 1}\left\|W\Lambda_{\beta_k}(\beta)S^{-1}(\beta) + S^{-1}(\beta)^\top\Lambda_{\beta_k}(\beta)W^\top\right\|_1 < \infty.
$$

Accordingly, $A_1$ satisfies Conditions (1)-(2) in Lemma 4.

By using the same techniques as those used in the proof of (S1.9), we can show that, for any $\eta_1 > 0$,

$$
\sup_{n\geq 1} n^{-1}\left\|\left\{\alpha^\top\mathbb{X}^\top S^{-1}(\beta)^\top\Lambda_{\beta_k}(\beta)W^\top\right\}^\top\right\|_{2+\eta_1}^{2+\eta_1} < \infty.
$$

Thus, Condition (3) in Lemma 4 holds. Finally, applying Lemma 4, we obtain that the orders of the first two terms on the right-hand side of (S1.11) are $o_{\mathrm{P}}(1)$. Consequently,

$$
-n^{-1}\frac{\partial^2 \ell(\theta)}{\partial\beta_k\partial\sigma^2} = -n^{-1}\mathrm{E}\frac{\partial^2 \ell(\theta)}{\partial\beta_k\partial\sigma^2} + o_{\mathrm{P}}(1).
$$

STEP VI. After algebraic calculation, we obtain

$$-n^{-1}\frac{\partial^2\ell(\theta)}{\partial\sigma^2\partial\sigma^2} = n^{-1}\sigma^{-6}\mathrm{vec}^\top(I_n)\,\mathrm{vec}\left\{\mathcal{E}(\alpha,\beta)\mathcal{E}(\alpha,\beta)^\top - \sigma^2 I_n\right\} - n^{-1}\mathrm{E}\frac{\partial^2\ell(\theta)}{\partial\sigma^2\partial\sigma^2}. \tag{S1.12}$$

We employ Lemma 4 by defining $A_1 = \sigma^{-6}I_n$, $L = 1$, and $B = 0_{n\times 1}$. It is clear that $A_1$ is symmetric and $\sup_{n\geq 1}\|I_n\|_1 = 1$. Hence, $A_1$ satisfies Conditions (1)-(2) in Lemma 4. In addition, $B$ naturally satisfies Condition (3) in Lemma 4. Applying Lemma 4, we then obtain that the order of the first term on the right-hand side of equation (S1.12) is $o_{\mathrm{P}}(1)$. Consequently,

$$-n^{-1}\frac{\partial^2\ell(\theta)}{\partial\sigma^2\partial\sigma^2} = -n^{-1}\mathrm{E}\frac{\partial^2\ell(\theta)}{\partial\sigma^2\partial\sigma^2} + o_{\mathrm{P}}(1).$$

Finally, the results obtained from the above six steps, in conjunction with Condition (C5), imply

$$-n^{-1}\frac{\partial^2\ell(\theta)}{\partial\theta\partial\theta^\top} = \mathcal{I}_n(\theta) + o_{\mathrm{P}}(1) \xrightarrow{\mathrm{P}} \mathcal{I}(\theta),$$

which completes the entire proof of Lemma 5.

## S2 Discussions of Conditions in Appendix

It is worth noting that all of the conditions in Appendix are mild and sensible. Condition (C1) is a moment condition, which is weaker than commonly used distribution assumptions; see, for example, the normal distribution assumption in Zhou et al. (2017). Conditions (C2) and (C3) have been carefully studied in Lee (2004). Condition (C4) is critical for showing the asymptotic normality of the QMLE, and the four types of the link functions discussed in Section 2.1 satisfy this condition. Condition (C5) is a standard condition for establishing the convergence of the Fisher information matrix and the variance of the score function. For example, let $\mathcal{I}_{\alpha\alpha}$ be a matrix consisting of the first $p$ rows and $p$ columns of $\mathcal{I}(\theta)$. Then, by Condition (C5) and equation (**??**), we obtain $n^{-1}\sigma^{-2}\mathbb{X}^\top\mathbb{X} \to \mathcal{I}_{\alpha\alpha}$, which is a standard assumption in linear regression analysis.

# S3 Proofs of Theorems and Corollaries

**Proof of Theorem 1.** To prove this theorem, we consider two steps, namely showing $\hat{\theta}$ is $n^{1/2}$-consistent, as $n \to \infty$, and verifying $\hat{\theta}$ is asymptotically normal, respectively.

STEP I. To complete this step, it suffices to follow the techniques of Fan and Li (2001) to show that, for an arbitrarily small positive constant $\xi > 0$, there exists a constant $M_\xi > 0$ such that

$$\mathrm{P}\left\{\sup_{u \in \mathbb{R}^{p+d+1}:\|u\|_2 = M_\xi} \ell\left(\theta + n^{-1/2}u\right) < \ell(\theta)\right\} \geq 1 - \xi \tag{S3.13}$$

as $n$ is sufficiently large enough. To this end, we employ the Taylor series expansion and obtain that

$$\begin{aligned}
& \sup_{u \in \mathbb{R}^{p+d+1}:\|u\|_2 = M_\xi} \left\{\ell\left(\theta + n^{-1/2}u\right) - \ell(\theta)\right\} \\
= & \sup_{u \in \mathbb{R}^{p+d+1}:\|u\|_2 = M_\xi} \left[n^{-1/2}\frac{\partial \ell(\theta)}{\partial \theta^\top}u - \frac{1}{2}n^{-1}u^\top\left\{-\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta^\top}\right\}u + R_n(u)\right] \\
\leq & \; M_\xi O_{\mathrm{P}}(1) - \frac{1}{2}\lambda_{\min}\left\{\mathcal{I}(\theta)\right\}M_\xi^2 + o_{\mathrm{P}}(1),
\end{aligned} \tag{S3.14}$$

where $\lambda_{\min}\{\cdot\}$ is the smallest eigenvalue of the matrix inside the braces and

$$R_n(u) = \frac{1}{6}n^{-3/2}u^\top\left(I_{p+d+1} \otimes u^\top\right)\frac{\partial}{\partial\theta^\top}\mathrm{vec}\left\{\frac{\partial^2\ell(\theta + n^{-1/2}\bar{c}u)}{\partial\theta\partial\theta^\top}\right\}u$$

with $0 \leq \bar{c} \leq 1$. Applying similar techniques to those used in the proof of Lemma 5, we obtain $R_n(u) = O_{\mathrm{P}}(n^{-1/2})$ for $\|u\|_2 = M_\xi$. This, together with Lemma 5, leads to the inequality in (S3.14). Note that $M_\xi O_{\mathrm{P}}(1) - \lambda_{\min}\left\{\mathcal{I}(\theta)\right\}M_\xi^2/2$ in (S3.14) is a quadratic function of $M_\xi$, and Condition (C5) implies its quadratic coefficient

$-\lambda_{\min}\{\mathcal{I}(\theta)\}/2 < 0$. Hence, as long as $M_\xi$ is sufficient large, we have

$$\sup_{u\in\mathbb{R}^{p+d+1}:\|u\|_2=M_\xi}\left\{\ell\left(\theta+n^{-1/2}u\right)-\ell(\theta)\right\}<0, \tag{S3.15}$$

with probability tending to 1, which demonstrates (S3.13). Based on the result of (S3.15), there exists a local maximizer $\hat{\theta}$ such that $\|\hat{\theta}-\theta\|_2 \le n^{-1/2}M_\xi$ as $n$ is large enough. This, in conjunction with (S3.13), implies

$$\mathrm{P}\left(\|\hat{\theta}-\theta\|_2 \le n^{-1/2}M_\xi\right) \ge \mathrm{P}\left\{\sup_{u\in\mathbb{R}^{p+d+1}:\|u\|_2=M_\xi}\ell\left(\theta+n^{-1/2}u\right)<\ell(\theta)\right\} \ge 1-\xi.$$

As a result, $n^{1/2}\|\hat{\theta}-\theta\|_2 = O_{\mathrm{P}}(1)$, which completes the proof of Step I.

