Supplementary Materials to “Portal Nodes Screening for Large Scale Social Networks”

APPENDIX A

Appendix A.1: Useful Lemmas

In this section we present and prove five useful lemmas, which could be employed as tools in later proofs.

Lemma 1. *Assume X follows sub-Gaussian distribution with mean 0 and moment generating function satisfying $E\{\exp(sX)\} \le \exp(\sigma^2 s^2/2)$. Then the random variable $Z = X^2 - E(X^2)$ follows sub-exponential distribution with mean 0, and the moment generating function satisfies $E\{\exp(sZ)\} \le \exp(c_z^2 s^2)$ for all $|s| \le 1/c_z$ where c_z is a positive constant.*

Proof: The proof can be found in Proposition 2.7.1 of Vershynin (2017).

Lemma 2. *Let $X = (X_1, \cdots, X_n)^\top \in \mathbb{R}^n$ and $Y = (Y_1, \cdots, Y_n)^\top \in \mathbb{R}^n$ be sub-Gaussian random vectors, with each element X_i and Y_i following sub-Gaussian distributions. Specifically, let $E(X) = \mathbf{0} \in \mathbb{R}^n$, $E(Y) = \mathbf{0} \in \mathbb{R}^n$, $cov(X) = \Sigma_x \in \mathbb{R}^{n\times n}$, $cov(Y) = \Sigma_y \in \mathbb{R}^{n\times n}$, and $cov(X,Y) = \Sigma_{xy} \in \mathbb{R}^{n\times n}$. Then, for any matrix $M \in \mathbb{R}^{n\times n}$, there exists positive constants ν, c_1, c_2, c_3, and c_4 that*

$$P\Big\{\Big|m^{-1}(Y^\top MY) - \sigma_y^{(m)}\Big| \ge \delta\Big\} \le c_1 \exp\Big\{ - c_2\sigma_{2y}^{-1} m^2\delta^2\Big\}, \tag{A.1}$$

$$P\Big\{\Big|m^{-1}(X^\top MY) - \sigma_{xy}^{(m)}\Big| \ge \delta\Big\} \le c_3 \exp\Big\{ - c_4\sigma_{2xy}^{-1} m^2\delta^2\Big\}, \tag{A.2}$$

for any $0 < \delta < \nu$, where $\sigma_y^{(m)} = m^{-1}tr(M\Sigma_y)$, $\sigma_{xy}^{(m)} = m^{-1}tr(M\Sigma_{xy})$, $\sigma_{2y} = tr(M\Sigma_y M \Sigma_y) + tr(M\Sigma_y M^\top\Sigma_y)$, $\sigma_{2xy} = tr(\Sigma_x M\Sigma_y M^\top) + tr(\Sigma_{xy} M^\top \Sigma_{xy} M^\top)$, and m is a normalizing constant.

Proof of (A.1): Note that $Y^\top MY = 2^{-1}Y^\top(M+M^\top)Y$. Let $\widetilde{Y} = \Sigma_y^{-1/2}Y$. It can be concluded $\widetilde{Y}$ follows sub-Gaussian distribution. Let $\mathbb{M} = M + M^\top$. It can be derived $Y^\top \mathbb{M} Y = \widetilde{Y}^\top(\Sigma_y^{1/2})^\top \mathbb{M}(\Sigma_y^{1/2})\widetilde{Y}$. In addition, let $\widetilde{\mathbb{M}} = (\Sigma_y^{1/2})^\top \mathbb{M}(\Sigma_y^{1/2})$, which takes a symmetric form. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of $\widetilde{\mathbb{M}}$. Since $\widetilde{\mathbb{M}}$ is a symmetric matrix, we could have the eigenvalue decomposition as $\widetilde{\mathbb{M}} = U^\top \Lambda U$, where $U = (U_1, \cdots, U_n)^\top \in \mathbb{R}^{n\times n}$ is an orthogonal matrix and $\Lambda = \text{diag}\{\lambda_1, \cdots, \lambda_n\}$. As a consequence, we have $Y^\top \mathbb{M} Y = \sum_i \lambda_i \zeta_i^2$, where $\zeta_i = U_i^\top \widetilde{Y}$ and ζ_is are *i.i.d.* from the standard sub-Gaussian distribution. It can be verified $\zeta_i^2 - 1$ satisfies sub-exponential distribution by Lemma 1. Next, one could easily verify that the sub-exponential distribution satisfies condition (P) on page 45 of Saulis and Statuleviveccius (2012), thus we have

$$P\{|m^{-1}(Y^\top MY) - \sigma_y^{(m)}| \geq \delta\} = P\{\textstyle\sum_i \lambda_i(\zeta_i^2 - 1)| \geq 2m\delta\}$$
$$\leq c_1 \exp\{-c_2(\textstyle\sum_i \lambda_i^2)^{-1} m^2\delta^2\} = c_1 \exp\{-c_2 \text{tr}^{-1}(\mathbb{M}\Sigma_y \mathbb{M}\Sigma_y) m^2\delta^2\}.$$

By noticing that $\text{tr}(\mathbb{M}\Sigma_y\mathbb{M}\Sigma_y) = 2\{\text{tr}(M\Sigma_y M\Sigma_y) + \text{tr}(M\Sigma_y M^\top \Sigma_y)\} = 2\sigma_{2y}$, (A.1) can be obtained.

Proof of (A.2): Let $Z = (X^\top, Y^\top)^\top \in \mathbb{R}^{2n}$ and $\mathbb{M}^* = (\mathbf{0}, M; M^\top, \mathbf{0}) \in \mathbb{R}^{(2n)\times(2n)}$. Then we have $X^\top MY = 2^{-1}(Z^\top \mathbb{M}^* Z)$. Therefore, (A.1) can be readily applied. Let $\Sigma_z = \text{cov}(Z) = (\Sigma_x, \Sigma_{xy}; \Sigma_{xy}^\top, \Sigma_y) \in \mathbb{R}^{(2n)\times(2n)}$. It can be verified $\text{tr}(\Sigma_z \mathbb{M}^* \Sigma_z \mathbb{M}^*) = 2\{\text{tr}(\Sigma_{xy} M^\top \Sigma_{xy} M^\top) + \text{tr}(\Sigma_x M \Sigma_y M^\top)\}$. Consequently, the desired result (A.2) can be obtained.

Lemma 3. *Assume conditions (C1)–(C3) hold for the model (2.4). Let* $\mathbb{Y}$ *and* $\mathbb{Z}$ *follow the model (2.4) with* $\Sigma_{zy} = (\mathbb{Z}^\top\mathbb{Z} - \hat{c}_y^{-1}\mathbb{Z}^\top\mathbb{Y}\mathbb{Y}^\top\mathbb{Z})/(NT)$ *and* $\widetilde{\Sigma}_{zy} = \Sigma_Z - TN^{-1}c_y^{-1}tr^2(S^{-1})\Sigma_Z\gamma\gamma^\top\Sigma_Z$, *where* $\hat{c}_y = \mathbb{Y}^\top\mathbb{Y}$ *and* $c_y = Ttr(\Sigma_Y)$. *Then it can be concluded* $\widetilde{\Sigma}_{zy}$ *is a positive definite matrix and*

$$P(\|\Sigma_{zy}^{-1} - \widetilde{\Sigma}_{zy}^{-1}\| > \epsilon) \leq \delta_{1y}^* \exp\big(-\delta_{2y}^* N^{1-2\tau}T\epsilon^2\big) + c_{1yz}^* \exp\big(-c_{2yz}^* NT\epsilon^2\big) \qquad \text{(A.3)}$$

where δ^*_{1y}, δ^*_{2y}, c^*_{1yz}, and c^*_{2yz} are finite constants, and $\|\cdot\|$ denotes the Frobenius norm of a matrix, i.e., $\|M\| = tr^{1/2}(M^\top M)$.

Proof: We separate the proof into three steps. In the first step, we prove that $\widetilde{\Sigma}_{zy}$ is positive definite. Second, we show that

$$P(\|\Sigma_{zy} - \widetilde{\Sigma}_{zy}\| > \epsilon) \leq \delta_{1y} \exp\big(- \delta_{2y} N^{1-2\tau} T \epsilon^2\big) + c_{1yz} \exp\big(- c_{2yz} N T \epsilon^2\big), \quad \text{(A.4)}$$

where δ_{1y}, δ_{2y}, c_{1yz}, and c_{2yz} are finite constants. Lastly, we prove the results of (A.3).

Step 1. ($\widetilde{\Sigma}_{zy}$ is positive definite) It suffices to prove for any $\eta \in \mathbb{R}^p$,

$$\eta^\top \Sigma_Z \eta - TN^{-1} c_y^{-1} \text{tr}^2(S^{-1})(\eta^\top \Sigma_Z \gamma)^2 > 0. \quad \text{(A.5)}$$

To this end, we derive the upper bound for $TN^{-1}c_y^{-1}\text{tr}^2(S^{-1})(\eta^\top \Sigma_Z \gamma)^2$. First by Von Neumann's trace inequality, we have $\text{tr}(S^{-1}) \leq \sum_{i=1}^N \sigma_i(S^{-1})$, where $\sigma_i(M)$ denotes the singular value of arbitrary matrix M. It can be further derived $\{\sum_i \sigma_i(S^{-1})\}^2 \leq N\{\sum_i \sigma_i^2(S^{-1})\} = N \sum_i \lambda_i\{S^{-1}(S^{-1})^\top\} = N\text{tr}(\Sigma_Y)/c_{\gamma e}$ by Cauchy inequality, where $c_{\gamma e} = \gamma^\top \Sigma_Z \gamma + \sigma_e^2$. In addition, by Cauchy inequality, we have $(\eta^\top \Sigma_Z \gamma)^2 \leq (\eta^\top \Sigma_Z \eta)$ $(\gamma^\top \Sigma_Z \gamma)$. As a result, we have $TN^{-1}c_y^{-1}\text{tr}^2(S^{-1})(\eta^\top \Sigma_Z \gamma)^2 \leq (\eta^\top \Sigma_Z \eta)(\gamma^\top \Sigma_Z \gamma)/c_{\gamma e}$. Consequently, we have $\eta^\top \Sigma_Z \eta - TN^{-1}c_y^{-1}\text{tr}^2(S^{-1})(\eta^\top \Sigma_Z \gamma)^2 \geq (1 - \gamma^\top \Sigma_Z \gamma / c_{\gamma e})(\eta^\top \Sigma_Z \eta) = \sigma_e^2 / c_{\gamma e}(\eta^\top \Sigma_Z \eta) > 0$. The desired result (A.5) can be obtained.

Step 2. (Proof (A.4)) It can be shown that $P(\|\Sigma_{zy} - \widetilde{\Sigma}_{zy}\| > \epsilon) \leq P(\|\mathbb{Z}^\top \mathbb{Z}/(NT)$ $- \Sigma_z\| > \epsilon/2) + P(\|\hat{c}_y^{-1}(NT)^{-1}\mathbb{Z}^\top \mathbb{Y}\mathbb{Y}^\top \mathbb{Z} - c_y^{-1}N^{-1}T\text{tr}^2(S^{-1})\Sigma_Z \gamma\gamma^\top \Sigma_Z\| > \epsilon/2)$. Since we have $\text{cov}(\mathbb{Z}_{k_1}, \mathbb{Z}_{k_2}) = \sigma_{Z,k_1k_2} I_{NT}$, then we have $P(\|\mathbb{Z}^\top \mathbb{Z}/(NT) - \Sigma_Z\| > \epsilon/2) \leq c_{1z}\exp(-c_{2z}NT\epsilon^2)$, where c_{1z} and c_{2z} are finite constants. Next, note that $\text{cov}(\mathbb{Y}, Z_k) = S^{-1}(\gamma^\top \Sigma_Z e_k)$ and we have $P(|\hat{c}_y/(NT) - c_y/(NT)| > \epsilon) \leq \delta_{1y}\exp(-\delta_{2y}N^2T/c_{2y}\epsilon^2)$ by Lemma 2, where $c_{2y} = \text{tr}(\Sigma_Y^2)$, δ_{1y} and δ_{2y} are finite constants. Therefore, it can be derived $P(\|\hat{c}_y^{-1}(NT)^{-1}\mathbb{Z}^\top \mathbb{Y}\mathbb{Y}^\top \mathbb{Z} - c_y^{-1}N^{-1}T\text{tr}^2(S^{-1})\Sigma_Z\gamma\gamma^\top \Sigma_Z\| > \epsilon/2) \leq \delta_{1y}\exp(-\delta_{2y}c_y^2$

$/(Tc_{2y})\epsilon^2) + 2c_{1yz}\exp\{-c_{2yz}Nc_y/\sigma_{yz}\epsilon^2\}$, where $c_{2y} = \text{tr}(\Sigma_Y^2)$ and $\sigma_{yz} = \text{tr}(\Sigma_Y) + \text{tr}(S^{-2})$. It can be derived $c_{2y} \leq N\lambda^2_{\max}(\Sigma_Y)$, $c_y \geq N\lambda_{\min}(\Sigma_Y)$, and $\sigma_{yz} \leq \text{tr}(\Sigma_Y) + \text{tr}\{S^{-1}(S^{-1})^\top\} = c_{3\gamma}\text{tr}(\Sigma_Y)$, where $c_{3\gamma} = 1 + (\gamma^\top\Sigma_Z\gamma + \sigma_e^2)^{-1}$. Note $\lambda_{\max}(\Sigma_Y) = N^\tau$ and $\lambda_{\min}(\Sigma_Y) > \tau_{\min}$ by condition (C3). Then (A.4) can be obtained by adjusting the constants.

