# Supplementary Materials to "Mixed Graphical Model with Ordinal Data: Parameter Estimation and Statistical Inference"

Huijie Feng* Yang Ning†

## 1 Proofs of Main Results

### 1.1 Proof of Proposition 1

*Proof.* This proof is based upon the technical results in Fan et al. [2017].
When both variable $j$ and $k$ are ordinal, by definition and Theorem 1 in Fan et al. [2017] we have

$$\begin{aligned}\mathbb{P}(\sup_{1\le j,k\le d_1}|\widehat{R}_{jk}-\Sigma^*_{jk}|>t) &= \mathbb{P}(\sup_{1\le j,k\le d_1}|\sum_{\substack{1\le p\le N_j\\1\le q\le N_k}}(\widehat{R}^{(p,q)}_{jk}-\Sigma^*_{jk})|>N_jN_kt)\\ &\le \sum_{\substack{1\le p\le \max N_j\\1\le q\le \max N_k}}\mathbb{P}(\sup_{1\le j,k\le d_1}|\widehat{R}^{(p,q)}_{jk}-\Sigma^*_{jk}|>t)\\ &\le \max_{j,k}N_jN_kd_1^2\left\{2\exp(-\frac{nt^2}{8L_2^2})+4\exp(-\frac{nt^2\pi}{16L_1^2L_2^2})+4\exp(-\frac{M^2n}{2L_1^2})\right\},\end{aligned}$$

where $L_1, L_2$ are constants and $M$ is defined in Assumption 2. Therefore for some constant $C$ independent of $(n, d)$, we have

$$\mathbb{P}(\sup_{1\le j,k\le d_1}|\widehat{R}_{jk}-\Sigma^*_{jk}|\le C\sqrt{\frac{log(d)}{n}})\ge 1-d^{-1}.$$

Similarly, when variable $j$ is ordinal and variable $k$ is continuous, Theorem 2 in Fan et al. [2017] implies

$$\mathbb{P}(\sup_{1\le j\le d_1,d_1+1\le k\le d}|\widehat{R}_{jk}-\Sigma^*_{jk}|>t)\le 2\max_{j}N_jd_1d_2\left\{\exp(-\frac{nt^2}{8L_3^2})+\exp(-\frac{nt^2\pi}{12L_1^2L_3^2})+\exp(-\frac{M^2n}{2L_1^2})\right\},$$

where $L_3$ is a constant. Therefore for some constant $C$ independent of $(n, d)$, we have

$$\mathbb{P}(\sup_{1\le j\le d_1,d_1+1\le k\le d}|\widehat{R}_{jk}-\Sigma^*_{jk}|\le C\sqrt{\frac{log(d)}{n}})\ge 1-d^{-1}.$$

*Department of Statistics, Cornell University, Ithaca, NY 14850, USA; e-mail: hf279@cornell.edu.
†Department of Statistics, Cornell University, Ithaca, NY 14850, USA; e-mail: yn265@cornell.edu.

When both variable $j$ and $k$ are continuous, Liu et al. [2012] proves that

$$\mathbb{P}(\sup_{d_1+1\le j,k\le d} |\widehat{R}_{jk} - \Sigma^*_{jk}| \le 2.45\pi\sqrt{\frac{log(d_2)}{n}}) \ge 1 - d_2^{-1}.$$

Combining all three cases implies the desired result. $\square$

## 1.2 Proof of Theorem 1

We first introduce a Corollary proved by Fan et al. [2017]:

**Corollary 1.** Under the same condition in Proposition 1, it holds that

$$\mathbb{P}(||\widetilde{\boldsymbol{R}} - \widehat{\boldsymbol{R}}||_{\max} \le C\sqrt{\frac{log(d)}{n}}) \ge 1 - d^{-1},$$
$$\mathbb{P}(||\widetilde{\boldsymbol{R}} - \boldsymbol{\Sigma}^*||_{\max} \le 2C\sqrt{\frac{log(d)}{n}}) \ge 1 - d^{-1}.$$

*Proof of Theorem 1.* By definition, we have

$$\frac{1}{2}\widehat{\boldsymbol{\beta}}^T\widetilde{\boldsymbol{R}}\widehat{\boldsymbol{\beta}} - \boldsymbol{e}_k^T\widehat{\boldsymbol{\beta}} + \boldsymbol{\lambda}||\widehat{\boldsymbol{\beta}}||_1 \le \frac{1}{2}\boldsymbol{\beta}^{*T}\widetilde{\boldsymbol{R}}\boldsymbol{\beta}^* - \boldsymbol{e}_k^T\boldsymbol{\beta}^* + \lambda||\boldsymbol{\beta}^*||_1.$$

Rearranging terms and applying duality bound $\boldsymbol{x}^T\boldsymbol{y} \le ||\boldsymbol{x}||_\infty||\boldsymbol{y}||_1$ we have

$$\begin{aligned}\frac{1}{2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^T\widetilde{\boldsymbol{R}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) &\le (\boldsymbol{e}_k - \widetilde{\boldsymbol{R}}\boldsymbol{\beta}^*)^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) + \lambda(||\boldsymbol{\beta}^*||_1 - ||\widehat{\boldsymbol{\beta}}||_1) \\ &\le ||\widetilde{\boldsymbol{R}}\boldsymbol{\beta}^* - \boldsymbol{e}_k||_\infty||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*||_1 + \lambda(||\boldsymbol{\beta}^*||_1 - ||\widehat{\boldsymbol{\beta}}||_1)\end{aligned}.$$