Step II. By the result of Step I and the Taylor series expansion, we have that $0=\partial\ell(\hat{\theta})/\partial\theta = \partial\ell(\theta)/\partial\theta + \{\partial^2\ell(\theta)/\partial\theta\partial\theta^\top\}(\hat{\theta}-\theta)+\bar{R}_n$, where

$$\bar{R}_n = \frac{1}{2}\left\{I_{p+d+1}\otimes\left(\hat{\theta}-\theta\right)^\top\right\}\frac{\partial}{\partial\theta^\top}\mathrm{vec}\left\{\frac{\partial^2\ell(\bar{\theta})}{\partial\theta\partial\theta^\top}\right\}(\hat{\theta}-\theta)$$

and $\bar{\theta}$ lies between $\hat{\theta}$ and $\theta$. Then employing similar techniques to those used in the proof of Lemma 5, we obtain $n^{-1}\partial\,\mathrm{vec}\{\partial^2\ell(\bar{\theta})/\partial\theta\partial\theta^\top\}/\partial\theta^\top = O_{\mathrm{P}}(1)$. This, together with Lemma 5, implies

$$\begin{aligned} n^{1/2}(\hat{\theta}-\theta) &= n^{-1/2}\mathcal{I}^{-1}(\theta)\frac{\partial\ell(\theta)}{\partial\theta}+o_{\mathrm{P}}(1) \\ &\xrightarrow{d} N\left(0,\mathcal{I}^{-1}(\theta)+\mathcal{I}^{-1}(\theta)\mathcal{J}(\theta,\mu^{(3)},\mu^{(4)})\mathcal{I}^{-1}(\theta)\right), \end{aligned} \tag{S3.16}$$

which completes the entire proof.

**Proof of Theorem 2.** To facilitate this proof, we slightly abuse the notation $\theta=(\alpha^\top,\beta^\top,\sigma^2)^\top$ by arranging it to be $\theta=(\sigma^2,\alpha^\top,\beta^\top)^\top=(\theta_1^\top,\theta_2^\top)^\top$, where $\theta_1=(\sigma^2,\alpha^\top,\beta_1)^\top$ and $\theta_2=(\beta_2,\cdots,\beta_d)^\top$. Accordingly, the associated quantities are also changed, such as the score function $\partial\ell(\theta)/\partial\theta$ and information matrix $\mathcal{I}(\theta)$.

Under the null hypothesis, $\mathrm{H}_0 : \theta_2 = 0_{(d-1)\times 1} \in \mathbb{R}^{(d-1)\times 1}$, we denote the resulting constrained QMLE of $\theta$ as $\hat{\theta}^{(r)}$. Consider

$$\mathcal{I}(\theta) = \begin{pmatrix} \mathcal{I}_{11}(\theta) & \mathcal{I}_{12}(\theta) \\ \mathcal{I}_{21}(\theta) & \mathcal{I}_{22}(\theta) \end{pmatrix},$$

where $\mathcal{I}_{ij}(\theta)$ is the convergence of its corresponding information matrix with respect to $\theta_i$ and $\theta_j$ for $i, j \in \{1, 2\}$. Then employing similar techniques to those used for proving (S3.16), we obtain that

$$n^{1/2}(\hat{\theta}^{(r)} - \theta) = n^{-1/2}\mathcal{I}_1(\theta)\frac{\partial \ell(\theta)}{\partial \theta} + o_{\mathrm{P}}(1), \tag{S3.17}$$

where $\mathcal{I}_1(\theta) = \begin{pmatrix} \mathcal{I}_{11}^{-1}(\theta) & 0_{(p+2)\times(d-1)} \\ 0_{(d-1)\times(p+2)} & 0_{(d-1)\times(d-1)} \end{pmatrix}$. This, together with the result of Theorem 1, implies that both $\hat{\theta}$ and $\hat{\theta}^{(r)}$ are $n^{1/2}$-consistent. By (S3.16) and (S3.17), we obtain

$$\sqrt{n}(\hat{\theta}^{(r)} - \hat{\theta}) = n^{-1/2}\left\{\mathcal{I}_1(\theta) - \mathcal{I}^{-1}(\theta)\right\}\frac{\partial \ell(\theta)}{\partial \theta} + o_{\mathrm{P}}(1) = O_{\mathrm{P}}(1). \tag{S3.18}$$

Applying the Taylor series expansion, we have

$$T_{lr} = (\hat{\theta}^{(r)} - \hat{\theta})^\top \left\{-\frac{\partial^2 \ell(\tilde{\theta})}{\partial \theta \partial \theta^\top}\right\}(\hat{\theta}^{(r)} - \hat{\theta}),$$

where $\tilde{\theta}$ lies between $\hat{\theta}$ and $\hat{\theta}^{(r)}$ and it is also $n^{1/2}$-consistent. In addition, by Conditions (C1)-(C5) and applying the similar techniques to those used in the proof of Lemma 5, we have

$$-n^{-1}\frac{\partial^2 \ell(\tilde{\theta})}{\partial \theta \partial \theta^\top} \xrightarrow{\mathrm{P}} \mathcal{I}(\theta).$$

Accordingly,

$$T_{lr} = \sqrt{n}(\hat{\theta}^{(r)} - \hat{\theta})^\top \mathcal{I}(\theta)\sqrt{n}(\hat{\theta}^{(r)} - \hat{\theta}) + o_{\mathrm{P}}(1). \tag{S3.19}$$

For the sake of simplicity, denote $\mathcal{K}(\theta, \mu^{(3)}, \mu^{(4)}) = \mathcal{I}(\theta) + \mathcal{J}(\theta, \mu^{(3)}, \mu^{(4)}) = \mathcal{K}$. By Lemma 5, we have that

$$n^{-1/2}\mathcal{K}^{-1/2}\frac{\partial \ell(\theta)}{\partial \theta} \xrightarrow{d} N(0, I_{p+d+1}).$$