Step 3. (Proof of (A.3)) Note $\Sigma_{zy}^{-1} - \widetilde{\Sigma}_{zy}^{-1} = \widetilde{\Sigma}_{zy}^{-1}(\widetilde{\Sigma}_{zy} - \Sigma_{zy})\Sigma_{zy}^{-1}$. Let $\Delta_{zy} = \widetilde{\Sigma}_{zy} - \Sigma_{zy}$, $\lambda_{zy} = \lambda_{\max}(\Sigma_{zy})$, and $\widetilde{\lambda}_{zy} = \lambda_{\max}(\widetilde{\Sigma}_{zy})$ Then we have $\|\Sigma_{zy}^{-1} - \widetilde{\Sigma}_{zy}^{-1}\|^2 = \text{tr}(\Sigma_{zy}^{-1}\Delta_{zy}\widetilde{\Sigma}_{zy}^{-2}\Delta_{zy}\Sigma_{zy}^{-1}) \geq \lambda_{zy}^{-2}\widetilde{\lambda}_{zy}^{-2}\|\Delta_{zy}\|^2$. Therefore we have $P(\|\Sigma_{zy}^{-1} - \widetilde{\Sigma}_{zy}^{-1}\| > \delta) \leq P\{\|\Delta_{zy}\| > \delta\lambda_{zy}\widetilde{\lambda}_{zy}\}$. Suppose δ is small that $\delta < \widetilde{\lambda}_{zy}$. Consequently we have $P\{\|\Delta_{zy}\| > \delta\lambda_{zy}\widetilde{\lambda}_{zy}\} \leq P\{\|\Delta_{zy}\| > \delta(\widetilde{\lambda}_{zy} - \delta)\widetilde{\lambda}_{zy}\} + P(\lambda_{zy} < \widetilde{\lambda}_{zy} - \delta)$. According to the Wielandt-Hoffman Theorem (Izenman, 2008), one could obtain that $|\lambda_{zy} - \widetilde{\lambda}_{zy}| \leq \|\Sigma_{zy} - \widetilde{\Sigma}_{zy}\|_2$. Therefore it can be implied that $P(\lambda_{zy} < \widetilde{\lambda}_{zy} - \delta) \leq P(|\lambda_{zy} - \widetilde{\lambda}_{zy}| > \delta) \leq P(\|\Sigma_{zy} - \widetilde{\Sigma}_{zy}\| > \delta)$. Together by (A.4), (A.3) can be obtained.

Lemma 4. *Let $Y \in \mathbb{R}^{N_y}$, $X_1 \in \mathbb{R}^{N_{1x}}$, and $X_2 \in \mathbb{R}^{N_{2x}}$ are sub-Gaussian random vectors with $cov(Y) = \Sigma_y$, $cov(X_1) = \Sigma_{1x}$, and $cov(X_2) = \Sigma_{2x}$. In addition, let $M_1 \in \mathbb{R}^{N_y \times N_y}$ and $M_2 \in \mathbb{R}^{N_y \times N_y}$. Define $\widehat{\xi}_{1y} = (Y^\top M_1 Y)/N_{1m}$, $\widehat{\xi}_{2y} = (Y^\top M_2 Y)/N_{2m}$, $\widehat{\xi}_{1x} = (X_1^\top X_1)/N_{1x}$, and $\widehat{\xi}_{2x} = (X_2^\top X_2)/N_{2x}$, where N_{1m} and N_{2m} are normalizing constants. Accordingly, let $\mu_{1y} = E(\widehat{\xi}_{1y})$, $\mu_{2y} = E(\widehat{\xi}_{2y})$, $\mu_{1x} = E(\widehat{\xi}_{1x}) > 0$, $\mu_{2x} = E(\widehat{\xi}_{3x}) > 0$.*

(a) Then for a sufficiently small δ, we then have

$$P\Big(\big|\widehat{\xi}_{1x}^{-1}\widehat{\xi}_{1y} - \mu_{1x}^{-1}\mu_{1y}\big| > \delta\Big) \leq \Delta_{1m} + \Delta_{1x} \tag{A.6}$$

$$P\Big(\big|\widehat{\xi}_{1y}\widehat{\xi}_{2y}/(\widehat{\xi}_{1x}\widehat{\xi}_{2x}) - \mu_{1y}\mu_{2y}/(\mu_{1x}\mu_{2x})\big| > \delta\Big) \leq \Delta_{1m} + \Delta_{2m} + \Delta_{1x} + \Delta_{2x} + \widetilde{\Delta}_{1m} + \widetilde{\Delta}_{2m}, \tag{A.7}$$

where $\Delta_{1m} = c_1\exp(-c_2\sigma_{1m}^{-1}N_{1m}^2\mu_{1x}^2\delta^2)$, $\Delta_{2m} = c_5\exp(-c_6\sigma_{2m}^{-1}N_{2m}^2\mu_{2x}^2\delta^2)$, $\Delta_{1x} = c_3\exp(-c_4\sigma_{1x}^{-1}N_{1x}^2\mu_{1x}^2)$, $\Delta_{2x} = c_7\exp(-c_8\sigma_{2x}^{-1}N_{2x}^2\mu_{2x}^2)$, $\widetilde{\Delta}_{1m} = c_1\exp(-c_2\sigma_{1m}^{-1}N_{1m}^2\mu_{1x}^2\mu_{2x}^2\mu_{2y}^{-2}\delta^2)$, $\widetilde{\Delta}_{2m} = c_5\exp(-c_6\sigma_{2m}^{-1}N_{2m}^2\mu_{2x}^2\mu_{1x}^2\mu_{1y}^{-2}\delta^2)$, $\sigma_{1m} = tr(M_1\Sigma_y M_1\Sigma_y) + tr(M_1\Sigma_y M_1^\top\Sigma_y)$,

$\sigma_{2m} = tr(M_2\Sigma_y M_2\Sigma_y) + tr(M_2\Sigma_y M_2^\top\Sigma_y)$, $\sigma_{1x} = tr(\Sigma_{1x}^2)$, $\sigma_{2x} = tr(\Sigma_{2x}^2)$, and c_j ($1 \le j \le 8$) are finite positive constants.

(b) Let $\mathbb{Z} = (Z_k) \in \mathbb{R}^{N_y\times p}$, where Z_k following sub-Gaussian distribution with $E(Z_k) = \mathbf{0}$ and $cov(Z_{k_1}, Z_{k_2}) = \sigma_{z,k_1k_2} I_{N_y}$. In addition, let $\Sigma_z = (\sigma_{z,k_1k_2}) \in \mathbb{R}^{p\times p}$ and assume $cov(Y, Z_k) = (e_k^\top\Sigma_z\gamma)\Sigma_{zy}$, where $e_k \in \mathbb{R}^p$ is a p-dimensional zero vector except the kth element being 1, $\gamma \in \mathbb{R}^p$ is a p-dimensional constant vector, and $\Sigma_{zy} \in \mathbb{R}^{N_y\times N_y}$. Define $\widehat{\xi}_{1yz} = \mathbb{Z}^\top M_1 Y/N_y$, $\widehat{\xi}_{2yz} = \mathbb{Z}^\top M_2 Y/N_y$, and accordingly $\mu_{1yz} = E(\widehat{\xi}_{1yz})$, $\mu_{2yz} = E(\widehat{\xi}_{2yz})$. In addition, assume $\widehat{\Omega} \in \mathbb{R}^{p\times p}$ be a random matrix and for a sufficiently small $\epsilon > 0$, it holds

$$P(\|\widehat{\Omega} - \Omega\| > \epsilon) \le \Delta_\omega(\epsilon), \tag{A.8}$$

where $\Delta_\omega(\epsilon)$ is a positive constant related to ϵ. It is assumed $\omega_{\min} \le \lambda_{\min}(\Omega) \le \lambda_{\max}(\Omega) \le \omega_{\max}$, where $\omega_{\min}$ and $\omega_{\max}$ are finite positive constants. Then for a sufficiently small δ, we then have

$$\begin{aligned} P\Big(\big\|\widehat{\xi}_{1x}^{-1}\widehat{\Omega}(\widehat{\xi}_{1yz}\widehat{\xi}_{2yz}^\top) - \mu_{1x}^{-1}\Omega(\mu_{1yz}\mu_{2yz}^\top)\big\| > \delta\Big) \le& \Delta_\omega(\delta) + \Delta_\omega\Big(\frac{\delta\mu_{1x}}{\|\mu_{1yz}\|\|\mu_{2yz}\|}\Big) \\ &+ \Delta_x + \Delta_{1yz} + \Delta_{2yz}, \end{aligned} \tag{A.9}$$

where $\Delta_x = c_{1x}\exp(-c_{2x}\sigma_{1x}^{-1}N_{1x}^2\mu_{1x}^2)$, $\Delta_{1yz} = c_{1yz}^a\exp(-c_{1yz}^b N_y^2\sigma_{1yz}^{-1}\mu_{1x}\delta^2)$, $\Delta_{2yz} = c_{2yz}^a\exp(-c_{2yz}^b N_y^2\sigma_{2yz}^{-1}\mu_{1x}\delta^2)$, $\sigma_{1yz} = tr(M_1\Sigma_y M_1^\top) + tr(\Sigma_{zy}M_1^\top\Sigma_{zy}M_1^\top)$, $\sigma_{2yz} = tr(M_2\Sigma_y M_2^\top) + tr(\Sigma_{zy}M_2^\top\Sigma_{zy}M_2^\top)$, and $c_{1\omega}$, $c_{2\omega}$, c_{1x}, c_{2x}, c_{1yz}^a, c_{1yz}^b, c_{2yz}^a, c_{2yz}^b are finite constants.

Proof of (a): For simplicity, we only prove the first inequality of (A.6). The second one can be obtained by iteratively applying the same technique.

PROOF OF (A.6). First we have $|\widehat{\xi}_{1y}/\widehat{\xi}_{1x} - \mu_{1y}/\mu_{1x}| \le |\widehat{\xi}_{1y}/\widehat{\xi}_{1x} - \mu_{1y}/\widehat{\xi}_{1x}| + |\mu_{1y}/\widehat{\xi}_{1x} - \mu_{1y}/\mu_{1x}|$. It can be concluded $P(|\widehat{\xi}_{1y}/\widehat{\xi}_{1x} - \mu_{1y}/\mu_{1x}| > \delta) \le P(|\widehat{\xi}_{1y}/\widehat{\xi}_{1x} - \mu_{1y}/\widehat{\xi}_{1x}| > \delta/2) + P(|\mu_{1y}/\widehat{\xi}_{1x} - \mu_{1y}/\mu_{1x}| > \delta/2)$. We then derive the upper bound for the two parts respectively in the following.

Part I. It can be derived

$$P(|\widehat{\xi}_{1y}/\widehat{\xi}_{1x} - \mu_{1y}/\widehat{\xi}_{1x}| > \delta/2) = P\Big(\frac{|\widehat{\xi}_{1y} - \mu_{1y}|}{\mu_{1x}}\frac{\mu_{1x}}{\widehat{\xi}_{1x}} > \delta/2\Big)$$
$$\leq P\Big(\frac{|\widehat{\xi}_{1y} - \mu_{1y}|}{\mu_{1x}} > \delta/4\Big) + P\Big(\frac{\mu_{1x}}{\widehat{\xi}_{1x}} > 2\Big) \tag{A.10}$$

By Lemma 2, it can be derived $P(|\widehat{\xi}_{1y} - \mu_{1y}| > \delta\mu_{1x}/4) \leq \alpha_1 \exp(-\alpha_2\sigma_{1m}^{-1}N_{1m}^2\mu_{1x}^2\delta^2)$, where α_1 and α_2 are finite constants. Next, we have $P(\mu_{1x} > 2\widehat{\xi}_{1x}) = P\{2(\widehat{\xi}_{1x} - \mu_{1x}) < -\mu_{1x}\} \leq P\{|\widehat{\xi}_{1x} - \mu_{1x}| > 1/2\mu_{1x}\}$. By Lemma 2, we have $P\{|\widehat{\xi}_{1x} - \mu_{1x}| > 1/2\mu_{1x}\} \leq c_3 \exp(-c_4\sigma_{1x}^{-1}N_{1x}^2\mu_{1x}^2)$. By summarizing the results in Part I and Part II and rearranging the constants, the desired results in (A.6) can be obtained.

Part II. Without loss of generality, we assume $\mu_{1y} > 0$. Let $\delta^* = (2\mu_{1y}/\mu_{1x})/(1 + 2\mu_{1y}/\mu_{1x})\delta$. Therefore, we have $\delta^* < \delta$ and hence $P(|\mu_{1y}/\widehat{\xi}_{1x} - \mu_{1y}/\mu_{1x}| > \delta/2) \leq P(|\mu_{1y}/\widehat{\xi}_{1x} - \mu_{1y}/\mu_{1x}| > \delta^*/2) \leq P(\mu_{1y}/\widehat{\xi}_{1x} > \delta^*/2 + \mu_{1y}/\mu_{1x}) + P(\mu_{1y}/\widehat{\xi}_{1x} < -\delta^*/2 + \mu_{1y}/\mu_{1x})$. Then we have $P(\mu_{1y}/\widehat{\xi}_{1x} > \delta^*/2 + \mu_{1y}/\mu_{1x}) = P(\mu_{1y}/\widehat{\xi}_{1x} > \{1 + \delta/(1 + 2\mu_{1y}/\mu_{1x})\}\mu_{1y}/\mu_{1x}) = P(\widehat{\xi}_{1x} - \mu_{1x} < -\delta/(1 + 2\mu_{1y}/\mu_{1x} + \delta)\mu_{1x})$. Similarly we can obtain $P(\mu_{1y}/\widehat{\xi}_{1x} < -\delta^*/2 + \mu_{1y}/\mu_{1x}) = P(\widehat{\xi}_{1x} - \mu_{1x} > \delta\mu_{1x}/(1 + 2\mu_{1y}/\mu_{1x} - \delta))$. Consequently we obtain $P(|\mu_{1y}/\widehat{\xi}_{1x} - \mu_{1y}/\mu_{1x}| > \delta/2) \leq \alpha_5 \exp(-\alpha_6\sigma_{1x}^{-1}N_{1x}^2\mu_{1x}^2\delta^2)$, where α_5 and α_6 are finite constants.