Since

$$||\widetilde{\boldsymbol{R}}\boldsymbol{\beta}^* - \boldsymbol{e}_k||_\infty = ||(\widetilde{\boldsymbol{R}} - \boldsymbol{\Sigma}^*)\boldsymbol{\beta}^*||_\infty \le ||\boldsymbol{\beta}^*||_1||(\widetilde{\boldsymbol{R}} - \boldsymbol{\Sigma}^*)||_{max}$$

Corollary 1 together with Assumption 3 implies that $||\widetilde{\boldsymbol{R}}\boldsymbol{\beta}^* - \boldsymbol{e}_k||_\infty \le CM_n\sqrt{\frac{log(d)}{n}}$ with probability grater than $1 - d^{-1}$. Now choose $\lambda = 2CM_n\sqrt{\frac{log(d)}{n}}$, we get

$$\begin{aligned}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^T\widetilde{\boldsymbol{R}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) &\le \lambda(||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*||_1 + 2||\boldsymbol{\beta}^*||_1 - 2||\widehat{\boldsymbol{\beta}}||_1) \\ &\le \lambda(3||(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_S||_1 - ||(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_{S^C}||_1)\end{aligned}$$

This implies that

$$||(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_{S^C}||_1 \le 3||(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_S||_1,$$

which further implies

$$||(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)||_1 \le 4||(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_S||_1 \le 4s^{*1/2}||(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)||_2. \tag{1.1}$$

On the other hand, we have

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^T\widetilde{\boldsymbol{R}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \le 3\lambda||(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_S||_1 \le 3\lambda s^{*1/2}||(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)||_2 \tag{1.2}$$

By Assumption 4, we get

$$||(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)||_2 \leq 6C\tau M_n\sqrt{\frac{s^*log(d)}{n}}.$$

Together with (1.1), it holds that

$$||(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)||_1 \leq 24C\tau M_n s^*\sqrt{\frac{log(d)}{n}},$$

and

$$(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)^T\widetilde{\boldsymbol{R}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*) \leq 36C\tau M_n\frac{s^*log(d)}{n}.$$

□

### 1.3 Proof of Theorem 2

Before we prove the main theorem, we firstly introduce and prove the following lemmas:

**Lemma 1.** Under Assumption 1 and 2, when both $X_{ij}$ and $X_{ik}$ are ordinal, it holds that

$$\frac{N_jN_k\sqrt{n}(\widehat{R}_{jk}-\Sigma^*_{jk})}{(\triangledown M^T_{jk}(\boldsymbol{v}_{jk})\boldsymbol{\Psi}_{jk}\triangledown M_{jk}(\boldsymbol{v}_{jk}))^{1/2}} \rightarrow_d N(0,1),$$

where $\triangledown M_{jk}(\boldsymbol{v}_{jk})$ and $\boldsymbol{\Psi}_{jk}$ are defined in the proof.

When $X_{ij}$ is ordinal and $X_{ik}$ is continuous, it holds that

$$\frac{N_j\sqrt{n}(\widehat{R}_{jk}-\Sigma^*_{jk})}{(\triangledown M^T_{j}(\boldsymbol{v}_{j})\boldsymbol{\Psi}_{j}\triangledown M_{j}(\boldsymbol{v}_{j}))^{1/2}} \rightarrow_d N(0,1),$$

where $\triangledown M_{j}(\boldsymbol{v}_{j})$ and $\boldsymbol{\Psi}_{j}$ are defined in the proof.

*Proof.* **Case 1:** When both $X_{ij}$ and $X_{ik}$ are ordinal, let $\boldsymbol{v}_{jk} = (\boldsymbol{\tau}_{jk}, \boldsymbol{\Phi}_j, \boldsymbol{\Phi}_k)^T \in \mathbb{R}^{N_jN_k+N_j+N_k}$, where $\boldsymbol{\tau}_{jk} = (\tau^{(1,1)}_{jk}, \tau^{(1,2)}_{jk}, \dots, \tau^{(N_j,N_k)}_{jk})$ and $\boldsymbol{\Phi}_j = (\Phi(\triangle^{(1)}_j), \dots, \Phi(\triangle^{(N_j)}_j))$ $\forall j = 1, \dots, d_1$. Define the function

$$M_{jk}(t_{11}, t_{12}, \dots, t_{N_jN_k}, x_1, \dots, x_{N_j}, y_1, \dots, y_{N_k}) = \sum_{\substack{1\leq p\leq N_j\\ 1\leq q\leq N_k}} F^{-1}(t_{pq}, \Phi^{-1}(x_p), \Phi^{-1}(y_q)).$$