This, in conjunction with (S3.18), leads to

$$\begin{aligned} T_{lr} &= \left\{n^{-1/2}\mathcal{K}^{-1/2}\frac{\partial \ell(\theta)}{\partial \theta}\right\}^{\top} \mathcal{K}^{1/2}\left\{\mathcal{I}_1(\theta) - \mathcal{I}^{-1}(\theta)\right\}\mathcal{I}(\theta) \\ &\quad \times \left\{\mathcal{I}_1(\theta) - \mathcal{I}^{-1}(\theta)\right\}\mathcal{K}^{1/2}\left\{n^{-1/2}\mathcal{K}^{-1/2}\frac{\partial \ell(\theta)}{\partial \theta}\right\} + o_{\mathrm{P}}(1). \end{aligned}$$

Using the fact that $\mathcal{I}_1(\theta)\mathcal{I}(\theta)\mathcal{I}_1(\theta) = \mathcal{I}_1(\theta)$ and $\{\mathcal{I}_1(\theta) - \mathcal{I}^{-1}(\theta)\}\mathcal{I}(\theta)\{\mathcal{I}_1(\theta) - \mathcal{I}^{-1}(\theta)\} = \mathcal{I}^{-1}(\theta) - \mathcal{I}_1(\theta)$, we further obtain

$$T_{lr} = \left\{n^{-1/2}\mathcal{K}^{-1/2}\frac{\partial \ell(\theta)}{\partial \theta}\right\}^{\top} \mathcal{K}^{1/2}\left\{\mathcal{I}^{-1}(\theta) - \mathcal{I}_1(\theta)\right\}\mathcal{K}^{1/2}\left\{n^{-1/2}\mathcal{K}^{-1/2}\frac{\partial \ell(\theta)}{\partial \theta}\right\} + o_{\mathrm{P}}(1).$$

Let $\lambda_1(\theta), \cdots, \lambda_{p+d+1}(\theta)$ be the eigenvalues of $\mathcal{K}^{1/2}\left\{\mathcal{I}^{-1}(\theta) - \mathcal{I}_1(\theta)\right\}\mathcal{K}^{1/2}$. The above results, together with the continuous mapping theorem and Slutsky's theorem, imply that $T_{lr}$ follows a weighted chi-square distribution $\sum_{l=1}^{p+d+1}\lambda_l(\theta, \mu^{(3)}, \mu^{(4)})\chi_l^2(1)$ asymptotically. This completes the first part of the proof.

Under the normal assumption of $\mathcal{E}$, the matrix $\mathcal{J}_n(\theta, \mu^{(3)}, \mu^{(4)})$ defined above Theorem 1 is 0. By Condition (C5), we have $\mathcal{J}(\theta, \mu^{(3)}, \mu^{(4)}) = 0$, which leads to $\mathcal{K} = \mathcal{I}(\theta)$. Using the fact that $\{\mathcal{I}_1(\theta) - \mathcal{I}^{-1}(\theta)\}\mathcal{I}(\theta)\{\mathcal{I}_1(\theta) - \mathcal{I}^{-1}(\theta)\} = \mathcal{I}^{-1}(\theta) - \mathcal{I}_1(\theta)$, the symmetric matrix $\mathcal{K}^{1/2}\{\mathcal{I}^{-1}(\theta) - \mathcal{I}_1(\theta)\}\mathcal{K}^{1/2} = \mathcal{I}^{1/2}(\theta)\{\mathcal{I}^{-1}(\theta) - \mathcal{I}_1(\theta)\}\mathcal{I}^{1/2}(\theta)$ is idempotent. In addition,

$$\begin{aligned} &\operatorname{tr}\left[\mathcal{I}^{1/2}(\theta)\left\{\mathcal{I}^{-1}(\theta) - \mathcal{I}_1(\theta)\right\}\mathcal{I}^{1/2}(\theta)\right] = \operatorname{tr}\left[\left\{\mathcal{I}^{-1}(\theta) - \mathcal{I}_1(\theta)\right\}\mathcal{I}(\theta)\right] \\ = &\operatorname{tr}\left\{I_{p+d+1} - \mathcal{I}_1(\theta)\mathcal{I}(\theta)\right\} = \operatorname{tr}\left\{\begin{pmatrix} 0_{(p+2)\times(p+2)} & -\mathcal{I}_{11}^{-1}\mathcal{I}_{12} \\ 0_{(d-1)\times(p+2)} & I_{d-1} \end{pmatrix}\right\} = d-1. \end{aligned}$$

The above results, together with the normality assumption, imply that $T_{lr} \xrightarrow{d} \chi^2(d-1)$, which completes the entire proof.

**Proof of Corollary 1.** By Conditions (C1)-(C5) and applying the similar techniques to those used in the proof of Lemma 5, we obtain $\mathcal{I}_n^{-1}(\hat{\theta})\mathcal{K}_n(\hat{\theta}, \hat{\mu}^{(3)}, \hat{\mu}^{(4)})\mathcal{I}_n^{-1}(\hat{\theta}) \xrightarrow{\mathrm{P}} \mathcal{I}^{-1}(\theta)\mathcal{K}\mathcal{I}^{-1}(\theta)$. In addition, by Theorem 1 and the continuous mapping theorem, we have, under the null hypothesis, $n^{1/2}(\Delta\hat{\theta}) \xrightarrow{d} N\big(0, \Delta\mathcal{I}^{-1}(\theta)\mathcal{K}\mathcal{I}^{-1}(\theta)\Delta^\top\big)$. The above results, together with Slutsky's theorem and the continuous mapping theorem, imply

$$T_w = (\Delta\hat{\theta})^\top \left[\Delta \left\{n^{-1}\mathcal{I}_n^{-1}(\hat{\theta})\mathcal{K}_n(\hat{\theta}, \hat{\mu}^{(3)}, \hat{\mu}^{(4)})\mathcal{I}_n^{-1}(\hat{\theta})\right\} \Delta^\top\right]^{-1} \Delta\hat{\theta} \xrightarrow{d} \chi^2(d-1).$$

This completes the proof.

**Proof of Corollary 2.** Let $\theta$ be defined as in the beginning of the proof of Theorem 2. Employing the Taylor series expansion, we have that

$$n^{-1/2}\frac{\partial \ell(\hat{\theta}^{(r)})}{\partial \theta} = \frac{1}{n}\frac{\partial \ell^2(\breve{\theta})}{\partial\theta\partial\theta^\top} n^{1/2}(\hat{\theta}^{(r)} - \hat{\theta}),$$

where $\breve{\theta}$ lies between $\hat{\theta}^{(r)}$ and $\hat{\theta}$. Since both $\hat{\theta}^{(r)}$ and $\hat{\theta}$ are $n^{1/2}$-consistent under the null hypothesis, $\breve{\theta}$ is also $n^{1/2}$-consistent. By Conditions (C1)-(C5) and applying similar techniques to those used in the proof of Lemma 5, we obtain, under the null hypothesis,

$$-n^{-1}\frac{\partial^2 \ell(\breve{\theta})}{\partial\theta\partial\theta^\top} \xrightarrow{\mathrm{P}} \mathcal{I}(\theta) \text{ and } \mathcal{I}_n^{-1}(\hat{\theta}^{(r)}) \xrightarrow{\mathrm{P}} \mathcal{I}^{-1}(\theta).$$

These results, together with (S3.19), imply that

$$\begin{aligned} T_s &= n^{-1}\left\{\frac{\partial \ell(\hat{\theta}^{(r)})}{\partial\theta}\right\}^\top \mathcal{I}_n^{-1}(\hat{\theta}^{(r)})\frac{\partial \ell(\hat{\theta}^{(r)})}{\partial\theta} \\ &= \sqrt{n}(\hat{\theta}^{(r)} - \hat{\theta})^\top \mathcal{I}(\theta)\sqrt{n}(\hat{\theta}^{(r)} - \hat{\theta}) + o_{\mathrm{P}}(1) = T_{lr} + o_{\mathrm{P}}(1), \end{aligned}$$

which completes the proof.