Proof of (A.7). It can be noted that

$$\frac{\widehat{\xi}_{1y}\widehat{\xi}_{2y}}{\widehat{\xi}_{1x}\widehat{\xi}_{2x}} - \frac{\mu_{1y}\mu_{2y}}{\mu_{1x}\mu_{2x}} = \Big(\frac{\widehat{\xi}_{1y}}{\widehat{\xi}_{1x}} - \frac{\mu_{1y}}{\mu_{1x}}\Big)\Big(\frac{\widehat{\xi}_{2y}}{\widehat{\xi}_{2x}} - \frac{\mu_{2y}}{\mu_{2x}}\Big) + \frac{\mu_{1y}}{\mu_{1x}}\Big(\frac{\widehat{\xi}_{2y}}{\widehat{\xi}_{2x}} - \frac{\mu_{2y}}{\mu_{2x}}\Big) + \frac{\mu_{2y}}{\mu_{2x}}\Big(\frac{\widehat{\xi}_{1y}}{\widehat{\xi}_{1x}} - \frac{\mu_{1y}}{\mu_{1x}}\Big).$$

Consequently, (A.7) can be obtained by applying the same proof technique of (A.6) to each part separately.

Proof of (b): Let $\widehat{\xi}^*_{1yz} = \widehat{\xi}_{1x}^{-1/2}\widehat{\xi}_{1yz}$ and $\widehat{\xi}^*_{2yz} = \widehat{\xi}_{1x}^{-1/2}\widehat{\xi}_{2yz}$. Accordingly, let $\mu^*_{1yz} = \mu_{1x}^{-1/2}\mu_{1yz}$ and $\mu^*_{2yz} = \mu_{1x}^{-1/2}\mu_{2yz}$. In this part, we derive upper bound for $P\big(\|\widehat{\Omega}(\widehat{\xi}^*_{1yz}\widehat{\xi}^{*\top}_{2yz}) -$

$\Omega(\mu^*_{1yz}\mu^{*\top}_{2yz})\big\| > \delta\big)$. Then the results can be obtained by using (A.6). It can be noted $\widehat{\Omega}(\widehat{\xi}^*_{1yz}\widehat{\xi}^\top_{2yz}) - \Omega(\mu^*_{1yz}\mu^{*\top}_{2yz}) = (\widehat{\Omega}-\Omega)(\widehat{\xi}^*_{1yz}\widehat{\xi}^\top_{2yz} - \mu^*_{1yz}\mu^{*\top}_{2yz}) + \Omega(\widehat{\xi}^*_{1yz}\widehat{\xi}^\top_{2yz} - \mu^*_{1yz}\mu^{*\top}_{2yz}) + (\widehat{\Omega} - \Omega)\mu^*_{1yz}\mu^{*\top}_{2yz}$. Therefore we have

$$P\big(\big\|\widehat{\Omega}(\widehat{\xi}^*_{1yz}\widehat{\xi}^\top_{2yz}) - \Omega(\mu^*_{1yz}\mu^{*\top}_{2yz})\big\| > \delta\big) \le P(\|(\widehat{\Omega}-\Omega)\mu^*_{1yz}\mu^{*\top}_{2yz}\| > \delta/3)$$
$$+ P(\|\Omega(\widehat{\xi}^*_{1yz}\widehat{\xi}^{*\top}_{2yz} - \mu^*_{1yz}\mu^{*\top}_{2yz})\| > \delta/3) + P(\|(\widehat{\Omega}-\Omega)(\widehat{\xi}^*_{1yz}\widehat{\xi}^{*\top}_{2yz} - \mu^*_{1yz}\mu^{*\top}_{2yz})\| > \delta/3).$$

We next look at the above three terms one by one. Without loss of generality, we assume $\mu^*_{1yz}\mu^{*\top}_{2yz} \neq \mathbf{0}$. Then we have $\|(\widehat{\Omega}-\Omega)\mu^*_{1yz}\mu^{*\top}_{2yz}\| = (\mu^{*\top}_{2yz}\mu_{2yz})^{1/2}\mathrm{tr}^{1/2}\{(\widehat{\Omega} - \Omega)\mu^*_{1yz}\mu^{*\top}_{1yz}(\widehat{\Omega}-\Omega)\} = (\mu^{*\top}_{2yz}\mu_{2yz})^{1/2}\{\mu^{*\top}_{1yz}(\widehat{\Omega}-\Omega)^2\mu^*_{1yz}\}^{1/2} \ge \|\mu^*_{2yz}\|\|\mu^*_{1yz}\||\lambda_{\min}(\widehat{\Omega}-\Omega)|$. Therefore we have $P(\|(\widehat{\Omega}-\Omega)\mu^*_{1yz}\mu^{*\top}_{2yz}\| > \delta/3) \le P(|\lambda_{\min}(\widehat{\Omega}-\Omega)| > 3^{-1}\|\mu^*_{1yz}\|^{-1}\|\mu^*_{2yz}\|^{-1}\delta)$. By (A.8), $P(|\lambda_{\min}(\widehat{\Omega}-\Omega)| > 3^{-1}\|\mu^*_{1yz}\|^{-1}\|\mu^*_{2yz}\|^{-1}\delta) \le \Delta_\omega(\delta\mu_{1x}\|\mu_{1yz}\|^{-1}\|\mu_{2yz}\|^{-1})$. Next, let $U_{yz} = \widehat{\xi}^*_{1yz}\widehat{\xi}^{*\top}_{2yz} - \mu^*_{1yz}\mu^{*\top}_{2yz}$, where ω^*_2 is a positive constant. Then we have $\|\Omega U_{yz}\| = \mathrm{tr}^{1/2}\{U^\top_{yz}\Omega^2 U_{yz}\} \ge \lambda_{\min}(\Omega)\|U_{yz}\|$. Therefore we have $P(\|\Omega U_{yz}\| > \delta/3) \le P(\|U_{yz}\| > 3^{-1}\lambda^{-1}_{\min}(\Omega)\delta)$. Lastly, for the last term we have $P(\|(\widehat{\Omega}-\Omega)U_{yz}\| > \delta/3) \le P(\|\widehat{\Omega}-\Omega\| > \sqrt{\delta/3}) + P(\|U_{yz}\| > \sqrt{\delta/3})$. Consequently, it suffices to derive the rate of

$$P(\|\widehat{\xi}^*_{1yz}\widehat{\xi}^{*\top}_{2yz} - \mu^*_{1yz}\mu^{*\top}_{2yz}\| > \delta_1), \tag{A.11}$$

where $\delta_1 = \min\{\sqrt{\delta/3}, \delta/(3\lambda_{\min}(\Omega))\}$. In other words, it suffices to derive $P(|\eta^\top\widehat{\xi}^*_{1yz}\widehat{\xi}^{*\top}_{2yz}\eta - \eta^\top\mu^*_{1yz}\mu^{*\top}_{2yz}\eta| > \delta_1)$ for any $\eta \in \mathbb{R}^p$ with $\|\eta\| = 1$. By similar arguments, it can be derived that $P(|\eta^\top\widehat{\xi}^*_{1yz}\widehat{\xi}^{*\top}_{2yz}\eta - \eta^\top\mu^*_{1yz}\mu^{*\top}_{2yz}\eta| > \delta_1) \le P(|\eta^\top\widehat{\xi}^*_{1yz} - \eta^\top\mu^*_{1yz}| > \delta_2) + P(|\eta^\top\widehat{\xi}^*_{2yz} - \eta^\top\mu^*_{2yz}| > \delta_2)$, where δ_2 is a finite positive constant. Note $\eta^\top\widehat{\xi}_{1yz} = (\eta^\top\mathbb{Z}^\top M_1 Y)/N_y$. Let $\mathcal{Y} = ((\mathbb{Z}\eta)^\top, Y^\top)^\top \in \mathbb{R}^{(2N_y)}$. We then have $\eta^\top\widehat{\xi}_{1yz} = \mathcal{Y}^\top M^*_1\mathcal{Y}/2$, where $M^*_1 = (0, M_1; M^\top_1, 0) \in \mathbb{R}^{(2N_y)\times(2N_y)}$. It can be derived $\Sigma_{\mathcal{Y}} = \mathrm{cov}(\mathcal{Y}) = (\sigma_\eta I_{N_y}, \Sigma^\eta_{zy}; \Sigma^{\eta\top}_{zy}, \Sigma_Y)$, where $\sigma_\eta = \eta^\top\Sigma_z\eta$ and $\Sigma^\eta_{zy} = \mathrm{cov}(\mathbb{Z}\eta, Y) = (\eta^\top\Sigma_Z\gamma)\Sigma_{zy}$. We then have $\mathrm{tr}(\Sigma_{\mathcal{Y}}M^*_1\Sigma_{\mathcal{Y}}M^*_1) = 2\sigma_\eta\mathrm{tr}(M_1\Sigma_Y M^\top_1) + 2(\eta^\top\Sigma_Z\gamma)^2\mathrm{tr}(\Sigma_{zy}M^\top_1\Sigma_{zy}M^\top_1)$. Moreover, we have $\eta^\top\mu_{1yz} = \mathrm{cov}(M_1 Y, \mathbb{Z}\eta)/N_y = (\eta^\top\Sigma_Z\gamma)\mathrm{tr}(M_1\Sigma_{zy})/N_y$. Then by

(A.2) of Lemma 2 and (A.6) of Lemma 4, we have $P(|\eta^\top \widehat{\xi}^*_{1yz} - \eta^\top \mu^*_{1yz}| > \delta_2) \le c^*_{1yz} \exp(-c_{2yz} N_y^2 \sigma^{-1}_{1yz} \mu_{1x} \delta_2^2) + \Delta_{1x}$. Consequently, (A.9) can be obtained.

Lemma 5. *Let $\Sigma \in \mathbb{R}^{m\times m}$, and $\widehat{\Sigma}$ be its estimate. Assume for any $\epsilon > 0$, Σ and $\widehat{\Sigma}$ satisfy*

$$\tau_{\min} \le \lambda_{\min}(\Sigma) \le \lambda_{\max}(\Sigma) \le \tau_{\max}, \tag{A.12}$$

$$\textit{and} \quad P\{\|\widehat{\Sigma} - \Sigma\|_\infty \ge \epsilon\} \le c_1 \exp(-c_2 T \epsilon^2 + c_3 \log m) \tag{A.13}$$

where $0 < \tau_{\min} < \tau_{\max}$, c_1, c_2, c_3 are positive constants. In addition, if $m = O(T^{\delta_1})$ with $0 \le \delta_1 < 1/2$, then we have for a positive constant c_4,

$$P\Big(\sup_{\|r\|=1} \big|r^\top (\widehat{\Sigma} - \Sigma) r\big| > \epsilon\Big) \le c_1 \exp(-c_2 T m^{-2} \epsilon^2 + c_3 \log m + c_4 m) \tag{A.14}$$

$$\textit{and} \quad \tau_{\min}/2 \le \lambda_{\min}(\widehat{\Sigma}) \le \lambda_{\max}(\widehat{\Sigma}) \le 2\tau_{\max} \textit{ with probability tending to 1,} \tag{A.15}$$

where c_1, c_2, c_3 are positive constants.

Proof: Note that by (A.12) and (A.14), the conclusion (A.15) is implied by the condition that $m = O(T^{\delta_1})$ with $0 \le \delta_1 < 1/2$. Thus, let us prove (A.14).

For any $\|r\| = 1$, we have

$$\begin{aligned} |r^\top (\widehat{\Sigma} - \Sigma) r| &\le \sum_{j_1, j_2} |r_{j_1} r_{j_2}| |\widehat{\sigma}_{j_1 j_2} - \sigma_{j_1 j_2}| \\ &\le \|\widehat{\Sigma} - \Sigma\|_\infty \sum_{j_1, j_2} |r_{j_1} r_{j_2}| = \|\widehat{\Sigma} - \Sigma\|_\infty (\sum_{j=1}^m |r_j|)^2 \\ &= \|\widehat{\Sigma} - \Sigma\|_\infty \|r\|_1^2 \le m \|\widehat{\Sigma} - \Sigma\|_\infty \end{aligned}$$

Therefore, we have $P\{|r^\top \{\widehat{\Sigma} - \Sigma\} r| > \epsilon\} \le P\{\|\widehat{\Sigma} - \Sigma\|_\infty > \epsilon/m\} \le c_1 \exp(-c_2 T m^{-2} \epsilon^2 + c_3 \log m)$. Lastly, we apply the discretization argument (Lemma F.2 of Basu et al.

(2015)) and then the result (A.14) could be obtained.

