Supplements

A. Proof of Theorem 1

Following the definition, we can derive the PL for the PLA-GGM as:

$$\ell_{PL}\left(\{\mathbf{z}_i, g_i\}_{i\in[n]}; \mathbf{R}(\cdot), \boldsymbol{\Omega}_0\right) \propto \sum_{i=1}^{n}\sum_{j=1}^{p}\left\{ z_{ij}\left(\Omega_{ijj} + \sum_{j'\neq j}\Omega_{ijj'}z_{ij}z_{ij'}\right) - \frac{1}{2}z_{ij}^2 - \frac{1}{2}\left(\Omega_{ijj} + \sum_{j'\neq j}\Omega_{ijj'}z_{ij}z_{ij'}\right)^2\right\}.$$

Then Lemma 1 can be proved by the definition of $\mathbf{z}_{i,-j}$.

B. Proof of Lemma 2

According to the analysis in Section 3.1, we treat the PL as p partially-linear additive linear regressions. Then, for each regression, we can derive $\hat{M}_{ij}$ as the estimation to the smooth part following the rationale in (Fan et al., 2005). Combining the results for every regression, we can derive Lemma 2.

C. Proof of Theorem 1

In this Section, we prove the $\sqrt{n}$-sparsistency of the L_1-regularized MPPLE by following the widely-used primal-dual witness proof technique (Wainwright, 2009; Ravikumar et al., 2010; Yang & Ravikumar, 2011; Yang et al., 2015a). PDW is characterized by the following Lemma 3:

Lemma 3. *Let $\hat{\boldsymbol{\Omega}}_0$ be an optimal solution to (5), and $\hat{\mathbf{Z}}$ be the corresponding dual solution. If $\hat{\mathbf{Z}}$ satisfies $\left\|\hat{\mathbf{Z}}_N\right\|_\infty < 1$, then any given optimal solution to (5) $\tilde{\boldsymbol{\Omega}}_0$ satisfies $\tilde{\boldsymbol{\Omega}}_{0I} = \mathbf{0}$. Moreover, if $\mathbf{H}_{SS}$ is positive definite, then the solution to (5) is unique.*

Proof. Specifically, following the same rationale as Lemma 1 in Wainwright 2009, Lemma 1 in Ravikumar et al. 2010, and Lemma 2 in Yang & Ravikumar 2011, we can derive Lemma 3 characterizing the optimal solution of (5). ☐

Bound $\|\boldsymbol{\nabla} F(\boldsymbol{\Omega}_0^*)\|_\infty$

Before we use the PDW, we first provide a Lemma bounding $\|\boldsymbol{\nabla} F(\boldsymbol{\Omega}_0^*)\|_\infty$, which has been shown to be vital for PDW (Wainwright, 2009; Ravikumar et al., 2010; Yang & Ravikumar, 2011; Yang et al., 2015a).

Lemma 4. *Let $r := 4C_5\lambda$. For any $\epsilon_d > 0$, with probability of at least $1-\epsilon_d$, there exists $C_4 > 0$ and $N_d > 0$ satisfying the following two inequalities:*

$$\|\boldsymbol{\nabla} F(\boldsymbol{\Omega}_0^*)\|_\infty \leq C_4\sqrt{\frac{\log p}{n}}, \tag{8}$$

$$\left\|\tilde{\boldsymbol{\Theta}}_S - \boldsymbol{\Theta}_S^*\right\|_\infty \leq r, \tag{9}$$

for $n > N_d$.

Proof. We prove (8) and (9) in turn.