## S4 Simulation and Empirical Results

To further examine the robustness of parameter estimates with respect to random error distributions, we simulated the independent and identically distributed random errors from $\sigma\zeta$, where $\zeta$ follows a standard normal distribution, a mixture normal distribution, a standardized $t_3$ distribution and a standardized exponential distribution, respectively. The rest of the model simulation setting is the same as that in Section 3 of the paper. Tables S.1-S.4 report results for these four distributions, respectively.

As suggested by an anonymous referee, we further adopt the network settings from Ma et al. (2019) and conduct simulation experiments under the power-law network structure and the network structure generated from the stochastic block model (SBM). The rest of the model simulation setting is the same as that in Table S.2. Tables S.5-S.6 present the results for these two network settings, respectively.

The results in Tables S.2-S.6 yield qualitatively similar findings to those obtained from the Gaussian error and original network setting in Table S.1. Hence, our proposed estimates still exhibit nice properties under these three non-normal errors and two different network structures. Note that the average execution times (in seconds) are not reported in Tables S.3-S.4 since they are quite similar to those in Tables S.1-S.2.

To measure the computation efficiency of the proposed QMLE under large-scale networks, we follow another referee's suggestion to conduct the simulation experiment with $n$ =10,000. The rest of the model simulation setting is the same as that in Table S.2. The results are reported in Table S.7. Apparently, the computation efficiency under $n$ =10,000 is limited since QMLE involves huge matrix operations under the large-scale network. Hence, we only conduct 100 realizations in this study. The discussion of possible approaches to deal with large-scale network computations is in Section 5 of the paper.

According to an anonymous referee's suggestion, we report the empirical coverages of a 95%

confidence interval constructed by QMLEs and their asymptotic normal distributions; see Table S.8. The model simulation setting is the same as that in Table S.1. Table S.8 indicates that empirical coverages are close to 95% when $n$ is large.

Table S.9 provides additional empirical results mentioned in Section 4 of the paper.

Table S.1: Comparison of the QMLEs of the parameters ($\alpha_1 = 2$, $\alpha_2 = 1$, $\beta_1 = -1$, $\beta_2 = 5$, $\beta_3 = -2$ and $\sigma^2 = 1$) for the exponential, logistic, inverse of the probit, and inverse of the log-log link functions, respectively. The independent and identically distributed random errors are simulated from the normal distribution $N(0, \sigma^2)$. Three measures are considered: the averaged bias of the estimate (BIAS), the standard deviation of the estimate (SD), and the root mean squared error of the estimate (RMSE). The average execution times (in seconds) for computing the parameter estimates are in parentheses.

| Link | $n$ | Measure | $\hat{\alpha}_1$ | $\hat{\alpha}_2$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|
| Exponential | $n = 200$ | BIAS | 0.0339 | -0.0051 | -0.0523 | 0.1760 | -0.0517 | -0.0156 |
| | | SD | 0.2925 | 0.0746 | 0.2264 | 0.8390 | 0.3425 | 0.1032 |
| | (0.61) | RMSE | 0.2945 | 0.0748 | 0.2323 | 0.8573 | 0.3464 | 0.1044 |
| | $n = 500$ | BIAS | 0.0129 | 0.0002 | -0.0141 | 0.0389 | -0.0086 | -0.0048 |
| | | SD | 0.1570 | 0.0490 | 0.1093 | 0.4021 | 0.1574 | 0.0669 |
| | (2.88) | RMSE | 0.1575 | 0.0490 | 0.1102 | 0.4040 | 0.1576 | 0.0671 |
| | $n$ =1,000 | BIAS | 0.0009 | 0.0013 | -0.0025 | -0.0017 | -0.0028 | -0.0027 |
| | | SD | 0.1109 | 0.0310 | 0.0788 | 0.3056 | 0.1340 | 0.0463 |
| | (14.20) | RMSE | 0.1109 | 0.0310 | 0.0788 | 0.3056 | 0.1340 | 0.0463 |
| Logistic | $n = 200$ | BIAS | 0.0168 | -0.0045 | -0.3548 | 1.4773 | -0.5960 | -0.0164 |
| | | SD | 0.2800 | 0.0746 | 1.7801 | 5.6597 | 4.8306 | 0.1032 |
| | (0.59) | RMSE | 0.2806 | 0.0747 | 1.8151 | 5.8493 | 4.8672 | 0.1045 |
| | $n = 500$ | BIAS | 0.0088 | 0.0003 | -0.0659 | 0.2344 | -0.1000 | -0.0051 |
| | | SD | 0.1644 | 0.0489 | 0.3870 | 1.4612 | 0.9375 | 0.0669 |
| | (2.36) | RMSE | 0.1647 | 0.0489 | 0.3926 | 1.4799 | 0.9428 | 0.0670 |
| | $n$ =1,000 | BIAS | 0.0052 | 0.0012 | -0.0261 | 0.0736 | -0.0477 | -0.0027 |
| | | SD | 0.1105 | 0.0310 | 0.2215 | 0.8997 | 0.6125 | 0.0464 |
| | (11.54) | RMSE | 0.1106 | 0.0310 | 0.2230 | 0.9027 | 0.6144 | 0.0464 |
| Inverse Probit | $n = 200$ | BIAS | -0.0150 | -0.0038 | -0.3570 | 1.7669 | -0.6655 | -0.0165 |
| | | SD | 0.2118 | 0.0744 | 1.8620 | 8.4554 | 3.9599 | 0.1035 |
| | (0.84) | RMSE | 0.2123 | 0.0745 | 1.8960 | 8.6381 | 4.0155 | 0.1048 |
| | $n = 500$ | BIAS | -0.0021 | 0.0005 | -0.0593 | 0.2846 | -0.1078 | -0.0050 |
| | | SD | 0.1268 | 0.0488 | 0.3154 | 1.4857 | 0.8097 | 0.0670 |
| | (3.27) | RMSE | 0.1268 | 0.0488 | 0.3209 | 1.5127 | 0.8169 | 0.0671 |
| | $n$ =1,000 | BIAS | -0.0007 | 0.0012 | -0.0230 | 0.1013 | -0.0522 | -0.0028 |
| | | SD | 0.0878 | 0.0310 | 0.1880 | 0.9448 | 0.5241 | 0.0464 |
| | (14.64) | RMSE | 0.0878 | 0.0310 | 0.1894 | 0.9503 | 0.5267 | 0.0464 |
| Inverse Log-Log | $n = 200$ | BIAS | 0.0225 | -0.0047 | -0.1829 | 0.8806 | -0.2919 | -0.0163 |
| | | SD | 0.2898 | 0.0746 | 0.8125 | 3.7780 | 1.3693 | 0.1033 |
| | (0.61) | RMSE | 0.2907 | 0.0747 | 0.8328 | 3.8793 | 1.4001 | 0.1046 |
| | $n = 500$ | BIAS | 0.0114 | 0.0003 | -0.0455 | 0.1679 | -0.0743 | -0.0051 |
| | | SD | 0.1658 | 0.0489 | 0.2668 | 1.0556 | 0.6540 | 0.0669 |
| | (2.54) | RMSE | 0.1662 | 0.0489 | 0.2707 | 1.0689 | 0.6582 | 0.0671 |
| | $n$ =1,000 | BIAS | 0.0048 | 0.0012 | -0.0171 | 0.0518 | -0.0355 | -0.0027 |
| | | SD | 0.1123 | 0.0310 | 0.1621 | 0.6705 | 0.4323 | 0.0463 |
| | (12.22) | RMSE | 0.1124 | 0.0310 | 0.1630 | 0.6725 | 0.4337 | 0.0464 |