Appendix A.2: Proof of Proposition 1

It suffices to show for a sufficiently small δ_1, we have $P\big\{\max_j \big|\widehat{R}_j^2 - R_j^2\big| > \delta_1\big\} \to 0$. We first derive the form of $\widehat{R}_j^2$. To this end, we first give $(\widetilde{\mathbb{Y}}^\top\widetilde{\mathbb{Y}})^{-1}$. Let $\hat{c}_y = \mathbb{Y}^\top\mathbb{Y}$ and $\Omega_{zy} = (\mathbb{Z}^\top\mathbb{Z} - \hat{c}_y^{-1}\mathbb{Z}^\top\mathbb{Y}\mathbb{Y}^\top\mathbb{Z})^{-1}$. We then have

$$(\widetilde{\mathbb{Y}}^\top\widetilde{\mathbb{Y}})^{-1} = \begin{pmatrix} \hat{c}_y^{-1} + \hat{c}_y^{-2}\mathbb{Y}^\top\mathbb{Z}\Omega_{zy}\mathbb{Z}^\top\mathbb{Y} & -\hat{c}_y^{-1}\mathbb{Y}^\top\mathbb{Z}\Omega_{zy} \\ -\hat{c}_y^{-1}\Omega_{zy}\mathbb{Z}^\top\mathbb{Y} & \Omega_{zy} \end{pmatrix}. \tag{A.16}$$

It can be noted $X_t^{(j)} = W_{\cdot j}Y_{jt} = (W_{\cdot j}e_j^\top)Y_t$, where $X_t^{(j)}$ is the jth column of X_t. Let $\mathbb{M}_j = I_T \otimes (W_{\cdot j}e_j^\top)$, $\xi_{1j} = \mathbb{X}_j^\top\mathbb{X}_j$, and $\xi_{2j} = \mathbb{Y}^\top\mathbb{M}_j\mathbb{Y}$, where $e_j \in \mathbb{R}^N$ is a vector with the jth element being 1 and others being 0. Define

$$\widehat{R}_{1j} = \xi_{1j}^{-1}\hat{c}_y^{-1}\xi_{2j}^2, \quad \widehat{R}_{2j} = \xi_{1j}^{-1}(\mathbb{Y}^\top\mathbb{M}_j\mathbb{Z}\Omega_{zy}\mathbb{Z}^\top\mathbb{M}_j^\top\mathbb{Y}),$$
$$\widehat{R}_{3j} = -2\xi_{1j}^{-1}\hat{c}_y^{-1}\xi_{2j}(\mathbb{Y}^\top\mathbb{M}_j\mathbb{Z}\Omega_{zy}\mathbb{Z}^\top\mathbb{Y}), \quad \widehat{R}_{4j} = \xi_{1j}^{-1}\hat{c}_y^{-2}(\mathbb{Y}^\top\mathbb{Z}\Omega_{zy}\mathbb{Z}^\top\mathbb{Y}).$$

Consequently, $\widehat{R}_j^2$ can be expressed as $\widehat{R}_j^2 = \widehat{R}_{1j} + \widehat{R}_{2j} + \widehat{R}_{3j} + \widehat{R}_{4j}$. Accordingly, define $R_{1j} = (\kappa_{1j}\sigma_{Y,jj}c_y)^{-1}\kappa_{2j}^2$, $R_{2j} = (N\kappa_{1j}\sigma_{Y,jj})^{-1}\kappa_{3j}^2c_z$, $R_{3j} = (N\kappa_{1j}\sigma_{Y,jj}c_y^2)^{-1}(\kappa_{2j}^2c_s^2c_z)$, and $R_{4j} = -2(N\kappa_{1j}\sigma_{Y,jj}c_y)^{-1}\kappa_{2j}\kappa_{3j}c_sc_z$. Hence we have $R_j^2 = R_{1j} + R_{2j} + R_{3j} + R_{4j}$. Therefore we have

$$P\big\{\big|\widehat{R}_j^2 - R_j^2\big| > \delta_1\big\} \le \sum_{k=1}^4 P\big\{\big|\widehat{R}_{kj} - R_{kj}\big| > \delta_1/4\big\}. \tag{A.17}$$

It suffices to show $\sum_{j=1}^N P\big\{\big|\widehat{R}_{kj} - R_{kj}\big| > \delta_1/4\big\} \to 0$ for $1 \le k \le 4$. For the sake of similarity, we prove the case for $k = 1, 2$ in the following two parts.

Part 1. (Proof of $\sum_{j=1}^N P\big\{\big|\widehat{R}_{1j} - R_{1j}\big| > \delta_1/4\big\} \to 0$). Let $\widehat{R}_{1j}^* = (NT^2)^{-1}(W_{\cdot j}^\top W_{\cdot j})^{-1}$ $(\mathbb{Y}^\top\mathbb{M}_j\mathbb{Y})^2$, $\widehat{\sigma}_{Y,jj} = T^{-1}\sum_t Y_{jt}^2$, and $\widehat{\sigma}_Y^2 = \mathbb{Y}^\top\mathbb{Y}/(NT)$. Accordingly, set $R_{1j}^* =$

$(NT^2)^{-1}\kappa_{1j}^{-1}\kappa_{2j}^2$, $\sigma_Y^2 = N^{-1}\mathrm{tr}(\Sigma_Y)$. Consequently we have $|\widehat{R}_{1j} - R_{1j}| = |\widehat{\sigma}_{Y,jj}^{-1}\widehat{\sigma}_Y^{-2}\widehat{R}_{1j}^* - \sigma_{Y,jj}^{-1}\sigma_Y^{-2}R_{1j}^*|$, where $\widehat{R}_{1j}^* = \{(T^{-1}N^{-1/2}\kappa_{1j}^{-1/2})(\mathbb{Y}^\top\mathbb{M}_j\mathbb{Y})\}^2$. Note that we have $\widehat{R}_{1j} = (\widehat{R}_{1j}^{*1/2}/\widehat{\sigma}_{Y,jj})(\widehat{R}_{1j}^{*1/2}/\widehat{\sigma}_Y^2)$. Therefore by Lemma 4,

$$
\begin{aligned}
P(|\widehat{\sigma}_{Y,jj}^{-1}\widehat{\sigma}_Y^{-2}\widehat{R}_{1j}^* - \sigma_{Y,jj}^{-1}\sigma_Y^{-2}R_{1j}^*| > \delta_1/4) \le & \underbrace{c_1\exp(-c_2TN\kappa_{1j}\sigma_{1m}^{-1}\sigma_{Y,jj}^2\delta_1^2)}_{:=\Delta_1} \\
& + \underbrace{c_3\exp(-c_4TN\kappa_{1j}\sigma_{1m}^{-1}\sigma_Y^4\delta_1^2)}_{:=\Delta_2} + \underbrace{2c_5\exp(-c_6TN\kappa_{1j}\sigma_{1m}^{-1}\sigma_Y^4\sigma_{Y,jj}^2R_{1j}^{*-1}\delta_1^2)}_{:=\Delta_3} \\
& + \underbrace{c_7\exp\{-c_8T\mathrm{tr}^{-1}(\Sigma_Y^2)\mathrm{tr}^2(\Sigma_Y)\}}_{:=\Delta_4} + \underbrace{c_9\exp(-c_{10}T)}_{:=\Delta_5}
\end{aligned}
$$

where $\sigma_{1m} = (W_{\cdot j}^\top\Sigma_Y e_j)^2 + (W_{\cdot j}^\top\Sigma_Y W_{\cdot j})\sigma_{Y,jj}$, c_js $(1 \le j \le 6)$ are finite constants. Further it can be calculated that $\sigma_{1m} \le 2(W_{\cdot j}^\top\Sigma_Y W_{\cdot j})\sigma_{Y,jj} \le 2(W_{\cdot j}^\top W_{\cdot j})\sigma_{Y,jj}\lambda_{\max}(\Sigma_Y)$. Moreover, we have and $\sigma_{Y,jj} \ge \lambda_{\min}(\Sigma_Y)$ and $\sigma_Y^2 \ge \lambda_{\min}(\Sigma_Y)$. Therefore, it can be shown that $\Delta_1 \le c_1\exp(-c_2TN\tau_{\max}^{-1}\tau_{\min}\delta_1^2)$ (by (C3)). Similarly, we have $\Delta_2 \le c_3\exp(-c_4TN\tau_{\max}^{-2}\tau_{\min}^2\delta_1^2)$, $\Delta_3 \le c_5\exp(-c_6TN\tau_{\max}^{-2}\tau_{\min}^2\delta_1^2)$, and $\Delta_4 \le c_7\exp(-c_8TN\tau_{\max}^{-2}$ $\tau_{\min}^2\delta_1^2)$. Consequently, it can be derived $P(|\widehat{\sigma}_{Y,jj}^{-1}\widehat{\sigma}_Y^{-2}\widehat{R}_{1j}^* - \sigma_{Y,jj}^{-1}\sigma_Y^{-2}R_{1j}^*| > \delta_1/4) \le \alpha_1\exp(-\alpha_2TN^{1-2\tau}\delta_1^2) + \alpha_3\exp(-\alpha_4T\delta_1^2)$, where α_j for $1 \le j \le 4$ are finite constants. Note that $\tau < 1/2$ and $T = O((N^{2(1-\zeta)}\log N)^\xi)$ for $\xi > 1$, we then have $\sum_{j=1}^N P(|\widehat{R}_{1j} - R_{1j}| > \delta_1/4) \to 0$.

Part 2. (Proof of $\sum_{j=1}^N P\{|\widehat{R}_{2j} - R_{2j}| > \delta_1/4\} \to 0$) We re-write $\widehat{R}_{2j}$ as

$$
\xi_{1j}^{-1}\mathbb{Y}^\top\mathbb{M}_j\mathbb{Z}\Omega_{zy}\mathbb{Z}^\top\mathbb{M}_j^\top\mathbb{Y} = \xi_{1j}^{-1}\mathrm{tr}\{\Omega_{zy}(\mathbb{Z}^\top\mathbb{M}_j^\top\mathbb{Y}\mathbb{Y}^\top\mathbb{M}_j\mathbb{Z})\} = \mathrm{tr}\Big(\widehat{\sigma}_{Y,jj}^{-1}\Sigma_{zy}^{-1}\widehat{R}_{2j}^*\Big) \quad (A.18)
$$

where $\Sigma_{zy} = \Omega_{zy}^{-1}/(NT)$, $\widehat{R}_{2j}^* = \kappa_{1j}^{-1}(NT^2)^{-1}(\mathbb{Z}^\top\mathbb{M}_j^\top\mathbb{Y}\mathbb{Y}^\top\mathbb{M}_j\mathbb{Z})$. Note we have $E(\mathbb{Z}^\top\mathbb{M}_j^\top$ $\mathbb{Y}) = \mathrm{tr}(\mathbb{M}_j^\top S^{-1})\Sigma_Z\gamma = T(W_{\cdot j}^\top S^{-1}e_j)\Sigma_Z\gamma = T\kappa_{3j}\Sigma_Z\gamma$. Consequently, one could verify that $R_{2j} = \mathrm{tr}(\sigma_{Y,jj}^{-1}\widehat{\Sigma}_{zy}^{-1}R_{2j}^*)$, where $R_{2j}^* = \kappa_{1j}^{-1}N^{-1}\kappa_{3j}^2\Sigma_Z\gamma\gamma^\top\Sigma_Z$. Next, we apply (A.9) to obtain the results that $P\{\|\widehat{\sigma}_{Y,jj}^{-1}\Sigma_{zy}^{-1}\widehat{R}_{2j}^* - \sigma_{Y,jj}^{-1}\Sigma_{zy}^{-1}R_{2j}^*\| > \delta_1/4\} \le \Delta_\omega(\delta_1) + \Delta_\omega(\sigma_{Y,jj}\kappa_{1j}\kappa_{3j}^{-2}N\|\Sigma_Z\gamma\|^{-2}\delta_1) + \Delta_x + \Delta_{1yz} + \Delta_{2yz}$, where in this case we have