Consider the estimator of $\boldsymbol{v}_{jk}$

$$\widehat{\boldsymbol{v}}_{jk} = (\widehat{\boldsymbol{\tau}}_{jk}, \widehat{\boldsymbol{X}}_j, \widehat{\boldsymbol{X}}_k)^T,$$

where

$$\widehat{\boldsymbol{\tau}}_{jk} = (\widehat{\tau}^{(1,1)}_{jk}, \widehat{\tau}^{(1,2)}_{jk}, \dots, \widehat{\tau}^{(N_j,N_k)}_{jk}) \;\text{ and }\; \widehat{\boldsymbol{X}}_j = (1-\bar{X}^{(1)}_j, \dots, 1-\bar{X}^{(N_j)}_j)\ \forall j = 1, \dots, d_1.$$

For notation simplicity, we write

$$\phi^{(p,q)}_{jk}(\boldsymbol{X}_i) = \mathbb{E}\Big[\text{sign}\{(X^{(p)}_{ij}-X^{(p)}_{i'j})(X^{(q)}_{ik}-X^{(q)}_{i'k})\}|X_{ij}, X_{ik}\Big],$$

where $\boldsymbol{X}_i$ denotes the $i^{th}$ sample and $\boldsymbol{X}_{i'}$ is an independent copy $\boldsymbol{X}_i$.

By the property of Hájek projection of U-statistics, we have

$$\sqrt{n}(\widehat{\boldsymbol{v}}_{jk} - \boldsymbol{v}_{jk}) = \sqrt{n}\widetilde{\boldsymbol{U}}_n + o_{\mathbb{P}}(1),$$

where

$$\widetilde{\boldsymbol{U}}_n = \frac{1}{n}\sum_{i=1}^{n}(\underbrace{\begin{bmatrix} 2\phi_{jk}^{(1,1)}(\boldsymbol{X}_i) \\ 2\phi_{jk}^{(1,2)}(\boldsymbol{X}_i) \\ \vdots \\ 2\phi_{jk}^{(N_j,N_k)}(\boldsymbol{X}_i) \\ 1 - X_{ij}^{(1)} \\ \vdots \\ 1 - X_{ij}^{(N_j)} \\ 1 - X_{ik}^{(1)} \\ \vdots \\ 1 - X_{ik}^{(N_k)} \end{bmatrix}}_{\widehat{\boldsymbol{\theta}}_i^{(jk)}} - \underbrace{\begin{bmatrix} 2\tau_{jk}^{(1,1)} \\ 2\tau_{jk}^{(1,2)} \\ \vdots \\ 2\tau_{jk}^{(N_j,N_k)} \\ \Phi(\triangle_j^{(1)}) \\ \vdots \\ \Phi(\triangle_j^{(N_j)}) \\ \Phi(\triangle_k^{(1)}) \\ \vdots \\ \Phi(\triangle_j^{(N_k)}) \end{bmatrix}}_{\boldsymbol{\theta}^{(jk)}}).$$

Since the sign function and $X_j^{(p)}$ are bounded, by the central limit theorem we get

$$\sqrt{n}\widetilde{\boldsymbol{U}}_n \to_d N(0, \boldsymbol{\Psi}_{jk}),$$

where $\boldsymbol{\Psi}_{jk} = \mathbb{E}(\widehat{\boldsymbol{\theta}}_i^{(jk)} - \boldsymbol{\theta}^{(jk)})(\widehat{\boldsymbol{\theta}}_i^{(jk)} - \boldsymbol{\theta}^{(jk)})^T$. On the other hand, we have by definition

$$\sqrt{n}(\widehat{R}_{jk} - \Sigma_{jk}^*) = \frac{\sqrt{n}}{N_j N_k}(M_{jk}(\widehat{\boldsymbol{v}}_{jk}) - M_{jk}(\boldsymbol{v}_{jk})).$$

By Lemma 2 of Fan et al. [2017], $\frac{\partial M_{jk}}{\partial t_{pq}}$ is bounded and well defined at $\boldsymbol{v}_{jk}$. In addition, we can show that

$$\frac{\partial M_{jk}}{\partial x_p}(\boldsymbol{v}_{jk}) = \sum_{q=1}^{N_k}\left[\Phi(\triangle_k^{(q)}) - \Phi(\frac{\triangle_k^{(q)} - \Sigma_{jk}\triangle_j^{(p)}}{\sqrt{1-\Sigma_{jk}^2}})\right] \cdot \frac{\partial M_{jk}}{\partial t_{pq}}(\boldsymbol{v}_{jk}),$$

which is also bounded and well defined, and similarly for $\frac{\partial M_{jk}}{\partial y_q}(\boldsymbol{v}_{jk})$. This implies that the gradient of $M_{jk}$ at $\boldsymbol{v}_{jk}$, denoted by $\triangledown M_{jk}(\boldsymbol{v}_{jk})$, is bounded and well defined. This completes the proof by applying the delta method.