Table S.2: Comparison of the QMLEs of the parameters ($\alpha_1 = 2$, $\alpha_2 = 1$, $\beta_1 = -1$, $\beta_2 = 5$, $\beta_3 = -2$ and $\sigma^2 = 1$) for the exponential, logistic, inverse of the probit, and inverse of the log-log link functions, respectively. The independent and identically distributed random errors are simulated from $\sigma\zeta$, where $\zeta$ follows a mixture normal distribution $0.9N(0, 5/9) + 0.1N(0, 5)$. Three measures are considered: the averaged bias of the estimate (BIAS), the standard deviation of the estimate (SD), and the root mean squared error of the estimate (RMSE). The average execution times (in seconds) for computing the parameter estimates are in parentheses.

| Link | $n$ | Measure | $\hat{\alpha}_1$ | $\hat{\alpha}_2$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|
| Exponential | $n = 200$ | BIAS | 0.0380 | -0.0014 | -0.0579 | 0.1971 | -0.0509 | -0.0099 |
| | | SD | 0.3027 | 0.0747 | 0.2360 | 0.8825 | 0.3495 | 0.1902 |
| | (0.83) | RMSE | 0.3050 | 0.0747 | 0.2430 | 0.9042 | 0.3532 | 0.1905 |
| | $n = 500$ | BIAS | 0.0126 | 0.0000 | -0.0152 | 0.0473 | -0.0090 | -0.0007 |
| | | SD | 0.1609 | 0.0491 | 0.1111 | 0.4097 | 0.1577 | 0.1275 |
| | (3.25) | RMSE | 0.1614 | 0.0491 | 0.1122 | 0.4124 | 0.1580 | 0.1275 |
| | $n =$1,000 | BIAS | 0.0038 | 0.0001 | -0.0044 | 0.0075 | -0.0042 | 0.0003 |
| | | SD | 0.1083 | 0.0312 | 0.0767 | 0.3054 | 0.1358 | 0.0856 |
| | (14.65) | RMSE | 0.1084 | 0.0312 | 0.0768 | 0.3055 | 0.1358 | 0.0856 |
| Logistic | $n = 200$ | BIAS | 0.0239 | -0.0009 | -0.4441 | 1.6682 | -0.7582 | -0.0103 |
| | | SD | 0.2839 | 0.0747 | 2.6011 | 7.6328 | 6.4832 | 0.1903 |
| | (0.75) | RMSE | 0.2849 | 0.0747 | 2.6387 | 7.8130 | 6.5274 | 0.1906 |
| | $n = 500$ | BIAS | 0.0090 | 0.0002 | -0.0739 | 0.3190 | -0.0826 | -0.0009 |
| | | SD | 0.1673 | 0.0489 | 0.4228 | 1.7378 | 0.9219 | 0.1273 |
| | (2.77) | RMSE | 0.1676 | 0.0489 | 0.4292 | 1.7668 | 0.9256 | 0.1273 |
| | $n =$1,000 | BIAS | 0.0059 | 0.0001 | -0.0258 | 0.0780 | -0.0363 | 0.0003 |
| | | SD | 0.1101 | 0.0312 | 0.2189 | 0.9108 | 0.5976 | 0.0856 |
| | (11.74) | RMSE | 0.1102 | 0.0312 | 0.2204 | 0.9142 | 0.5987 | 0.0856 |
| Inverse Probit | $n = 200$ | BIAS | -0.0108 | -0.0005 | -0.4591 | 1.9509 | -0.9679 | -0.0110 |
| | | SD | 0.2189 | 0.0748 | 2.6830 | 8.0535 | 7.7695 | 0.1899 |
| | (1.04) | RMSE | 0.2192 | 0.0748 | 2.7220 | 8.2864 | 7.8296 | 0.1902 |
| | $n = 500$ | BIAS | -0.0017 | 0.0003 | -0.0714 | 0.3794 | -0.0956 | -0.0009 |
| | | SD | 0.1306 | 0.0489 | 0.3463 | 1.7206 | 0.7980 | 0.1274 |
| | (3.78) | RMSE | 0.1306 | 0.0489 | 0.3536 | 1.7619 | 0.8037 | 0.1274 |
| | $n =$1,000 | BIAS | 0.0008 | 0.0001 | -0.0247 | 0.1121 | -0.0440 | 0.0003 |
| | | SD | 0.0879 | 0.0312 | 0.1877 | 0.9442 | 0.5198 | 0.0856 |
| | (15.03) | RMSE | 0.0879 | 0.0312 | 0.1894 | 0.9508 | 0.5217 | 0.0856 |
| Inverse Log-Log | $n = 200$ | BIAS | 0.0286 | -0.0011 | -0.1906 | 0.8914 | -0.2981 | -0.0102 |
| | | SD | 0.2927 | 0.0747 | 0.6886 | 3.0889 | 1.4761 | 0.1905 |
| | (0.81) | RMSE | 0.2941 | 0.0747 | 0.7145 | 3.2150 | 1.5059 | 0.1908 |
| | $n = 500$ | BIAS | 0.0120 | 0.0002 | -0.0498 | 0.2131 | -0.0623 | -0.0008 |
| | | SD | 0.1688 | 0.0489 | 0.2785 | 1.1581 | 0.6418 | 0.1274 |
| | (2.95) | RMSE | 0.1692 | 0.0489 | 0.2829 | 1.1776 | 0.6448 | 0.1274 |
| | $n =$1,000 | BIAS | 0.0062 | 0.0001 | -0.0179 | 0.0566 | -0.0283 | 0.0003 |
| | | SD | 0.1114 | 0.0312 | 0.1601 | 0.6765 | 0.4233 | 0.0856 |
| | (12.53) | RMSE | 0.1115 | 0.0312 | 0.1611 | 0.6789 | 0.4243 | 0.0856 |