$\Delta_{1yz} = \Delta_{2yz}$. Note here we have $P(\|NT\Omega_{zy} - \Sigma_{zy}\| > \epsilon) \leq \Delta_\omega(\delta_1)$ by Lemma 3, where $\Delta_\omega(\delta_1) = \delta_{1y}^* \exp\big(-\delta_{2y}^* N^{1-2\tau} T\delta_1^2\big) + c_{1yz}^* \exp\big(-c_{2yz}^* NT\delta_1^2\big) \to 0$. It can be derived $\kappa_{3j}^2 \leq e_j^\top S^{-1} S^{-1\top} e_j (W_{\cdot j}^\top W_{\cdot j}) = \sigma_{Y,jj}\kappa_{1j}/c_{\gamma e}$. Therefore we have $\sigma_{Y,jj}\kappa_{1j}\kappa_{3j}^{-2} \geq c_{\gamma e}$. As a result, we have $\Delta_\omega(\sigma_{Y,jj}\kappa_{1j}\kappa_{3j}^{-2} N\|\Sigma_Z\gamma\|^{-2}\delta_1) \leq \Delta_\omega(c_{\gamma e}\|\Sigma_Z\gamma\|^{-2}N\delta_1) \to 0$. Next, we have $\Delta_x = c_{1x}\exp(-c_{2x}\sigma_{1x}^{-1}T\mu_{1x}^2\delta_1^2)$, where $\sigma_{1x} = \sigma_{Y,jj}^2$ and $\mu_{1x} = \sigma_{Y,jj}$. Consequently, we have $\Delta_x = c_{1x}\exp(-c_{2x}T\delta_1^2)$. Next, $\mathrm{cov}(Z_k, \mathbb{Y}) = (I_T \otimes S^{-1})(e_k^\top \Sigma_Z \gamma)$, where $e_k \in \mathbb{R}^p$ is a vector with the kth element being 1 and others being 0. Let $\Sigma_{\mathbb{Y}} = I_T \otimes \Sigma_Y$. Consequently, we have $\Sigma_{zy} = I_T \otimes S^{-1}$ and $\Delta_{1yz} = \Delta_{2yz} = c_{yz}^a \exp(-c_{yz}^b NT^2\kappa_{1j}\sigma_{yz}^{-1}\mu_{1x}\delta^2)$, where $\sigma_{yz} = \mathrm{tr}(\mathbb{M}_j^\top \Sigma_{\mathbb{Y}} \mathbb{M}_j) + \mathrm{tr}(\Sigma_{zy}\mathbb{M}_j\Sigma_{zy}\mathbb{M}_j) = T(W_{\cdot j}^\top \Sigma_Y e_j)^2 + T(e_j^\top S^{-1} W_{\cdot j})^2 \leq T(W_{\cdot j}^\top W_{\cdot j})(e_j^\top \Sigma_Y^2 e_j) + T(W_{\cdot j}^\top W_{\cdot j})\{e_j^\top S^{-1}(S^{-1})^\top e_j\}$. It can be further derived $T(W_{\cdot j}^\top W_{\cdot j})(e_j^\top \Sigma_Y^2 e_j) \leq \kappa_{1j} TN\lambda_{\max}^2(\Sigma_Y)$ and $e_j^\top S^{-1}(S^{-1})^\top e_j = c_{\gamma e}^{-1} e_j^\top \Sigma_Y e_j \leq c_{\gamma e}^{-1}\lambda_{\max}(\Sigma_Y)$. Therefore, $\sigma_{yz} \leq \kappa_{1j}TN\{\lambda_{\max}^2(\Sigma_Y) + c_{\gamma e}^{-1}\lambda_{\max}(\Sigma_Y)\}$ In addition, we have $\sigma_{Y,jj} \geq \lambda_{\min}(\Sigma_Y)$. Consequently, it can be derived $\Delta_{1yz} \leq c_{yz}^{a*}\exp(-c_{yz}^{b*} N^{1-2\tau}T\delta_1^2)$ by condition (C3), where c_{yz}^{a*} and c_{yz}^{b*} are finite constants. Lastly, note by condition (C3) we have $\tau < 1/2$ and $T = O((N^{2(1-\zeta)}\log N)^\xi)$ with $\xi > 1$, we have $\sum_{j=1}^N P(|\widehat{R}_{2j} - R_{2j}| > \delta_1/4) \to 0$. This completes the proof.

Appendix A.3: Proof of Theorem 1

In this proof, we separate the proof into three steps. In the first step, we show that the total amount of signal $\sum_{j=1}^N R_j^2$ is of $O(N^\tau)$. Second, we prove the set $\mathcal{M}$ can be covered by $\widehat{\mathcal{M}} = \{1 \leq j \leq N : \widehat{R}_j^2 > c_{\min}/2\}$. Lastly, we show that the size of $\widehat{\mathcal{M}}$ can be bounded by $m_{\max}$, which takes order of $O(N^{1+\tau-\zeta})$.

Step 1. We first prove that $\sum_{j=1}^N R_j^2 \leq C_r = O(N^\tau)$. It suffices to show the upper bound of each term in (2.7). Specifically, we reconsider that $R_{1j} = (\kappa_{1j}\sigma_{Y,jj}c_y)^{-1}\kappa_{2j}^2$, $R_{2j} = (N\kappa_{1j}\sigma_{Y,jj})^{-1}\kappa_{3j}^2 c_z$, $R_{3j} = (N\kappa_{1j}\sigma_{Y,jj}c_y^2)^{-1}(\kappa_{2j}^2 c_s^2 c_z)$, and $R_{4j} = -2(N\kappa_{1j}\sigma_{Y,jj}c_y)^{-1}$ $\kappa_{2j}\kappa_{3j}c_s c_z$, and we have $R_j^2 = R_{1j} + R_{2j} + R_{3j} + R_{4j}$. We next investigate each of them separately. By Cauchy inequality we have

$$c_y \geq \lambda_{\min}(\Sigma_Y), \sigma_{Y,jj} \geq \lambda_{\min}(\Sigma_Y) \tag{A.19}$$

$$|\kappa_{2j}| \leq (e_j^\top \Sigma_Y e_j)^{1/2}(W_{\cdot j}^\top \Sigma_Y W_{\cdot j})^{1/2} \leq \sigma_{Y,jj}^{1/2}\kappa_{1j}^{1/2}\lambda_{\max}^{1/2}(\Sigma_Y), \tag{A.20}$$

$$c_s \leq N\lambda_{\max}^{1/2}\{S^{-1}(S^{-1})^\top\} = N\lambda_{\max}^{1/2}(\Sigma_Y)/c_{\gamma e}, \tag{A.21}$$

$$|\kappa_{3j}| \leq [e_j^\top \{S^{-1}(S^{-1})^\top\}e_j]^{1/2}(W_{\cdot j}^\top W_{\cdot j})^{1/2} = \sigma_{Y,jj}^{1/2}\kappa_{1j}^{1/2}/c_{\gamma e}^{1/2} \tag{A.22}$$

It can be shown that $\max\{|R_{1j}|, |R_{2j}|, |R_{3j}|, |R_{4j}|\} \leq c_r\lambda_{\max}(\Sigma_Y)/N$, where c_r is a finite positive constant. For simplicity, we only verify R_{1j} for illustration propose. It can be derived $|R_{1j}| \leq (e_j^\top\Sigma_Y e_j)(W_{\cdot j}^\top \Sigma_Y W_{\cdot j})/(\kappa_{1j}\sigma_{Y,jj}c_y) \leq \lambda_{\max}(\Sigma_Y)/\{N\lambda_{\min}(\Sigma_Y)\}$ by (A.19) and (A.20). Consequently, by condition (C2), we have $\sum_j R_j^2 \leq C_r$, where $C_r = O(N^\tau)$.

Step 2. Recall $c_{\min} = \min_{j\in\mathcal{M}} R_j^2$ and $\mathcal{M} \subset \{j : R_j^2 \geq c_{\min}\}$. Define $\widehat{\mathcal{M}} = \{j : \widehat{R}_j^2 \geq c_{\min}/2\}$. In this step, we show that $\widehat{\mathcal{M}}$ should uniformly cover $\mathcal{M}$ with probability tending to 1. Otherwise, there must exist at least one $j^* \in \mathcal{M}$ not included in $\widehat{\mathcal{M}}$. By the definition, we know $\widehat{R}_{j^*}^2 < 2^{-1}c_{\min}$. In the meanwhile, if $j^* \in \mathcal{M}$, we should have $R_{j^*}^2 \geq c_{\min}$. This implies that $|\widehat{R}_j^2 - R_j^2| > 2^{-1}c_{\min}$. As a result, if $\mathcal{M} \not\subset \widehat{\mathcal{M}}$, it then could be concluded $\max_i |\widehat{R}_j^2 - R_j^2| > 2^{-1}c_{\min}$. We then have $P(\mathcal{M} \not\subset \widehat{\mathcal{M}}) \leq P(\max_i |\widehat{R}_j^2 - R_j^2| > c_{\min}/2)$. By condition (C2), we have $c_{\min} \geq c$ asymptotically, where $c = N^{\zeta-1}$. Then the desired results can be obtained by the conclusion of Proposition 1.

Step 3. Lastly, we verify that the size of $\widehat{\mathcal{M}}$ can be uniformly bounded. By the first step, we have $\sum_{j=1}^N R_j^2 \leq C_r = O(N^\tau)$. Define $\mathcal{M}_s = \{j : R_j^2 > c_{\min}/4\}$. It can be obtained $C_r \geq \sum_{j\in\mathcal{M}_s} R_j^2 \geq |\mathcal{M}_s|c_{\min}/4$. Then we have $|\mathcal{M}_s| \leq 4C_r/c_{\min} \overset{\text{def}}{=} m_{\max}$. By condition (C3) and the result in Step 1, it can be concluded that $m_{\max} = O(N^{1+\tau-\zeta})$. If $|\widehat{\mathcal{M}}| > |\mathcal{M}_s|$, we must have $\widehat{\mathcal{M}} \not\subset \mathcal{M}_s$. This implies there exists at least one $j \in \widehat{\mathcal{M}}$ with $\widehat{R}_j^2 \geq c_{\min}/2$ but $j \notin \mathcal{M}_s$ with $R_j^2 \leq c_{\min}/4$. Consequently we have

$\max_j |\hat{R}_j^2 - R_j^2| \geq 4^{-1}c_{\min}$. It can be concluded $P(|\widehat{\mathcal{M}}| > m_{\max}) \leq P(\max_j |\widehat{R}_j^2 - R_j^2| \geq 4^{-1}c_{\min})$. By Proposition 1, we have $P(\max_i |\widehat{R}_j^2 - R_j^2| \geq 4^{-1}c_{\min}) \to 0$. Immediately we know $P(|\widehat{\mathcal{M}}| \leq m_{\max}) \to 1$ as $N \to \infty$.

Appendix A.4: Proof of Proposition 2

Note the form of R_j^2 is given in (2.7) and recall $R_j^2 = R_{1j} + R_{2j} + R_{3j} + R_{4j}$. It can be derived $R_{2j} + R_{3j} + R_{4j} = c_z N^{-1}(c_s \kappa_{2j}/c_y - \kappa_{3j})^2$. Therefore, we have $R_j^2 \geq R_{1j}$. It then suffices to derive the order of R_{1j}. Before we go into details, we define some notations. For two arbitrary matrices $M_1 = (m_{1,ij}) \in \mathbb{R}^{N_1 \times N_2}$ and $M_2 = (m_{2,ij}) \in \mathbb{R}^{N_1 \times N_2}$, define $M_1 \succcurlyeq M_2$ if $m_{1,ij} \geq m_{2,ij}$ for $1 \leq i \leq N_1$ and $1 \leq j \leq N_2$. Similarly, we could define the notation "$\preccurlyeq$". In what follows, we first derive the lower bound of R_{1j} for $j \in \mathcal{M}$ as $R_{1j} \geq (\kappa_{1j}\sigma_{Y,jj}c_y)^{-1}\kappa_{5j}^2$, where

$$\kappa_{5j} = e_j^\top WD(I - WD)^{-1}(I - DW^\top)^{-1}DW^\top W_{\cdot j}. \tag{A.23}$$

Then we discuss the order of the lower bound.

Step 1. ($R_{1j} \geq (\kappa_{1j}\sigma_{Y,jj}c_y)^{-1}\kappa_{5j}^2$) First, we investigate the order of κ_{2j}. By performing a Taylor's expansion on Σ_Y, we have $\Sigma_Y = I + (I - WD)^{-1}WD + (I - DW^\top)^{-1}DW^\top + WD(I-WD)^{-1}(I-DW^\top)^{-1}DW^\top$. One can easily verify that $\kappa_{2j} = e_j^\top \Sigma_Y W_{\cdot j} = e_j^\top WD(I-WD)^{-1}W_{\cdot j} + e_j^\top DW^\top(I-DW^\top)^{-1}W_{\cdot j} + e_j^\top WD(I-WD)^{-1}(I-DW^\top)^{-1}DW^\top W_{\cdot j} = \kappa_{3j} + \kappa_{4j} + \kappa_{5j}$ due to $e_j^\top W_{\cdot j} = 0$, where $\kappa_{4j} = e_j^\top DW^\top(I - DW^\top)^{-1}W_{\cdot j}$ and κ_{5j} defined in (A.23). Due to that $d_{\min} > 0$, we have $\kappa_{3j} > 0$, $\kappa_{4j} > 0$, and $\kappa_{5j} > 0$. Therefore, we have $R_{1j} = (\kappa_{1j}\sigma_{Y,jj}c_y)^{-1}\kappa_{2j}^2 \geq (\kappa_{1j}\sigma_{Y,jj}c_y)^{-1}\kappa_{5j}^2$. It then suffices to derive the order of κ_{5j}.