**Case 2:** When $X_{ij}$ is ordinal and $X_{ik}$ is continuous. Let $\boldsymbol{v}_j = (\boldsymbol{\tau}_j, \boldsymbol{\Phi}_j)^T \in \mathbb{R}^{2N_j}$, where $\boldsymbol{\tau}_j = (\tau_{jk}^{(1)}, \ldots, \tau_{jk}^{(N_j)})$.. Define

$$M_j(t_1, \ldots, t_{N_j}, x_1, \ldots, x_{N_j}) = \sum_{1 \le p \le N_j} F^{-1}(t_p, \Phi^{-1}(x_p)).$$

x Let

$$\widehat{\boldsymbol{v}}_j = (\widehat{\boldsymbol{\tau}}_j, \widehat{\boldsymbol{X}}_j)^T,$$

where

$$\widehat{\boldsymbol{\tau}}_j = (\widehat{\tau}_j^{(1)}, \dots, \widehat{\tau}_j^{(N_j)}) \text{ and } \widehat{\boldsymbol{X}}_j = (1-\bar{X}_j^{(1)}, \dots, 1-\bar{X}_j^{(N_j)}) \ \forall j = 1, \dots, d_1.$$

Similarly, we write

$$phi_{jk}^{(p)}(\boldsymbol{X}_i) = \mathbb{E}\Big[\mathrm{sign}\{(X_{ij}^{(p)} - X_{i'j}^{(p)})(X_{ik} - X_{i'k})\}|X_{ij}, X_{ik}\Big],$$

The Hájek projection of U-statistics gives

$$\sqrt{n}(\widehat{\boldsymbol{v}}_j - \boldsymbol{v}_j) = \sqrt{n}\widetilde{\boldsymbol{U}}_n + o_{\mathbb{P}}(1),$$

where

$$\widetilde{\boldsymbol{U}}_n = \frac{1}{n}\sum_{i=1}^n(\underbrace{\begin{bmatrix} 2\phi_{jk}^{(1)}(\boldsymbol{X}_i) \\ \vdots \\ 2\phi_{jk}^{(N_j)}(\boldsymbol{X}_i) \\ 1 - X_{ij}^{(1)} \\ \vdots \\ 1 - X_{ij}^{(N_j)} \end{bmatrix}}_{\widehat{\boldsymbol{\theta}}_i^{(j)}} - \underbrace{\begin{bmatrix} 2\tau_{jk}^{(1)} \\ \vdots \\ 2\tau_{jk}^{(N_j)} \\ \Phi(\triangle_j^{(1)}) \\ \vdots \\ \Phi(\triangle_j^{(N_j)}) \end{bmatrix}}_{\boldsymbol{\theta}^{(j)}}).$$

Since the sign function and $X_j^{(p)}$ are bounded, by the central limit theorem we get

$$\sqrt{n}\widetilde{\boldsymbol{U}}_n \to_d N(0, \boldsymbol{\Psi}_j),$$

where $\boldsymbol{\Psi}_j = \mathbb{E}(\widehat{\boldsymbol{\theta}}_i^{(j)} - \boldsymbol{\theta}^{(j)})(\widehat{\boldsymbol{\theta}}_i^{(j)} - \boldsymbol{\theta}^{(j)})^T$. Therefore, it holds that

$$\sqrt{n}(\widehat{R}_{jk} - \Sigma_{jk}^*) = \frac{\sqrt{n}}{N_j}(M_j(\widehat{\boldsymbol{v}}_j) - M_j(\boldsymbol{v}_j)).$$

Similar to the previous case, it can be shown that $\triangledown M_j(\boldsymbol{v}_j)$ is bounded and well defined. This completes the proof by applying the delta method.

□

**Lemma 2.** Under Assumption 1 - 4, it holds that

$$(\widehat{\gamma} - \boldsymbol{\gamma}^*)^T \widetilde{\boldsymbol{R}}_{22}(\widehat{\gamma} - \boldsymbol{\gamma}^*) = \mathcal{O}_{\mathbb{P}}(M_n^2 \frac{s^* log(d)}{n}).$$

*Proof of Lemma 2.* The Proof of Theorem 1 implies that

$$\begin{aligned}(\widehat{\gamma} - \boldsymbol{\gamma}^*)^T \widetilde{\boldsymbol{R}}_{22}(\widehat{\gamma} - \boldsymbol{\gamma}^*) &\leq 2(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\widetilde{\boldsymbol{R}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) + 2(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^2 \\ &\leq 2(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\widetilde{\boldsymbol{R}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) + 2||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*||_2^2 \\ &\lesssim 36C\tau M_n(1 + C\tau M_n)\frac{s^* log(d)}{n}.\end{aligned}$$

□

**Lemma 3.** Under Assumption 1 - 4, with probability greater than $1-d^{-1}$, we have $||\widehat{\boldsymbol{w}}-\boldsymbol{w}^*||_1 \lesssim (C'+M_n)s^*\sqrt{\frac{log(d)}{n}}$, and $(\widehat{\boldsymbol{w}}-\boldsymbol{w}^*)^T\widetilde{\boldsymbol{R}}_{22}(\widehat{\boldsymbol{w}}-\boldsymbol{w}^*) \lesssim (C'+M_n)^2 s^*\frac{log(d)}{n}$, where $C'>0$ is a constant.