Table S.3: Comparison of the QMLEs of the parameters ($\alpha_1 = 2$, $\alpha_2 = 1$, $\beta_1 = -1$, $\beta_2 = 5$, $\beta_3 = -2$ and $\sigma^2 = 1$) for the exponential, logistic, inverse of the probit, and inverse of the log-log link functions, respectively. The independent and identically distributed random errors are simulated from $\sigma\zeta$, where $\zeta$ follows a standardized $t_3$ distribution. Three measures are considered: the averaged bias of the estimate (BIAS), the standard deviation of the estimate (SD), and the root mean squared error of the estimate (RMSE).

| Link | $n$ | Measure | $\hat{\alpha}_1$ | $\hat{\alpha}_2$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|
| Exponential | $n = 200$ | BIAS | 0.0388 | -0.0067 | -0.0572 | 0.1911 | -0.0568 | -0.0016 |
| | | SD | 0.2885 | 0.0770 | 0.2347 | 0.8643 | 0.3522 | 0.7034 |
| | | RMSE | 0.2911 | 0.0773 | 0.2416 | 0.8851 | 0.3568 | 0.7034 |
| | $n = 500$ | BIAS | 0.0133 | -0.0001 | -0.0137 | 0.0387 | -0.0077 | -0.0104 |
| | | SD | 0.1486 | 0.0477 | 0.1043 | 0.3981 | 0.1521 | 0.4769 |
| | | RMSE | 0.1492 | 0.0477 | 0.1052 | 0.4000 | 0.1523 | 0.4770 |
| | $n =1{,}000$ | BIAS | -0.0001 | 0.0007 | -0.0032 | 0.0087 | -0.0045 | 0.0091 |
| | | SD | 0.1112 | 0.0315 | 0.0782 | 0.2991 | 0.1320 | 0.4618 |
| | | RMSE | 0.1112 | 0.0315 | 0.0783 | 0.2992 | 0.1321 | 0.4619 |
| Logistic | $n = 200$ | BIAS | 0.0218 | -0.0063 | -0.4725 | 1.6868 | -0.7899 | -0.0032 |
| | | SD | 0.2758 | 0.0771 | 2.6725 | 12.1332 | 4.9082 | 0.7035 |
| | | RMSE | 0.2767 | 0.0774 | 2.7139 | 12.2498 | 4.9713 | 0.7035 |
| | $n = 500$ | BIAS | 0.0102 | 0.0000 | -0.0760 | 0.1716 | -0.1157 | -0.0110 |
| | | SD | 0.1551 | 0.0476 | 0.5967 | 1.9302 | 1.7875 | 0.4689 |
| | | RMSE | 0.1554 | 0.0476 | 0.6016 | 1.9378 | 1.7913 | 0.4690 |
| | $n =1{,}000$ | BIAS | 0.0050 | 0.0006 | -0.0288 | 0.1252 | -0.0631 | 0.0090 |
| | | SD | 0.1095 | 0.0315 | 0.2197 | 0.9156 | 0.6046 | 0.4615 |
| | | RMSE | 0.1096 | 0.0315 | 0.2216 | 0.9242 | 0.6079 | 0.4616 |
| Inverse Probit | $n = 200$ | BIAS | -0.0087 | -0.0057 | -0.4423 | 2.2639 | -0.8289 | -0.0035 |
| | | SD | 0.2082 | 0.0772 | 2.3112 | 11.4641 | 5.1989 | 0.7033 |
| | | RMSE | 0.2084 | 0.0774 | 2.3531 | 11.6854 | 5.2646 | 0.7033 |
| | $n = 500$ | BIAS | 0.0018 | 0.0002 | -0.0837 | 0.2823 | -0.1465 | -0.0110 |
| | | SD | 0.1225 | 0.0477 | 0.6208 | 2.0223 | 1.8435 | 0.4695 |
| | | RMSE | 0.1225 | 0.0477 | 0.6264 | 2.0419 | 1.8493 | 0.4697 |
| | $n =1{,}000$ | BIAS | -0.0020 | 0.0006 | -0.0245 | 0.1413 | -0.0609 | 0.0090 |
| | | SD | 0.0887 | 0.0315 | 0.1875 | 0.9448 | 0.5151 | 0.4616 |
| | | RMSE | 0.0887 | 0.0315 | 0.1891 | 0.9554 | 0.5187 | 0.4617 |
| Inverse Log-Log | $n = 200$ | BIAS | 0.0291 | -0.0063 | -0.2160 | 0.9841 | -0.3872 | -0.0024 |
| | | SD | 0.2820 | 0.0771 | 1.1084 | 5.5325 | 2.7047 | 0.7037 |
| | | RMSE | 0.2835 | 0.0774 | 1.1293 | 5.6193 | 2.7323 | 0.7037 |
| | $n = 500$ | BIAS | 0.0135 | -0.0000 | -0.0557 | 0.1282 | -0.0940 | -0.0110 |
| | | SD | 0.1582 | 0.0476 | 0.4459 | 1.4110 | 1.3360 | 0.4682 |
| | | RMSE | 0.1587 | 0.0476 | 0.4494 | 1.4169 | 1.3393 | 0.4684 |
| | $n =1{,}000$ | BIAS | 0.0047 | 0.0006 | -0.0193 | 0.0883 | -0.0468 | 0.0090 |
| | | SD | 0.1115 | 0.0315 | 0.1605 | 0.6734 | 0.4276 | 0.4615 |
| | | RMSE | 0.1116 | 0.0315 | 0.1617 | 0.6792 | 0.4301 | 0.4616 |

Table S.4: Comparison of the QMLEs of the parameters ($\alpha_1 = 2$, $\alpha_2 = 1$, $\beta_1 = -1$, $\beta_2 = 5$, $\beta_3 = -2$ and $\sigma^2 = 1$) for the exponential, logistic, inverse of the probit, and inverse of the log-log link functions, respectively. The independent and identically distributed random errors are simulated from $\sigma\zeta$, where $\zeta$ follows a standardized exponential distribution. Three measures are considered: the averaged bias of the estimate (BIAS), the standard deviation of the estimate (SD), and the root mean squared error of the estimate (RMSE).