Step 2. (The order of $(\kappa_{1j}\sigma_{Y,jj}c_y)^{-1}\kappa_{5j}^2$) Without loss of generality, we assume the first s elements of d are nonzero. Assume $c_{\gamma e} = 1$ for simplification in the following. Note κ_{5j} can be written as $\kappa_{5j} = (W_{j\cdot}^\top D)(\Sigma_Y)(DW^\top W_{\cdot j})$, where $W_{j\cdot}$ denotes the jth row vector of W. It can be easily verified that $W_{j\cdot}^\top D \succcurlyeq \mathbf{0}$ and $DW^\top W_{\cdot j} \succcurlyeq \mathbf{0}$. We next prove that $\Sigma_Y \succcurlyeq \mathbf{0}$. By applying Taylor's expansion on Σ_Y, we have $\Sigma_Y = \{\sum_{k=0}^\infty (WD)^k\}\{\sum_{k=0}^\infty (DW^\top)^k\}$. It can be noted under the assumption of Proposition 2 that $d_{\min} > 0$, we will have all the elements in Σ_Y to satisfy $(WD)^{k_1}(DW^\top)^{k_2} \succcurlyeq \mathbf{0}$. Then it can be shown the elementwise lower bound of $W_{j\cdot}^\top D$ and $DW^\top W_{\cdot j}$ are $W_{j\cdot}^\top D \succcurlyeq d_{\min} W_{j\cdot}^\top \widetilde{I}_s$, and $DW^\top W_{\cdot j} \succcurlyeq c_w^* D\mathbf{1} \succcurlyeq c_w^* d_{\min} \widetilde{I}_s \mathbf{1}_N$, where $c_w^* = \min_{j\in\mathcal{M}}(W_{\cdot j}^\top W_{\cdot j})$, $d_{\min} = \min_{j\in\mathcal{M}} d_j$, and $\widetilde{I}_s = \mathrm{diag}(\mathbf{1}_s, \mathbf{0}_{N-s}) \in \mathbb{R}^{N\times N}$. Consequently, we have

$$\begin{aligned}\kappa_{5j} &\ge c_w^* d_{\min}^2 (W_{j\cdot}^\top \widetilde{I}_s \Sigma_Y \widetilde{I}_s \mathbf{1}_N) \ge c_w^* d_{\min}^2 (W_{j\cdot}^\top \widetilde{I}_s \Sigma_Y \widetilde{I}_s W_{j\cdot})\\ &\ge c_w^* d_{\min}^2 (W_{j\cdot}^\top \widetilde{I}_s \mathrm{diag}(\Sigma_Y) \widetilde{I}_s W_{j\cdot}) \ge c_w^* c_w^2 d_{\min}^2 \min_{j\in\mathcal{M}} \sigma_{Y,jj},\end{aligned}$$

where the second inequality is due to $\mathbf{1}_N \succcurlyeq W_{j\cdot}$ and the last one is because $W_{j\cdot}^\top \widetilde{I}_s W_{j\cdot} \ge c_w^2$ by condition (2.12). For $j \in \mathcal{M}$, we have $c_1 N^\zeta \le \min\{c_w^*, \kappa_{1j}\} \le \max\{c_w^*, \kappa_{1j}\} \le c_2 N^\zeta$ by (2.10). Moreover, we have $c_3 N^{-1}\mathrm{tr}(\Sigma_Y) \le \min_{j\in\mathcal{M}} \sigma_{Y,jj} \le \max_{j\in\mathcal{M}} \sigma_{Y,jj} \le c_4 N^{-1}\mathrm{tr}(\Sigma_Y)$ by (2.11). Consequently, we have $(\kappa_{1j}\sigma_{Y,jj}c_y)^{-1}\kappa_{5j}^2 \ge c_1^2 c_2^{-1} c_4 c_3^{-2} c_w^4 d_{\min}^4 N^{\zeta-1}$. Consequently, the desired results can be obtained.

Appendix A.5: Matrix Forms and Notations

Denote $M_{\cdot j}$ to be the jth column vector of an arbitrary matrix M. The form of Σ_2 is given by

$$\Sigma_2 = \begin{pmatrix} \Sigma_{2d} & \Sigma_{2d\gamma} \\ \Sigma_{2d\gamma}^\top & \Sigma_{2\gamma} \end{pmatrix}, \tag{A.24}$$

$\Sigma_{2d} = (\Sigma_{2d,j_1j_2}) \in \mathbb{R}^{m\times m}$, $\Sigma_{2d\gamma} = (\Sigma_{2d_j\gamma} : 1 \le j \le m) \in \mathbb{R}^{m\times p}$ with

$$\Sigma_{2d,j_1j_2} = \lim_{N\to\infty} \{N^{-1}\delta_{j_1}\delta_{j_2} + \sigma_e^{-2} N^{-1} W_{\cdot j_1}^\top W_{\cdot j_2}(e_{j_1}^\top \Sigma_Y e_{j_2})\} \tag{A.25}$$

$$\Sigma_{2d_j\gamma} = \mathbf{0}, \quad \Sigma_{2\gamma} = \sigma_e^{-2}\Sigma_Z. \tag{A.26}$$

where $\delta_j = e_j^\top S_{\mathcal{M}}^{-1} W_{\cdot j}$. The form of Σ_1 is given as

$$\Sigma_1 = \Sigma_2 + \Delta\Sigma, \text{ where } \Delta\Sigma = \begin{pmatrix} \widetilde{\Delta}_d & \mathbf{0}_{m,p} \\ \mathbf{0}_{p,m} & \mathbf{0}_{p,p} \end{pmatrix}, \tag{A.27}$$

where $\mathbf{0}_{n_1,n_2}$ denotes a $n_1 \times n_2$ zero matrix. Here $\widetilde{\Delta}_d = (\Delta_{d,j_1j_2})$ and $\Delta_{d,j_1j_2} = \lim_{N\to\infty}\{N^{-1}\mathrm{tr}\{\mathrm{diag}(W_{\cdot j_1} e_{j_1}^\top S_{\mathcal{M}}^{-1})\mathrm{diag}(W_{\cdot j_2} e_{j_2}^\top S_{\mathcal{M}}^{-1})\}(\kappa_4 - 3\sigma_e^4)/\sigma_e^4\}$, where $\kappa_4 = E\varepsilon_{it}^4$.

Appendix A.6: Proof of Theorem 2

The proof is separated into the following two steps. In the first step, we prove that $\widehat{\theta}_{\mathcal{M}}$ is consistent with the rate $\alpha_{NT} = \sqrt{(NT)^{-1/2} m^{1/2}}$. In the second step, for each parameter $\widehat{d}_j$ ($j \in \mathcal{M}$) and $\widehat{\gamma}$, we show that they are asymptotic normal.

Step 1. To establish the consistency result, we follow Fan and Li (2001) to prove that for $\epsilon > 0$, there exists a constant $C > 0$ such that

$$\lim_{\min\{N,T\}\to\infty} P\Big\{\sup_{\|u\|=C} \ell(\theta_{\mathcal{M}} + \alpha_{NT} u) < \ell(\theta_{\mathcal{M}})\Big\} \geq 1 - \epsilon. \tag{A.28}$$

It is implied by (A.28) with probability at least $1-\epsilon$, there exists a local optimizer $\widehat{\theta}_{\mathcal{M}}$ in the ball $\{\theta_{\mathcal{M}} + C\alpha_{NT} u : \|u\| \leq 1\}$. Consequently, we will have $\|\widehat{\theta}_{\mathcal{M}} - \theta_{\mathcal{M}}\| = O_p(\alpha_{NT})$. Let $\dot{\ell}(\theta_{\mathcal{M}}) = \partial \ell(\theta_{\mathcal{M}})/\partial\theta_{\mathcal{M}} \in \mathbb{R}^m$ and $\ddot{\ell}(\theta_{\mathcal{M}}) = \partial^2\ell(\theta_{\mathcal{M}})/\partial\theta_{\mathcal{M}}\partial\theta_{\mathcal{M}}^\top \in \mathbb{R}^{m\times m}$ be the first and second order derivatives of $\ell(\theta_{\mathcal{M}})$ with respect to $\theta_{\mathcal{M}}$. We apply the Taylor's expansion to obtain that,

$$\sup_{\|u\|=C}\Big\{\ell(\theta_{\mathcal{M}} + C\alpha_{NT}u) - \ell(\theta_{\mathcal{M}})\Big\} = \sup_{\|u\|=C}\Big\{C\alpha_{NT}\dot{\ell}^\top(\theta_{\mathcal{M}})u + \frac{1}{2}C^2\alpha_{NT}^2 u^\top \ddot{\ell}(\theta_{\mathcal{M}})u + o_p(m)\Big\},$$
$$\leq C\|\alpha_{NT}\dot{\ell}(\theta_{\mathcal{M}})\| - 2^{-1}C^2 m\lambda_{\min}\{-(NT)^{-1}\ddot{\ell}(\theta_{\mathcal{M}})\} + o_p(m). \tag{A.29}$$

We then prove that (A.29) is asymptotically negative with probability 1.

Denote $\dot{\ell}_d(\theta_{\mathcal{M}}) = \partial \ell(\theta_{\mathcal{M}})/\partial d_{\mathcal{M}} \in \mathbb{R}^m$ and $\dot{\ell}_\gamma(\theta_{\mathcal{M}}) = \partial \ell(\theta_{\mathcal{M}})/\partial \gamma \in \mathbb{R}^p$. In addition, denote $\ddot{\ell}_d(\theta_{\mathcal{M}}) = (\ddot{\ell}_{d_{j_1} d_{j_2}}(\theta_{\mathcal{M}})) = \partial^2 \ell(\theta_{\mathcal{M}})/\partial d_{\mathcal{M}} \partial d_{\mathcal{M}}^\top \in \mathbb{R}^{m\times m}$, $\ddot{\ell}_{d\gamma}(\theta_{\mathcal{M}}) = (\ddot{\ell}_{d_j\gamma})^\top = \partial^2 \ell(\theta_{\mathcal{M}})/\partial d_{\mathcal{M}} \partial \gamma^\top \in \mathbb{R}^{m\times p}$, and $\ddot{\ell}_\gamma(\theta_{\mathcal{M}}) = \partial^2 \ell(\theta_{\mathcal{M}})/\partial \gamma \partial \gamma^\top \in \mathbb{R}^{p\times p}$. We then give the expressions of $\dot{\ell}(\theta_{\mathcal{M}})$ and $\ddot{\ell}(\theta_{\mathcal{M}})$ in the following as

$$\dot{\ell}_{d_j}(\theta_{\mathcal{M}}) = -T\delta_j + \sigma_e^{-2}\Delta_j, \tag{A.30}$$

$$\dot{\ell}_\gamma(\theta_{\mathcal{M}}) = \sigma_e^{-2}\sum_{t=1}^T Z_t^\top (SY_t - Z_t\gamma), \tag{A.31}$$

where $\delta_j = e_j^\top S^{-1} W_{\cdot j}$, $\Delta_j = \sum_{t=1}^T (SY_t - Z_t\gamma)^\top (W_{\cdot j} Y_{jt})$, and

$$\ddot{\ell}_{d_{j_1} d_{j_2}}(\theta_{\mathcal{M}}) = -T\delta_{j_1}\delta_{j_2} - \sigma_e^{-2}\sum_{t=1}^T (W_{\cdot j_1}^\top W_{\cdot j_2} Y_{j_1 t} Y_{j_2 t}), \tag{A.32}$$

$$\ddot{\ell}_{d_j\gamma}(\theta_{\mathcal{M}}) = -\sigma_e^{-2}\sum_{t=1}^T Z_t^\top W_{\cdot j} Y_{jt}, \quad \ddot{\ell}_\gamma(\theta_{\mathcal{M}}) = -\sigma_e^{-2}\sum_{t=1}^T Z_t^\top Z_t.$$

Next, we prove two important results: (1) $\alpha_{NT}\dot{\ell}_{d_j}(\theta_{\mathcal{M}}) = O_p(\sqrt{m})$ and $\alpha_{NT}\dot{\ell}_\gamma(\theta_{\mathcal{M}}) = O_p(\sqrt{m})$; (2) $P\{\| -(NT)^{-1}\ddot{\ell}(\theta_{\mathcal{M}}) - \Sigma_2\|_\infty > \epsilon_0\} \to 0$ for arbitrary $\epsilon_0 > 0$, where Σ_2 is given by (A.24). Next, we separate the proof of Step 1 into 3 parts in the following. In Step 1.1, we prove (1), in Step 1.2, we prove (2), and Step 1.3, we prove (3) $\lambda_{\min}(\Sigma_2) > \tau_0$, where $\tau_0 > 0$ is a constant. Then by applying Lemma 5 we have $\lambda_{\min}(-(NT)^{-1}\ddot{\ell}(\theta_{\mathcal{M}})) > \tau_0/2$. Consequently, by choosing C large enough, we could have (A.29) is negative with probability tending to 1. This completes the proof of Step 1.

Step 1.1. We firstly look at (A.30). Note that $E(\Delta_j) = T\mathrm{tr}(We_je_j^\top S^{-1}) = T\delta_j$. Therefore we have $E\{\dot{\ell}_{d_j}(\theta_{\mathcal{M}})\} = 0$. In addition, note that Z_t and $\mathcal{E}_t$ follow sub-Gaussian distribution and are independent over $1 \le t \le T$. Then we have $\mathrm{var}\{\alpha_{NT}\dot{\ell}_{d_j}(\theta_{\mathcal{M}})\} \le c\alpha_{NT}^2 T\sigma_e^2 \mathrm{tr}\{We_je_j^\top S^{-1}S^{\top -1}e_je_j^\top W^\top\} \le c_1 m N^{-1}(e_j^\top \Sigma_Y e_j)(W_{\cdot j}^\top W_{\cdot j})$ $\le c_1 m\lambda_{\max}(\Sigma_Y)(N^{-1}W_{\cdot j}^\top W_{\cdot j}) = O(m)$, which is due to $\max\{N^{-1}W_{\cdot j}^\top W_{\cdot j}, \lambda_{\max}(\Sigma_Y)\} = O(1)$ by (C5). Consequently we have $\alpha_{NT}\dot{\ell}_{d_j}(\theta_{\mathcal{M}}) = O_p(\sqrt{m})$. One could similarly verify that $\alpha_{NT}\dot{\ell}_\gamma(\theta_\mathcal{M}) = O_p(\sqrt{m})$, which is omitted here to save space.