*Proof of Lemma 3.* By definition, we know $||\widehat{\boldsymbol{w}}||_1 \leq ||\boldsymbol{w}^*||_1 = ||\boldsymbol{w}^*_{s'}||_1$. This together with $||(\widehat{\boldsymbol{w}}-\boldsymbol{w}^*))_{s'}||_1 \geq ||\boldsymbol{w}^*_{s'}||_1 - ||\widehat{\boldsymbol{w}}_{s'}||_1$ implies that $||(\widehat{\boldsymbol{w}}-\boldsymbol{w}^*))_{s'}||_1 \geq ||(\widehat{\boldsymbol{w}}-\boldsymbol{w}^*))_{s'^C}||_1$. In addition, we have $||(\widehat{\boldsymbol{w}}-\boldsymbol{w}^*))||_1 \leq 2||(\widehat{\boldsymbol{w}}-\boldsymbol{w}^*))_{s'}||_1$.
On the other hand, by Corollary 1 and Assumption 3 we have

$$\begin{aligned}
||\widetilde{\boldsymbol{R}}_{21}-\widetilde{\boldsymbol{R}}_{22}\boldsymbol{w}^*||_\infty &= ||\widetilde{\boldsymbol{R}}_{21}-\widetilde{\boldsymbol{R}}_{22}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}||_\infty \\
&= ||\widetilde{\boldsymbol{R}}_{21}-\boldsymbol{\Sigma}_{21}+(\boldsymbol{I}-\widetilde{\boldsymbol{R}}_{22}\boldsymbol{\Sigma}_{22}^{-1})\boldsymbol{\Sigma}_{21}||_\infty \\
&\leq ||\widetilde{\boldsymbol{R}}_{21}-\boldsymbol{\Sigma}_{21}||_\infty + ||(\boldsymbol{\Sigma}_{22}-\widetilde{\boldsymbol{R}}_{22})(\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})||_\infty \\
&\leq ||\widetilde{\boldsymbol{R}}-\boldsymbol{\Sigma}||_{\max} + ||\boldsymbol{w}^*||_1||\boldsymbol{R}_{22}-\widetilde{\boldsymbol{R}}_{22}||_{\max} \\
&\leq (2C+\tau M_n)\sqrt{\frac{log(d)}{n}}.
\end{aligned} \tag{1.3}$$

By choosing $\lambda = \tau M_n\sqrt{\frac{log(d)}{n}}$, (1.3) implies that $||\widetilde{\boldsymbol{R}}_{22}(\widehat{\boldsymbol{w}}-\boldsymbol{w}^*)||_\infty \leq ||\widetilde{\boldsymbol{R}}_{21}-\widetilde{\boldsymbol{R}}_{22}\boldsymbol{w}^*||_\infty + ||\widetilde{\boldsymbol{R}}_{21}-\widetilde{\boldsymbol{R}}_{22}\widehat{\boldsymbol{w}}||_\infty \leq 2(C+\tau M_n)\sqrt{\frac{log(d)}{n}}$. Therefore we get

$$\begin{aligned}
(\widehat{\boldsymbol{w}}-\boldsymbol{w}^*)^T\widetilde{\boldsymbol{R}}_{22}(\widehat{\boldsymbol{w}}-\boldsymbol{w}^*) &\leq ||\widehat{\boldsymbol{w}}-\boldsymbol{w}^*||_1||\widetilde{\boldsymbol{R}}_{22}(\widehat{\boldsymbol{w}}-\boldsymbol{w}^*)||_\infty \\
&\leq ||\widetilde{\boldsymbol{R}}_{22}(\widehat{\boldsymbol{w}}-\boldsymbol{w}^*)||_\infty||\widehat{\boldsymbol{w}}-\boldsymbol{w}^*||_1 \\
&\leq 2s'^{1/2}||\widehat{\boldsymbol{w}}-\boldsymbol{w}^*||_2||\widetilde{\boldsymbol{R}}_{22}(\widehat{\boldsymbol{w}}-\boldsymbol{w}^*)||_\infty \\
&\leq 4(C+\tau M_n)\sqrt{\frac{s^*log(d)}{n}}||\widehat{\boldsymbol{w}}-\boldsymbol{w}^*||_2.
\end{aligned} \tag{1.4}$$

This together with the Assumption 4 implies that

$$||\widehat{\boldsymbol{w}}-\boldsymbol{w}^*||_1 \leq 2||(\widehat{\boldsymbol{w}}-\boldsymbol{w}^*)_{S'}||_1 \leq 2s'||\widehat{\boldsymbol{w}}-\boldsymbol{w}^*||_2 \leq 8\tau(C+\tau M_n)s^*\sqrt{\frac{log(d)}{n}}.$$

In addition,

$$(\widehat{\boldsymbol{w}}-\boldsymbol{w}^*)^T\widetilde{\boldsymbol{R}}_{22}(\widehat{\boldsymbol{w}}-\boldsymbol{w}^*) \leq 16\tau(C+\tau M_n)^2\frac{s^*log(d)}{n}.$$

This completes the proof. □

**Lemma 4.** With probability greater than $1-d^{-1}$, we have $(\widehat{\boldsymbol{v}}-\boldsymbol{v}^*)^T\widetilde{\boldsymbol{R}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*) \lesssim (C'+M_n)M_n\frac{s^*log(d)}{n}$, where $C'>0$ is a constant.