| Link | $n$ | Measure | $\hat{\alpha}_1$ | $\hat{\alpha}_2$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|
| Exponential | $n = 200$ | BIAS | 0.0286 | 0.0016 | -0.0440 | 0.1522 | -0.0311 | -0.0242 |
| | | SD | 0.2908 | 0.0715 | 0.2197 | 0.8350 | 0.3331 | 0.1907 |
| | | RMSE | 0.2922 | 0.0715 | 0.2241 | 0.8487 | 0.3345 | 0.1922 |
| | $n = 500$ | BIAS | 0.0050 | 0.0004 | -0.0084 | 0.0161 | -0.0083 | -0.0105 |
| | | SD | 0.1566 | 0.0470 | 0.1085 | 0.4052 | 0.1562 | 0.1275 |
| | | RMSE | 0.1567 | 0.0470 | 0.1089 | 0.4056 | 0.1564 | 0.1279 |
| | $n$ =1,000 | BIAS | 0.0012 | 0.0013 | -0.0037 | -0.0010 | -0.0082 | -0.0067 |
| | | SD | 0.1118 | 0.0317 | 0.0823 | 0.3196 | 0.1388 | 0.0885 |
| | | RMSE | 0.1118 | 0.0317 | 0.0824 | 0.3196 | 0.1390 | 0.0887 |
| Logistic | $n = 200$ | BIAS | 0.0281 | 0.0018 | -0.4055 | 1.7055 | -0.4230 | -0.0243 |
| | | SD | 0.2815 | 0.0717 | 2.7929 | 11.0253 | 3.1882 | 0.1902 |
| | | RMSE | 0.2829 | 0.0717 | 2.8222 | 11.1565 | 3.2162 | 0.1917 |
| | $n = 500$ | BIAS | 0.0018 | 0.0003 | -0.0468 | 0.1821 | -0.1116 | -0.0105 |
| | | SD | 0.1568 | 0.0469 | 0.3589 | 1.4959 | 0.9479 | 0.1275 |
| | | RMSE | 0.1568 | 0.0469 | 0.3619 | 1.5069 | 0.9545 | 0.1279 |
| | $n$ =1,000 | BIAS | 0.0023 | 0.0012 | -0.0240 | 0.0643 | -0.0607 | -0.0067 |
| | | SD | 0.1138 | 0.0317 | 0.2369 | 0.9121 | 0.6103 | 0.0885 |
| | | RMSE | 0.1139 | 0.0317 | 0.2381 | 0.9144 | 0.6133 | 0.0887 |
| Inverse Probit | $n = 200$ | BIAS | -0.0058 | 0.0025 | -0.3522 | 1.6998 | -0.5352 | -0.0251 |
| | | SD | 0.2145 | 0.0715 | 2.1311 | 8.6770 | 4.6687 | 0.1900 |
| | | RMSE | 0.2146 | 0.0715 | 2.1600 | 8.8420 | 4.6993 | 0.1917 |
| | $n = 500$ | BIAS | -0.0069 | 0.0004 | -0.0555 | 0.2970 | -0.1036 | -0.0106 |
| | | SD | 0.1257 | 0.0470 | 0.3579 | 1.7949 | 0.8147 | 0.1274 |
| | | RMSE | 0.1259 | 0.0470 | 0.3622 | 1.8194 | 0.8212 | 0.1279 |
| | $n$ =1,000 | BIAS | -0.0029 | 0.0012 | -0.0216 | 0.0903 | -0.0539 | -0.0067 |
| | | SD | 0.0899 | 0.0317 | 0.1971 | 0.9366 | 0.5264 | 0.0885 |
| | | RMSE | 0.0900 | 0.0318 | 0.1982 | 0.9409 | 0.5292 | 0.0887 |
| Inverse Log-Log | $n = 200$ | BIAS | 0.0307 | 0.0018 | -0.1819 | 0.7451 | -0.3023 | -0.0247 |
| | | SD | 0.2876 | 0.0713 | 1.0704 | 2.8038 | 3.2294 | 0.1902 |
| | | RMSE | 0.2892 | 0.0713 | 1.0858 | 2.9011 | 3.2435 | 0.1918 |
| | $n = 500$ | BIAS | 0.0034 | 0.0003 | -0.0316 | 0.1286 | -0.0804 | -0.0104 |
| | | SD | 0.1593 | 0.0469 | 0.2553 | 1.0835 | 0.6632 | 0.1275 |
| | | RMSE | 0.1594 | 0.0469 | 0.2572 | 1.0911 | 0.6680 | 0.1279 |
| | $n$ =1,000 | BIAS | 0.0027 | 0.0012 | -0.0162 | 0.0435 | -0.0423 | -0.0067 |
| | | SD | 0.1154 | 0.0317 | 0.1709 | 0.6779 | 0.4242 | 0.0885 |
| | | RMSE | 0.1154 | 0.0317 | 0.1717 | 0.6793 | 0.4263 | 0.0887 |

Table S.5: Comparison of the QMLEs of the parameters ($\alpha_1 = 2$, $\alpha_2 = 1$, $\beta_1 = -1$, $\beta_2 = 5$, $\beta_3 = -2$ and $\sigma^2 = 1$) for the exponential link function. The independent and identically distributed random errors are simulated from $\sigma\zeta$, where $\zeta$ follows a mixture normal distribution $0.9N(0, 5/9) + 0.1N(0, 5)$. The weighting matrix $W$ is generated from the power-law network structure. Three measures are considered: the averaged bias of the estimate (BIAS), the standard deviation of the estimate (SD), and the root mean squared error of the estimate (RMSE). The average execution times (in seconds) for computing the parameter estimates are in parentheses.

| Link | $n$ | Measure | $\hat{\alpha}_1$ | $\hat{\alpha}_2$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|
| Exponential | $n = 200$ | BIAS | 0.0042 | -0.0009 | -0.0195 | 0.0608 | -0.0220 | -0.0095 |
| | | SD | 0.1277 | 0.0745 | 0.1416 | 0.6552 | 0.2953 | 0.1895 |
| | (0.14) | RMSE | 0.1278 | 0.0745 | 0.1429 | 0.6580 | 0.2961 | 0.1897 |
| | $n = 500$ | BIAS | 0.0023 | 0.0003 | -0.0013 | -0.0001 | -0.0021 | -0.0005 |
| | | SD | 0.0988 | 0.0489 | 0.0436 | 0.1305 | 0.0964 | 0.1273 |
| | (1.02) | RMSE | 0.0989 | 0.0489 | 0.0436 | 0.1305 | 0.0965 | 0.1273 |
| | $n$ =1,000 | BIAS | -0.0018 | -0.0010 | 0.0004 | 0.0035 | 0.0039 | 0.0007 |
| | | SD | 0.0626 | 0.0319 | 0.0368 | 0.1245 | 0.0922 | 0.0855 |
| | (4.60) | RMSE | 0.0627 | 0.0319 | 0.0368 | 0.1246 | 0.0923 | 0.0855 |

Table S.6: Comparison of the QMLEs of the parameters ($\alpha_1 = 2$, $\alpha_2 = 1$, $\beta_1 = -1$, $\beta_2 = 5$, $\beta_3 = -2$ and $\sigma^2 = 1$) for the exponential link function. The independent and identically distributed random errors are simulated from $\sigma\zeta$, where $\zeta$ follows a mixture normal distribution $0.9N(0, 5/9) + 0.1N(0, 5)$. The weighting matrix $W$ is based on the network structure generated from the stochastic block model (SBM). Three measures are considered: the averaged bias of the estimate (BIAS), the standard deviation of the estimate (SD), and the root mean squared error of the estimate (RMSE). The average execution times (in seconds) for computing the parameter estimates are in parentheses.