STEP 1.2. It suffices to show for any $\epsilon_0 > 0$

$$P\Big\{\Big\| - (NT)^{-1}\ddot{\ell}_d(\theta_\mathcal{M}) - \Sigma_{2d}\Big\|_\infty > \epsilon_0\Big\} \to 0 \tag{A.33}$$

$$P\Big\{\Big\| - \alpha_{NT}^2\ddot{\ell}_{d\gamma}(\theta_\mathcal{M})\Big\|_\infty > \epsilon_0\Big\} \to 0 \tag{A.34}$$

and $-(NT)^{-1}\ddot{\ell}_\gamma(\theta_\mathcal{M}) \to_p \sigma_e^{-2}\Sigma_Z$. Due to the similarity, we only prove (A.33) in the following. It suffices to show that

$$P\Big\{\max_{j_1,j_2\in\mathcal{M}}\Big|\frac{\sum_t W_{\cdot j_1}^\top W_{\cdot j_2} Y_{j_1t}Y_{j_2t}}{NT\sigma_e^2} - \frac{W_{\cdot j_1}^\top W_{\cdot j_2}\Sigma_{Y,j_1j_2}}{N\sigma_e^2}\Big| > \epsilon_1\Big\} \to 0, \tag{A.35}$$

where $\epsilon_1 = \epsilon_0/3$. Denote $\kappa_{j_1j_2} = \lim_{N\to\infty} N^{-1}W_{\cdot j_1}^\top W_{\cdot j_2}$. By (C5), we have

$$\kappa_{j_1j_2} \le \lim_{N\to\infty} N^{-1}(W_{\cdot j_1}^\top W_{\cdot j_1})^{1/2}(W_{\cdot j_2}^\top W_{\cdot j_2})^{1/2} \le \lambda_{\max}(\mathbb{W}_\mathcal{M}) < \infty.$$

By Lemma 2, we have that

$$\begin{aligned} p_{d,j_1j_2} &\stackrel{\text{def}}{=} P\big\{\kappa_{j_1j_2}\sigma_e^{-2}|T^{-1}\sum_t Y_{j_1t}Y_{j_2t} - \Sigma_{Y,j_1j_2}| > \epsilon_1\big\} \\ &\le c_1\exp\{-c_2\sigma_{y,j_1j_2}^{-1}T\epsilon_1^2\} \le c_1\exp\{-c_2\lambda_{\max}^{-2}(\Sigma_{Y,\mathcal{M}})T\epsilon_1^2\}. \end{aligned}$$

for arbitrary positive ϵ_1, where $\sigma_{y,j_1j_2} = \Sigma_{Y,j_1j_2}\Sigma_{Y,j_2j_1} + \Sigma_{Y,j_1j_1}\Sigma_{Y,j_2j_2}$, c_1, c_2 are finite positive constants. By (C5), we have $\lambda_{\max}(\Sigma_{Y,\mathcal{M}}) \le \tau_2 < \infty$. Therefore we have $P\Big\{\Big\| - (NT)^{-1}\ddot{\ell}_d(\theta_\mathcal{M}) - \Sigma_{2d}\Big\|_\infty > \epsilon_1\Big\} \le \sum_{j_1,j_2} p_{d,j_1j_2} \le m^2 c_1\exp(-c_2\lambda_{\max}^{-2}(\Sigma_Y)T\epsilon_1^2)$ $\to 0$ due to $\log(m) = o(T)$.

STEP 1.3. Note that we have $\lambda_{\min}(\Sigma_Z) > 0$, then we only need to prove that $\lambda_{\min}(\Sigma_{2d}) > \tau_0 > 0$. It suffices to show that for any $\eta = (\eta_j)^\top \in \mathbb{R}^m$, we have

$$N^{-1}\sum_{j_1,j_2}\eta_{j_1}\delta_{j_1}\eta_{j_2}\delta_{j_2} + \sigma_e^{-2}N^{-1}\sum_{j_1,j_2}\eta_{j_1}\eta_{j_2}W_{\cdot j_1}^\top W_{\cdot j_2}\Sigma_{Y,j_1j_2} > \tau_0 > 0, \tag{A.36}$$

where τ_0 is a positive constant. One should note that for the first part of (A.36) we have $\sum_{j_1,j_2}\eta_{j_1}\delta_{j_1}\eta_{j_2}\delta_{j_2} = (\sum_j \eta_j\delta_j)^2 \geq 0$. Let $\mathbb{W} = W^\top W/N$ and $\mathbb{W}_\mathcal{M} \in \mathbb{R}^{m\times m}$ denote the submatrix of $\mathbb{W}$ with row and column indexes in $\mathcal{M}$. By Hiai and Lin (2017), we have $\prod_{j=1}^m \lambda_j(\mathbb{W}_\mathcal{M}\circ\Sigma_{Y,\mathcal{M}}) \geq \prod_{j=1}^m \lambda_j(\mathbb{W}_\mathcal{M}\Sigma_{Y,\mathcal{M}}) \geq \lambda_{\min}^m(\mathbb{W}_\mathcal{M})\lambda_{\min}^m(\Sigma_{Y,\mathcal{M}})$. Since we have $\min\{\lambda_{\min}(\mathbb{W}_\mathcal{M}), \lambda_{\min}(\Sigma_{Y,\mathcal{M}})\} \geq \tau_1 > 0$ and $\lambda_{\max}(\mathbb{W}_\mathcal{M}\circ\Sigma_{Y,\mathcal{M}}) \leq \max_{j_1,j_2}(W_{\cdot j_1}^\top W_{\cdot j_2})\max_{\|\eta\|=1}(\eta^\top|\Sigma_{Y,\mathcal{M}}|_e\eta) \leq \lambda_{\max}(\mathbb{W}_\mathcal{M})\lambda_{\max}(|\Sigma_{Y,\mathcal{M}}|_e) < \infty$ by Condition (3.2), we could conclude that $\lambda_{\min}(\mathbb{W}_\mathcal{M}\circ\Sigma_{Y,\mathcal{M}}) \geq \tau_0$ This proves (A.36).

Step 2. The asymptotic normality of $\widehat{\gamma}$ is trivial by noting that $(NT)^{-1/2}\Sigma_{2\gamma}^{-1}\dot{\ell}_\gamma(\theta_\mathcal{M}) \to_d N(0,\sigma_e^2\Sigma_Z^{-1})$ and then use the Slutsky's Theorem. In the following we prove the asymptotic normality for $\widehat{d}_i$. Let $\eta^{(i)} = e_i^\top\widehat{\Sigma}_{2d}^{-1} \in \mathbb{R}^m$, where $\widehat{\Sigma}_{2d} = -(NT)^{-1}\ddot{\ell}(\theta_\mathcal{M})$. It suffices to show $(NT)^{-1/2}\eta^{(i)\top}\dot{\ell}_d(\theta_\mathcal{M}) \to_d N(0,\sigma_i^2)$. For convenience, we omit the index i in $\eta^{(i)}$ and write $\eta^{(i)}$ as $\eta = (\eta_j)$ in the following. Note that $(NT)^{-1/2}\eta^{(i)\top}\dot{\ell}_d(\theta_\mathcal{M}) = (NT)^{-1/2}e_i^\top\Sigma_{2d}^{-1}\dot{\ell}_d(\theta_\mathcal{M}) + (NT)^{-1/2}e_i^\top(\widehat{\Sigma}_{2d}^{-1} - \Sigma_{2d}^{-1})\dot{\ell}_d(\theta_\mathcal{M})$. We separate the goals into two steps: (1) we prove $(NT)^{-1/2}e_i^\top(\widehat{\Sigma}_{2d}^{-1}-\Sigma_{2d}^{-1})\dot{\ell}_d(\theta_\mathcal{M}) = o_p(1)$; and (2) $(NT)^{-1/2}e_i^\top\Sigma_{2d}^{-1}\dot{\ell}_d(\theta_\mathcal{M}) \to_d N(0,\sigma_i^2)$.

Step 2.1. We could write $(NT)^{-1/2}e_i^\top(\widehat{\Sigma}_{2d}^{-1}-\Sigma_{2d}^{-1})\dot{\ell}_d(\theta_\mathcal{M}) = (NT)^{-1/2}e_i^\top\widehat{\Sigma}_{2d}^{-1}(\widehat{\Sigma}_{2d}-\Sigma_{2d})\Sigma_{2d}^{-1}\dot{\ell}_d(\theta_\mathcal{M})$. By the Cauchy's inequality, one could derive that

$$\begin{aligned}(NT)^{-1/2}\big|e_i^\top\widehat{\Sigma}_{2d}^{-1}(\widehat{\Sigma}_{2d}-\Sigma_{2d})\Sigma_{2d}^{-1}\dot{\ell}_d(\theta_\mathcal{M})\big| &\leq \sqrt{NT}\big|\lambda_1\big\{\widehat{\Sigma}_{2d}^{-1}(\widehat{\Sigma}_{2d}-\Sigma_{2d})\Sigma_{2d}^{-1}\big\}\big|\big\|\dot{\ell}_d(\theta_\mathcal{M})\big\| \\ &\leq (NT)^{-1/2}\big|\lambda_1\big(\widehat{\Sigma}_{2d}-\Sigma_{2d}\big)\big|\lambda_{\min}^{-1}(\widehat{\Sigma}_{2d})\lambda_{\min}^{-1}(\Sigma_{2d})\big\|\dot{\ell}(\theta_\mathcal{M})\big\|,\end{aligned}$$

where $\lambda_1(M)$ denotes the eigenvalue with largest absolute value. From the Step 1 we know that $(NT)^{-1/2}\|\dot{\ell}(\theta_\mathcal{M})\| = O_p(\sqrt{m})$. Next, by (A.14) we know that

$$P\Big(\sup_{\|r\|=1}\big|r^\top(\widehat{\Sigma}-\Sigma)r\big| > \epsilon/\sqrt{m}\Big) \leq c_1\exp(-c_2Tm^{-3}\epsilon^2 + c_3\log m + c_4 m).$$

Since we have $m = o(T^{\delta_1})$ with $0 \leq \delta_1 < 1/4$, it could be concluded $\big|\lambda_1\big(\widehat{\Sigma}_{2d}-\Sigma_{2d}\big)\big| = o_p(1/\sqrt{m})$. This leads to the result that $(NT)^{-1/2}e_i^\top(\widehat{\Sigma}_{2d}^{-1}-\Sigma_{2d}^{-1})\dot{\ell}_d(\theta_\mathcal{M}) = o_p(1)$.

STEP 2.2. One could write $\dot{\ell}_{d_j}(\theta_{\mathcal{M}})$ as

$$\begin{aligned}\dot{\ell}_{d_j}(\theta_{\mathcal{M}}) &= -T\delta_j + \sigma_e^{-2}\sum_{t=1}^T \mathcal{E}_t^\top\{We_je_j^\top S^{-1}(\mathcal{E}_t + Z_t\gamma)\}\\ &= -T\delta_j + \sigma_e^{-2}\sum_{t=1}^T \mathcal{E}_t^\top We_je_j^\top S^{-1}\mathcal{E}_t + \sigma_e^{-2}\sum_{t=1}^T \mathcal{E}_t^\top (We_je_j^\top S^{-1})Z_t\gamma,\\ &\overset{\text{def}}{=} -T\delta_j + \sum_t \mathcal{E}_t^\top M_j\mathcal{E}_t + \sum_t \mathcal{E}_t^\top U_j(Z_t\gamma). \end{aligned} \tag{A.37}$$

One could verify that $\lim_{\min(N,T)\to\infty} \text{var}\{(NT)^{-1/2}\dot{\ell}_d(\theta_{\mathcal{M}})\} \to \Sigma_1$, where Σ_1 is given by (A.27). It can be derived $\eta^\top\dot{\ell}_d(\theta_{\mathcal{M}}) = -T\sum_j \eta_j\delta_j + \sum_t\sum_j \mathcal{E}_t^\top M_j\eta_j\mathcal{E}_t + \sum_t\sum_j \mathcal{E}_tU_j\eta_j(Z_t\gamma)$. Let $M_\eta = \sum_j M_j\eta_j$, $U_\eta = \sum_j U_j\eta_j$, and $\mathbb{M}_\eta = |M_\eta|_e$, $\mathbb{U}_\eta = |U_\eta|_e$. Since $\{\mathcal{E}_t\}$ is independent over $1 \le t \le T$, then by the central limit theorem for the linear-quadratic forms (Zhu et al., 2018), it suffices to show

$$T^{-1}N^{-2}\text{tr}\big\{\mathbb{M}_\eta\mathbb{M}_\eta^\top\mathbb{M}_\eta\mathbb{M}_\eta^\top\big\} \to 0 \tag{A.38}$$

$$T^{-1}N^{-1}\lambda_{\max}(\mathbb{U}_\eta\mathbb{U}_\eta^\top) \to 0 \tag{A.39}$$