*Proof.* By Cauchy-Schiwartz inequality and Cholesky decomposition, write $\widetilde{\boldsymbol{R}} = \boldsymbol{L}\boldsymbol{L}^T$, we have

$$\begin{aligned}
[(\widehat{\boldsymbol{v}}-\boldsymbol{v}^*)^T\widetilde{\boldsymbol{R}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)]^2 &= [(\widehat{\boldsymbol{v}}-\boldsymbol{v}^*)^T\boldsymbol{L}\boldsymbol{L}^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)]^2 \\
&\leq (\widehat{\boldsymbol{v}}-\boldsymbol{v}^*)^T\boldsymbol{L}\boldsymbol{L}^T(\widehat{\boldsymbol{v}}-\boldsymbol{v}^*)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)^T\boldsymbol{L}\boldsymbol{L}^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*). \\
&= (\widehat{\boldsymbol{v}}-\boldsymbol{v}^*)^T\widetilde{\boldsymbol{R}}(\widehat{\boldsymbol{v}}-\boldsymbol{v}^*)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)^T\widetilde{\boldsymbol{R}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)
\end{aligned}$$

Thus Theorem 1 and Lemma 3 imply the desired result. □

*Proof of Theorem 2.* By definition, we have

$$\begin{aligned}\sqrt{n}|\widehat{S}(\widehat{\boldsymbol\beta}_0)-S(\boldsymbol\beta^*)| &= \sqrt{n}|\widehat{\boldsymbol v}^T(\widehat{\boldsymbol R}\widehat{\boldsymbol\beta}_0-\boldsymbol e_k)-\boldsymbol v^{*T}(\widehat{\boldsymbol R}\boldsymbol\beta^*-\boldsymbol e_k)| \\ &\le \sqrt{n}|(\widehat{\boldsymbol v}-\boldsymbol v^*)^T(\widehat{\boldsymbol R}\widehat{\boldsymbol\beta}_0-\boldsymbol e_k)|+\sqrt{n}|\boldsymbol v^{*T}\widehat{\boldsymbol R}(\widehat{\boldsymbol\beta}_0-\boldsymbol\beta^*)| \\ &:= I_1+I_2.\end{aligned}$$

Here

$$\begin{aligned}|I_1| &\le \sqrt{n}|(\widehat{\boldsymbol v}-\boldsymbol v^*)^T(\widetilde{\boldsymbol R}\boldsymbol\beta^*-\boldsymbol e_k)|+\sqrt{n}|(\widehat{\boldsymbol v}-\boldsymbol v^*)^T(\widehat{\boldsymbol R}\widehat{\boldsymbol\beta}_0-\widetilde{\boldsymbol R}\boldsymbol\beta^*)| \\ &\le \underbrace{\sqrt{n}||\widehat{\boldsymbol v}-\boldsymbol v^*||_1||\widetilde{\boldsymbol R}\boldsymbol\beta^*-\boldsymbol e_k||_\infty}_{(1)}+\underbrace{\sqrt{n}|(\widehat{\boldsymbol v}-\boldsymbol v^*)^T\widetilde{\boldsymbol R}(\widehat{\boldsymbol\beta}_0-\boldsymbol\beta^*)|}_{(2)}+\underbrace{\sqrt{n}||\widehat{\boldsymbol v}-\boldsymbol v^*||_1||(\widehat{\boldsymbol R}-\widetilde{\boldsymbol R})\widehat{\boldsymbol\beta}_0||_\infty}_{(3)}\end{aligned}$$

By the proof of Theorem 1 and Lemma 3, we can show that

$$(1)\lesssim o_{\mathbb P}(1).$$

In addition, Lemma 4 implies that $(2)=o_{\mathbb P}(1)$. On the other hand,

$$||(\widehat{\boldsymbol R}-\widetilde{\boldsymbol R})\widehat{\boldsymbol\beta}_0||_\infty \le ||(\widehat{R}-\widetilde{\boldsymbol R})(\widehat{\boldsymbol\beta}_0-\boldsymbol\beta^*)||_\infty+||(\widehat{R}-\widetilde{\boldsymbol R})\boldsymbol\beta^*||_\infty.$$

By Corollary 1 and Theorem 1, we know $\sqrt{n}||(\widehat{\boldsymbol R}-\widetilde{\boldsymbol R})\widehat{\boldsymbol\beta}_0||_\infty\lesssim M_n\sqrt{log(d)}(s^*\sqrt{\frac{log(d)}{n}}+1)$ with probability greater than $1-d^{-1}$. Together with Lemma 4 we can show that

$$(3)\lesssim o_{\mathbb P}(1).$$

For $I_2$, we have

$$I_2=\underbrace{\sqrt{n}|\boldsymbol v^{*T}(\widehat{\boldsymbol R}-\widetilde{\boldsymbol R})(\widehat{\boldsymbol\beta}_0-\boldsymbol\beta^*)|}_{(4)}+\underbrace{\sqrt{n}|\boldsymbol v^{*T}\widetilde{\boldsymbol R}(\widehat{\boldsymbol\beta}_0-\boldsymbol\beta^*)|}_{(5)}.$$