| Link | $n$ | Measure | $\hat{\alpha}_1$ | $\hat{\alpha}_2$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|
| Exponential | $n = 200$ | BIAS | 0.0537 | -0.0019 | -0.0966 | 0.3359 | -0.0496 | -0.0096 |
| | | SD | 0.3323 | 0.0748 | 0.3111 | 1.3322 | 0.4980 | 0.1903 |
| | (0.47) | RMSE | 0.3366 | 0.0748 | 0.3258 | 1.3739 | 0.5005 | 0.1905 |
| | $n = 500$ | BIAS | 0.0267 | 0.0009 | -0.0438 | 0.1396 | -0.0248 | -0.0006 |
| | | SD | 0.2722 | 0.0493 | 0.2233 | 0.8579 | 0.3122 | 0.1275 |
| | (4.35) | RMSE | 0.2735 | 0.0494 | 0.2276 | 0.8692 | 0.3131 | 0.1275 |
| | $n$ =1,000 | BIAS | 0.0281 | -0.0000 | -0.0358 | 0.1131 | -0.0318 | 0.0004 |
| | | SD | 0.2578 | 0.0312 | 0.1827 | 0.6330 | 0.2900 | 0.0856 |
| | (47.60) | RMSE | 0.2593 | 0.0312 | 0.1862 | 0.6430 | 0.2917 | 0.0856 |

Table S.7: Comparison of the QMLEs of the parameters ($\alpha_1 = 2$, $\alpha_2 = 1$, $\beta_1 = -1$, $\beta_2 = 5$, $\beta_3 = -2$ and $\sigma^2 = 1$) for the exponential link function. The independent and identically distributed random errors are simulated from $\sigma\zeta$, where $\zeta$ follows a mixture normal distribution $0.9N(0, 5/9) + 0.1N(0, 5)$. Three measures are considered: the averaged bias of the estimate (BIAS), the standard deviation of the estimate (SD), and the root mean squared error of the estimate (RMSE). The average execution times (in seconds) for computing the parameter estimates are in parentheses.

| Link | $n$ | Measure | $\hat{\alpha}_1$ | $\hat{\alpha}_2$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\sigma}^2$ |
|---|---|---|---|---|---|---|---|---|
| Exponential | $n$ =10,000 | BIAS | -0.0024 | 0.0014 | 0.0008 | 0.0024 | 0.0009 | 0.0014 |
| | | SD | 0.0356 | 0.0089 | 0.0263 | 0.1038 | 0.0437 | 0.0290 |
| | (17,468.83) | RMSE | 0.0357 | 0.0090 | 0.0264 | 0.1038 | 0.0437 | 0.0290 |

Table S.8: Comparison of the QMLEs of the parameters ($\alpha_1 = 2$, $\alpha_2 = 1$, $\beta_1 = -1$, $\beta_2 = 5$, $\beta_3 = -2$ and $\sigma^2 = 1$) for the exponential, logistic, inverse of the probit, and inverse of the log-log link functions, respectively. The independent and identically distributed random errors are simulated from the normal distribution $N(0, \sigma^2)$. One measure is considered: the empirical coverages of a 95% confidence interval constructed by QMLEs and their asymptotic normal distributions.

| Link | $n$ | Measure | $\alpha_1$ | $\alpha_2$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|---|
| Exponential | $n = 200$ | Coverage | 0.9420 | 0.9450 | 0.9450 | 0.9450 | 0.9360 | 0.9200 |
| | $n = 500$ | Coverage | 0.9480 | 0.9360 | 0.9510 | 0.9490 | 0.9530 | 0.9270 |
| | $n =$1,000 | Coverage | 0.9630 | 0.9530 | 0.9550 | 0.9590 | 0.9520 | 0.9340 |
| Logistic | $n = 200$ | Coverage | 0.8960 | 0.9480 | 0.9660 | 0.9600 | 0.9740 | 0.9190 |
| | $n = 500$ | Coverage | 0.9360 | 0.9380 | 0.9520 | 0.9490 | 0.9470 | 0.9290 |
| | $n =$1,000 | Coverage | 0.9470 | 0.9530 | 0.9590 | 0.9650 | 0.9440 | 0.9310 |
| Inverse Probit | $n = 200$ | Coverage | 0.9120 | 0.9440 | 0.9460 | 0.9380 | 0.9690 | 0.9160 |
| | $n = 500$ | Coverage | 0.9360 | 0.9380 | 0.9470 | 0.9400 | 0.9450 | 0.9270 |
| | $n =$1,000 | Coverage | 0.9430 | 0.9540 | 0.9520 | 0.9480 | 0.9480 | 0.9310 |
| Inverse Log-Log | $n = 200$ | Coverage | 0.8990 | 0.9460 | 0.9490 | 0.9530 | 0.9650 | 0.9190 |
| | $n = 500$ | Coverage | 0.9360 | 0.9380 | 0.9550 | 0.9440 | 0.9500 | 0.9280 |
| | $n =$1,000 | Coverage | 0.9450 | 0.9530 | 0.9560 | 0.9620 | 0.9420 | 0.9310 |

Table S.9: The impact of covariates (Degree and Volatility) on the top eight influence indices.

| | Covariates | | Influence Index |
|---|---|---|---|
| No. | Degree | Volatility | |
| 1 | 1.0225 | -2.5817 | 21.4577 |
| 2 | 1.2326 | -2.0639 | 9.1265 |
| 3 | 1.2326 | -1.9946 | 8.2601 |
| 4 | 2.1569 | -1.1035 | 4.4743 |
| 5 | -0.2169 | -3.2567 | 4.2804 |
| 6 | -0.3219 | -3.1483 | 2.4625 |
| 7 | 1.6107 | -1.4622 | 1.9353 |
| 8 | 0.7705 | -2.2907 | 1.9017 |

# REFERENCES

Durrett, R. (2009) *Probability: Theory and Examples*, 4th ed, Cambridge University Press.

Fan, J. and Li, R. (2001) "Variable selection via nonconcave penalized likelihood and its oracle properties", *Journal of the American Statistical Association*, 96, 1348–1360.

Kelejian, H. and Prucha, I. (2001) "On the asymptotic distribution of the Moran I test statistic with applications", *Journal of Econometrics*, 104, 219–257.

Lee, L.F. (2004) "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models", *Econometrica*, 72, 1899–1925.

Ma, Y., Pan, R., Zou, T. and Wang, H. (2019) "A naive least squares method for spatial autoregression with covariates", *Statistica Sinica*, to appear.

Seber, G.A. (2008) *A Matrix Handbook for Statisticians*, John Wiley & Sons, Inc.

Zhou, J., Tu, Y., Chen, Y. and Wang, H. (2017) "Estimating spatial autocorrelation with sampled network data", *Journal of Business and Economics Statistics*, 35, 130–138.

Zou, T., Lan, W., Wang, H. and Tsai, C. L. (2017) "Covariance regression analysis", *Journal of the American Statistical Association*, 112, 266–281.