First we prove (A.38). It could be derived $\mathbb{M}_\eta \preccurlyeq \sum_j |\eta_j||We_je_j^\top S^{-1}|_e \overset{\text{def}}{=} \sum_j \mathbb{M}_{\eta j}$. It suffices to show $T^{-1}N^{-2}\sum_{j_1,j_2,j_3,j_4}|\eta_{j_1}\eta_{j_2}\eta_{j_3}\eta_{j_4}|\text{tr}\{\mathbb{M}_{\eta j_1}\mathbb{M}_{\eta j_2}^\top\mathbb{M}_{\eta j_3}\mathbb{M}_{\eta j_4}^\top\} \to 0$. Let $\eta_{j_1j_2j_3j_4} = \eta_{j_1}\eta_{j_2}\eta_{j_3}\eta_{j_4}$. It can be derived

$$\begin{aligned}&T^{-1}N^{-2}\sum_{j_1,j_2,j_3,j_4}|\eta_{j_1}\eta_{j_2}\eta_{j_3}\eta_{j_4}|\text{tr}\{\mathbb{M}_{\eta j_1}\mathbb{M}_{\eta j_2}^\top\mathbb{M}_{\eta j_3}\mathbb{M}_{\eta j_4}^\top\}\\ &\le \frac{1}{N^2T}\sum_{j_1,j_2,j_3,j_4}|\eta_{j_1j_2j_3j_4}|(W_{\cdot j_2}^\top W_{\cdot j_3})(W_{\cdot j_1}^\top W_{\cdot j_4})\{e_{j_1}^\top|S^{-1}|_e|S^{\top -1}|_e e_{j_2}\}\{e_{j_3}^\top|S^{-1}|_e|S^{\top-1}|_e e_{j_4}\}\\ &\le \frac{1}{N^2T}\sum_{j_1,j_2,j_3,j_4}|\eta_{j_1j_2j_3j_4}|\prod_{k=1}^4(W_{\cdot j_k}^\top W_{\cdot j_k})^{1/2}(e_{j_k}^\top|S^{-1}|_e|S^{\top-1}|_e e_{j_k})^{1/2}\\ &\le \sigma_Y^{-2}T^{-1}\lambda_{\max}^2(\mathbb{W}_{\mathcal{M}})\lambda_{\max}^2(\Sigma_{Y,\mathcal{M}}) \to 0 \end{aligned} \tag{A.40}$$

as $\min\{T,N\}\to\infty$, where the second inequality is due to the Cauchy inequality, and the last one is due to $\sum_{j_1,j_2,j_3,j_4}|\eta_{j_1}\eta_{j_2}\eta_{j_3}\eta_{j_4}| \le \sum_{j_1,j_2}|\eta_{j_2}\eta_{j_2}|\{\sum_{j_3,j_4}(\eta_{j_3}^2+\eta_{j_4}^2)/2\} =$

$c_\eta \sum_{j_1,j_2} |\eta_{j_2}\eta_{j_2}| \le c_\eta^2$, where c_η is a constant. Similar technique could be applied to prove (A.39) by noting that $(e_j^\top |S^{-1}|_e |S^{\top -1}|_e e_j)(W_{\cdot j}^\top W_{\cdot j}) = O(N)$.

APPENDIX B

In this appendix we provide some numerical procedures and results of the proposed screening and selection method.

Appendix B.1: Local Linear Approximation Algorithm

We first state the rough idea of the revised LLA algorithm. Generally, it breaks the estimation procedure into two steps. First, an initial Lasso type estimator is firstly obtained by imposing an L_1 penalty. Next, a local linear approximation is applied on the penalty as $p_\lambda(|d_j|) \approx |d_j| p'_\lambda(|d_j^{(0)}|)$, where $d_j^{(0)}$ denotes the estimator from the initial Lasso estimator. Consequently, the previous estimator is plugged in to continue estimation, which essentially leads to a weighted L_1 optimization problem. Here we borrow the idea of the LLA algorithm and illustrate the algorithm for the network data in the following.

Since the estimation of (3.3) does not take a closed form, the classical LARS algorithm (Efron et al., 2004) cannot be directly applied. Alternatively, we take the approach of the coordinate descent estimation (Breheny and Huang, 2011). That is, we optimize the objective function with respect to each parameter (i.e., d_j) at once and repeat the procedure sequentially. In each step, the second order approximation is applied to the quasi likelihood and then the objective function is analytically optimized.

For the jth parameter d_j, we introduce the notation $\theta_{\mathcal{M}}^{(-j)}$ as the remaining vector after d_j ($j \in \mathcal{M}$) is deleted in $\theta_{\mathcal{M}}$. Recall that $\ell^{(j)}(x) = \ell(x, \theta_{\mathcal{M}}^{(-j)})$ is a function of $\ell(\theta)$ at $d_j = x$ given the other parameters $\theta_{\mathcal{M}}^{(-j)}$ fixed, $\dot{\ell}^{(j)}(\cdot)$ and $\ddot{\ell}^{(j)}(\cdot)$ are the first and second derivative function of $\ell^{(j)}(\cdot)$. It can be derived

$$v_{1j} \stackrel{\text{def}}{=} \dot{\ell}^{(j)}(d_j) = -Te_j^\top S_{\mathcal{M}}^{-1} W_{\cdot j} + NT\delta_{1j},$$

$$v_{2j} \stackrel{\text{def}}{=} \ddot{\ell}^{(j)}(d_j) = T(e_j^\top S_{\mathcal{M}}^{-1} W_{\cdot j})^2 + \widehat{\sigma}_e^{-2}(W_{\cdot j}^\top W_{\cdot j}) \sum_{t=1}^{T} Y_{jt}^2 - 2NT(\delta_{1j})^2, \tag{B.1}$$

where $\delta_{1j} = (NT)^{-1}\widehat{\sigma}_e^{-2} \sum_{t=1}^{T}(S_{\mathcal{M}} Y_t - Z_t\gamma)^\top (W_{\cdot j} Y_{jt})$ and $\widehat{\sigma}_e^2 = (NT)^{-1} \sum_t (S_{\mathcal{M}} Y_t - Z_t\gamma)^\top (S_{\mathcal{M}} Y_t - Z_t\gamma)$. Given the mth estimator $\hat{d}_j^{(m)}$ for $j \in \mathcal{M}$, we could approximate the quasi log-likelihood function with respect to d_j at $\hat{d}_j^{(m)}$ by omitting some constatns as

$$\ell(\theta_{\mathcal{M}}|\theta_{\mathcal{M}}^{(-j)}) \approx \dot{\ell}^{(j)}(\hat{d}_j^{(m)}) + v_{1j}^{(m)}\Big(d_j - \hat{d}_j^{(m)}\Big) - 2^{-1} v_{2j}^{(m)} \Big(d_j - \hat{d}_j^{(m)}\Big)^2,$$
$$\approx -2^{-1} v_{2j}^{(m)} \Big(d_j - (v_{2j}^{(m)})^{-1} v_{1j}^{(m)} - \hat{d}_j^{(m)}\Big)^2,$$

where $v_{1j}^{(m)} = \dot{\ell}^{(j)}(\hat{d}_j^{(m)})$, and $v_{2j}^{(m)} = \ddot{\ell}^{(j)}(\hat{d}_j^{(m)})$. In addition, let $z_j^{(m)} = (v_{2j}^{(m)})^{-1} v_{1j}^{(m)} + \hat{d}_j^{(m)}$. The approximated objective function in the jth dimension takes the form

$$Q_a(d_j) = v_{2j}^{(m)}(d_j - z_j^{(m)})^2 + w_j^{(m)} \lambda |d_j|, \tag{B.2}$$

where $w_j^{(m)} = p'_\lambda(|\hat{d}_j^{(m)}|)$ is the weighted parameter. As a result, (B.2) takes an L_1 penalty form, which can be optimized and the closed form solution can be obtained. However, note in the approximated objective function the quadratic form $(d - z_j^{(m)})^2$ is weighted by the scaling value $v_{2j}^{(m)}$, which varies across different nodes. This could result in a unstable and discontinuous solution of the penalty function (Breheny and Huang, 2011). Moreover, it loses the consistent interpretation of penalty parameters. To solve this issue, we follow Breheny and Huang (2011) to adopt an adaptive rescaling technique by using a scaling parameter, which transforms the objective function in

(B.2) to the following one,

$$Q_a^*(d_j) = (d_j - z_j^{(m)})^2 + w_j^{(m)}|d_j|. \tag{B.3}$$

This is equivalent to solve a univariate Lasso problem and the closed form solution can be obtained as $\hat{d}_j^{(m+1)} = \text{sgn}(z_j^{(m)})(|z_j^{(m)}| - w_j^{(m)})_+$, where $\text{sgn}(\cdot)$ denotes the sign function and $(|z_j^{(m)}| - w_j^{(m)})_+ = \max(|z_j^{(m)}| - w_j^{(m)}, 0)$. The estimation procedure is summarized in Algorithm 1.

Remark. It should be noted that in the first step, solving (B.3) essentially yields the Lasso estimator. To avoid eliminating portal nodes at the beginning, it is recommended that the tuning parameter $\lambda^{(0)}$ should be sufficiently small. We follow the advice of Wang et al. (2013) to set $\lambda^{(0)} = \lambda\eta$ with a small $\eta = 1/\log(NT)$.

Appendix B.2: Simulation of the QMLE Estimation and Inference

In this section, we conduct the simulation experiment to verify the model inference result. We set the first $n_s = 10$ nodes to be the portal nodes. Next, we use the three examples in Section 4.1 to construct the network structure among the non-portal nodes. The other settings are the same with the simulation study in Section 4.1. The experiment is replicated for 100 times. In each replication, $\mathcal{M}$ is constructed by the all the portal nodes, and other 5 non-portal nodes with highest nodal in-degrees.

To evaluate the estimation performance, we calculate the average RMSE for the estimated parameters, i.e., $\text{RMSE}_d = \sum_{r=1}^{100}\{|\mathcal{M}|^{-1}\sum_{j\in\mathcal{M}}(\widehat{d}_j^{(r)} - d_j)^2/100\}^{1/2}$, $\text{RMSE}_\gamma = \sum_{r=1}^{100}\{p^{-1}\|\widehat{\gamma}^{(r)} - \gamma\|^2/100\}^{1/2}$, where $\widehat{d}_j^{(r)}$ and $\widehat{\gamma}^{(r)}$ is the QMLE estimation obtained at the rth replication. In addition, the 95% confidence interval is constructed for both d_j and γ_j as $\text{CI}_{d_j}^{(r)} = (\widehat{d}_j - z_{0.975}\widehat{\text{SE}}_{d_j}^{(r)}, \widehat{d}_j + z_{0.975}\widehat{\text{SE}}_{d_j}^{(r)})$, and $\text{CI}_{\gamma_j}^{(r)} = (\widehat{\gamma}_j - z_{0.975}\widehat{\text{SE}}_{\gamma_j}^{(r)}, \widehat{\gamma}_j + z_{0.975}\widehat{\text{SE}}_{\gamma_j}^{(r)})$, where $\widehat{\text{SE}}_{d_j}$ and $\widehat{\text{SE}}_{\gamma_j}$ are the root square of the diagonal elements of asymptotic covariance given in Theorem 2, and z_α is the αth quantile of the standard normal distribution. Then we report the average coverage probability (CP) for $d_{\mathcal{M}}$ and γ respectively as $\mathrm{CP}_d = \frac{1}{100|\mathcal{M}|}\sum_{j\in\mathcal{M}}\sum_{r=1}^{100} I(d_j \in \mathrm{CI}_{d_j}^{(r)})$ and $\mathrm{CP}_\gamma = \frac{1}{100p}\sum_{j=1}^{p}\sum_{r=1}^{100} I(\gamma_j \in \mathrm{CI}_{\gamma_j}^{(r)})$.

The results are summarized in Table 1. First, the RMSE values are decreased as N and T increase, which implies the consistency of the resulting QMLE estimator. Next, the coverage probabilities of both estimators are stable at 95% level. This corroborates with the asymptotic normality result given in Theorem 2.

References

Basu, S., Michailidis, G., et al. (2015), "Regularized estimation in sparse high-dimensional time series models," *The Annals of Statistics*, 43, 1535–1567.

Breheny, P. and Huang, J. (2011), "Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection," *The Annals of Applied Statistics*, 5, 232–253.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., et al. (2004), "Least angle regression," *Annals of Statistics*, 32, 407–499.

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.

Hiai, F. and Lin, M. (2017), "On an eigenvalue inequality involving the Hadamard product," *Linear Algebra and its Applications*, 515, 313–320.

Izenman, A. J. (2008), *Modern Multivariate Statistical Techniques*, Springer.

Saulis, L. and Statuleviveccius, V. (2012), *Limit Theorems for Large Deviations*, Springer Science & Business Media.

Vershynin, R. (2017), *High-Dimensional Probability: An Introduction with Applications*, Cambridge University Press.

Wang, L., Kim, Y., and Li, R. (2013), "Calibrating non-convex penalized regression in ultra-high dimension," *Annals of Statistics*, 41, 2505–2536.

Zhu, X., Huang, D., Pan, R., and Wang, H. (2018), "Multivariate Spatial Autoregression for Large Scale Social Networks," *Journal of Econometrics*, To appear.

Table 1: Simulation Results for the MEDIAN NETWORK with 100 Replications for three examples with $\delta = 1/2$ and $\delta = 1/4$. The RMSE_d, RMSE_γ are reported. In addition, the coverage probability (CP_d, CP_γ) and the network density (ND) are also reported in percentages.

	$\delta = 1/2$					$\delta = 1/4$				
(N, T)	RMSE_d	CP_d	RMSE_γ	CP_d	ND (%)	RMSE_d	CP_d	RMSE_γ	CP_d	ND (%)
Example 1 : Dyad Independence Network										
(100,50)	2.22	94.5	1.39	94.8	6.10	3.13	94.6	1.41	94.7	5.63
(200,100)	1.46	95.1	0.70	95.5	3.07	2.45	94.7	0.69	95.6	2.87
Example 2 : Stochastic Block Model										
(100,50)	1.86	94.8	1.42	94.5	3.28	1.64	94.5	1.38	95.4	2.28
(200,100)	1.11	95.3	0.71	95.2	1.41	1.03	95.0	0.70	95.2	1.03
Example 3 : Power-law Distribution Network										
(100,50)	2.25	94.5	1.40	95.4	4.78	2.92	95.2	1.42	94.7	4.20
(200,100)	1.47	95.3	0.72	94.9	2.43	2.11	95.0	0.69	95.2	2.18