By Theorem 1 and Corollary 1, we can show that $(4)\lesssim\sqrt{n}||\widehat{\boldsymbol R}-\widetilde{\boldsymbol R}||_{\max}||\widehat{\boldsymbol\beta}-\boldsymbol\beta^*||_1=o_{\mathbb P}(1)$, and $(5)\le||\widehat{\boldsymbol\beta}-\boldsymbol\beta^*||_1||\widetilde{\boldsymbol R}_{21}-\widetilde{\boldsymbol R}_{22}\boldsymbol w^*||_\infty=o_{\mathbb P}(1)$.
The bounds above imply that $\sqrt{n}|\widehat{S}(\widehat{\boldsymbol\beta}_0)-S(\boldsymbol\beta^*)|=o_{\mathbb P}(1)$. This together with the normality of $\widehat{\boldsymbol R}$ implied by Lemma 1, consistency of $\widehat\sigma$ and the assumption that $\sigma^*\ge K$ implies the desired result by applying Slutsky's theorem.
Note that in our case, we can estimate $\sigma^*$ by the plug-in estimator $\widehat\sigma=\widehat{\boldsymbol v}^T\widehat{\boldsymbol\Phi}\widehat{\boldsymbol v}$, where $\widehat{\boldsymbol\Phi}=\frac{1}{n}\sum_{i=1}^n(\widehat{\boldsymbol R}_i\widehat{\boldsymbol\beta}-\boldsymbol e_k)(\widehat{\boldsymbol R}_i\widehat{\boldsymbol\beta}-\boldsymbol e_k)^T$, and $(\widehat{\boldsymbol R}_i)_{jk}$ can be written as a function of $\widehat{\boldsymbol\theta}_i^{(jk)}$(or $\widehat{\boldsymbol\theta}_i^{(j)}$) that depends on the function $M_{jk}(.)$ $(orM_j(.))$, the type of $X_{ij}$, $X_{ik}$ and possibly the number of levels if at least one of the two is ordinal. This estimator is consistent by Theorem 1 and Lemma 3. □

## 2 Building confidence interval using the pseudo-score function

Without loss of generality, suppose we want to build a confidence interval for $\theta=\widehat\Omega_{1k}$. Consider a one-step type estimator:

$$\widetilde\theta:=\widehat\theta-\widehat{S}(\widehat{\boldsymbol\beta})\widehat\Omega_{11}.$$

The following Corollary shows that $\widetilde\theta$ is consistent and asymptotic normal:

**Corollary 2.** Suppose Assumption 1 - 5 hold. If $M_n^2 s^* log(d)/\sqrt{n} = o(1)$ and $\sigma^* \geq K$ for some constant $K$, then we have

$$\sqrt{n}(\widetilde{\theta} - \theta^*)/(\Omega_{11}^* \sigma^{*1/2}) \to_d N(0,1).$$

Therefore, a $(1-\alpha) \times 100\%$ confidence interval of $\theta^*$ is given by

$$[\,\widetilde{\theta} - \Phi^{-1}(1-\alpha/2)\widehat{\Omega}_{11}\widehat{\sigma}^{1/2}/\sqrt{n},\ \widetilde{\theta} + \Phi^{-1}(1-\alpha/2)\widehat{\Omega}_{11}\widehat{\sigma}^{1/2}/\sqrt{n}\,].$$

## 2.1 Proof of Corollary 2

*Proof.* It suffices to show that $\sqrt{n}|(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)/(\sigma^{*1/2}\Omega_{11}^*) + \boldsymbol{v}^{*T}(\widehat{\boldsymbol{R}}\boldsymbol{\beta}^* - \boldsymbol{e}_k)/\sigma^{*1/2}| = o_{\mathbb{P}}(1)$. By definition, we have

$$\begin{aligned}
\sqrt{n}|(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)/\Omega_{11}^* + \boldsymbol{v}^{*T}(\widehat{\boldsymbol{R}}\boldsymbol{\beta}^* - \boldsymbol{e}_k)| &= \sqrt{n}|(\widehat{\boldsymbol{\theta}} - \widehat{S}(\widehat{\boldsymbol{\beta}})\widehat{\Omega}_{11} - \boldsymbol{\theta}^*)/\Omega_{11}^* + \boldsymbol{v}^{*T}(\widehat{\boldsymbol{R}}\boldsymbol{\beta}^* - \boldsymbol{e}_k)| \\
&= \sqrt{n}|(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)/\Omega_{11}^* - \widehat{S}(\widehat{\boldsymbol{\beta}})\widehat{\Omega}_{11}/\Omega_{11}^* + \boldsymbol{v}^{*T}(\widehat{\boldsymbol{R}}\boldsymbol{\beta}^* - \boldsymbol{e}_k)| \\
&\leq \sqrt{n}|(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)/\Omega_{11}^* - \boldsymbol{v}^{*T}\widehat{\boldsymbol{R}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)| + \sqrt{n}|(\boldsymbol{v}^* - \widehat{\boldsymbol{v}})^T(\widehat{\boldsymbol{R}}\widehat{\boldsymbol{\beta}} - \boldsymbol{e}_k)| \\
&\quad + \sqrt{n}|(1 - \widehat{\Omega}_{11}/\Omega_{11}^*)\widehat{\boldsymbol{v}}^T(\widehat{\boldsymbol{R}}\widehat{\boldsymbol{\beta}} - \boldsymbol{e}_k)| \\
&:= I_1 + I_2 + I_3.
\end{aligned}$$

By Corollary 1 and Lemma 3, we can show that

$$|I_2| \leq ||\boldsymbol{v}^* - \widehat{\boldsymbol{v}}||_1 ||\widehat{\boldsymbol{R}}\widehat{\boldsymbol{\beta}} - \boldsymbol{e}_k||_\infty \leq o_{\mathbb{P}}(1).$$

For $I_1$, we have

$$|I_1| \leq \underbrace{\sqrt{n}|(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)/\Omega_{11}^* - \boldsymbol{v}^{*T}\widetilde{\boldsymbol{R}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)|}_{(1)} + \underbrace{\sqrt{n}|\boldsymbol{v}^{*T}(\widetilde{\boldsymbol{R}} - \widehat{\boldsymbol{R}})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)|}_{(2)}.$$

Here $(2) \lesssim \sqrt{n}||\widetilde{\boldsymbol{R}} - \widehat{\boldsymbol{R}}||_{\max}||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*||_1 = o_{\mathbb{P}}(1)$ by Corollary 1 and Theorem 1. In addition, we can show that

$$(1) \lesssim \sqrt{n}||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*||_1 ||\boldsymbol{T}||_\infty,$$

where $\boldsymbol{T} = [1/\Omega_{11}^* - (\widetilde{R}_{11} - \boldsymbol{w}^{*T}\widetilde{\boldsymbol{R}}_{21}), \widetilde{\boldsymbol{R}}_{12} - \boldsymbol{w}^*\widetilde{\boldsymbol{R}}_{22}]$. Theorem 1, the proof of Lemma 3 and the fact that

$$|1/\Omega_{11}^* - (\widetilde{R}_{11} - \boldsymbol{w}^{*T}\widetilde{\boldsymbol{R}}_{21})| \leq |\widetilde{R}_{11} - \Sigma_{11}| + |\boldsymbol{w}^{*T}(\widetilde{\boldsymbol{R}}_{21} - \boldsymbol{\Sigma}_{21})|$$

imply that $(1) = o_{\mathbb{P}}(1)$.
In addition, the proof of Theorem 2 implies that $\sqrt{n}|\widehat{v}^T(\widehat{R}\widehat{\boldsymbol{\beta}} - e_k)/\sigma^{*1/2}| = \mathcal{O}_{\mathbb{P}}(1)$. Therefore the consistency of $\widehat{\Omega}_{11}$ and the fact that $\Omega_{11}^* \geq \frac{1}{\tau}$ imply that $|I_3|/\sigma^{*1/2} = o_{\mathbb{P}}(1)$.
This completes the proof.

$\square$

# 3 Additional Simulation Details

## 3.1 RBE and EMLE method

For the rank-based estimator in Fan et al. [2017], due to the existence of multiple levels, the single level $p_j$ for $X_{ij}$ is chosen such that the data above or below this level are balanced as much as possible, i.e.,

$$\widehat{p}_j = \underset{1 \le k \le N_j}{\operatorname{argmin}} \left| \sum_{i=1}^{n} (\mathbb{1}(X_{ij} \ge k) - \mathbb{1}(X_{ij} < k)) \right|.$$

The RBE estimator in method (1) in the main text is defined as $\widehat{R}_{jk}^{(\widehat{p}_j, \widehat{q}_j)}$ for ordinal-by-ordinal entry and similarly for ordinal-by-continuous entry, which is often the best estimator among the class of base estimators in practice. The EMLE method is an extension to the direct estimation method Suggala et al. [2017] for ordinal graphical model, where the direct estimation for an ordinal-ny-continuous entry is given by

$$\widehat{\Sigma}_{jk} = \underset{-1 < s < 1}{\operatorname{argmax}} \prod_{i=1}^{n} \int_{u \in [\widehat{\triangle}_j^{(X_{ij})}, \widehat{\triangle}_j^{(X_{ij}+1)}]} \phi_2(x_1, X_{ik}; s) \mathrm{d}x_1.$$

For estimating the latent precision matrix, we compare the proposed estimator with the corresponding modified SCIO estimator by plugging in the RBE and EMLE estimates. For instance, the modified SCIO estimator with RBE is given by $\widehat{\boldsymbol{\Omega}}_k = \operatorname{argmin}_{\boldsymbol{\beta}} \{ \frac{1}{2} \boldsymbol{\beta}^T \widetilde{\boldsymbol{R}}_{\text{RBE}} \boldsymbol{\beta} - \boldsymbol{e}_k^T \boldsymbol{\beta} + \lambda_k ||\boldsymbol{\beta}||_1 \}$, where $\widehat{\boldsymbol{\Omega}}_k$ is the $k$th column of the estimated precision matrix. For all considered methods, the tuning parameter $\lambda_k$ is taken as $\lambda_k = \operatorname{argmin}_{\lambda_k \in \Lambda} ||\widehat{\boldsymbol{\Omega}}_k - \boldsymbol{\Omega}_k||_1$, where $\Lambda$ is a reasonable grid.

# References


Jianqing Fan, Han Liu, Yang Ning, and Hui Zou. High dimensional semiparametric latent graphical model for mixed data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(2):405–421, 2017.

Han Liu, Fang Han, Ming Yuan, John Lafferty, Larry Wasserman, et al. High-dimensional semiparametric gaussian copula graphical models. *The Annals of Statistics*, 40(4):2293–2326, 2012.

Arun Sai Suggala, Eunho Yang, and Pradeep Ravikumar. Ordinal graphical models: A tale of two approaches. In *International Conference on Machine Learning*, pages 3260–3269, 2017.