PROOF OF (8)

To begin with, we prove (8). We define

$$\lambda_{ij}^* = (\mathbf{1}_i - \mathbf{S}_{ij})^\top \mathbf{x}_j \boldsymbol{\Omega}_{0\cdot j}^*.$$

We use the F to denote the PPL defined in Definition 1. Then, the derivative of $F(\boldsymbol{\Omega}^*_{0\cdot j})$ is:

$$\frac{\partial F(\boldsymbol{\Omega}_0^*)}{\partial \Omega^*_{0j'j}} = \frac{\sum_{i=1}^n \left\{ -\left(\mathbf{1}_i - \mathbf{S}_{ij}\right)^\top \mathbf{y}_j \left[\left(\mathbf{1}_i - \mathbf{S}_{ij}\right)^\top \mathbf{x}_j\right]_{j'} + \lambda^*_{ij}\left[\left(\mathbf{1}_i - \mathbf{S}_{ij}\right)^\top \mathbf{x}_j\right]_{j'} \right\}}{n} + \frac{\sum_{i=1}^n \left\{ -\left(\mathbf{1}_i - \mathbf{S}_{ij'}\right)^\top \mathbf{y}_{j'} \left[\left(\mathbf{1}_i - \mathbf{S}_{ij'}\right)^\top \mathbf{x}_{j'}\right]_{j} + \lambda^*_{ij'}\left[\left(\mathbf{1}_i - \mathbf{S}_{ij}\right)^\top \mathbf{x}_{j'}\right]_{j} \right\}}{n}, \tag{10}$$

where $[\cdot]_j$ denotes the j^{th} component of the vector. For the ease of presentation, we define

$$\mathbf{y}'_j = \begin{bmatrix} \left(\mathbf{1}_1 - \mathbf{S}_{1j}\right)^\top \mathbf{y}_j \\ \vdots \\ \left(\mathbf{1}_n - \mathbf{S}_{nj}\right)^\top \mathbf{y}_j \end{bmatrix} \text{ and } \mathbf{x}'_j = \begin{bmatrix} \left(\mathbf{1}_1 - \mathbf{S}_{1j}\right)^\top \mathbf{x}_j \\ \vdots \\ \left(\mathbf{1}_n - \mathbf{S}_{nj}\right)^\top \mathbf{x}_j \end{bmatrix}.$$

Then, we consider

$$\frac{\mathbf{x}'^\top_j \mathbf{x}'_j \boldsymbol{\Omega}^*_{0\cdot j} - \mathbf{x}'^\top_j \mathbf{y}'_j}{n}, \tag{11}$$

whose j'^{th} component is just the target value

$$\frac{\sum_{i=1}^n \left\{ -\left(\mathbf{1}_i - \mathbf{S}_{ij}\right)^\top \mathbf{y}_j \left[\left(\mathbf{1}_i - \mathbf{S}_{ij}\right)^\top \mathbf{x}_j\right]_{j'} + \lambda^*_{ij}\left[\left(\mathbf{1}_i - \mathbf{S}_{ij}\right)^\top \mathbf{x}_j\right]_{j'} \right\}}{n}.$$

Therefore, we focus on bounding (11). Then,

$$\begin{aligned} \frac{\mathbf{x}'^\top_j \mathbf{x}'_j \boldsymbol{\Omega}^*_{0\cdot j} - \mathbf{x}'^\top_j \mathbf{y}'_j}{n} &= \frac{\mathbf{x}'^\top_j \mathbf{x}'_j \left[\boldsymbol{\Omega}^*_{0\cdot j} - \left(\mathbf{x}'^\top_j \mathbf{x}'_j\right)^{-1} \mathbf{x}'^\top_j \mathbf{y}'_j\right]}{n}, \\ &= \frac{\mathbf{x}'^\top_j \left(\mathbf{I} - \mathbf{S}_j\right)\left(\mathbf{M}_j + \boldsymbol{\epsilon}_j\right)}{n} \end{aligned}$$

where the second equality is due to Lemma 5.

We fist study $\frac{\mathbf{X}'^\top_j (\mathbf{I}-\mathbf{S}_j)\mathbf{M}_j}{n}$: according ot Lemma 8, for any $\epsilon_a > 0$, there exists $\delta_a > 0$ and $N_a > 0$ satisfying

$$\mathrm{P}\left\{ \left\| \frac{\mathbf{x}'^\top_j \left(\mathbf{I} - \mathbf{S}_j\right)\mathbf{M}_j}{n} \right\|_\infty > \delta_a \left[\frac{\log(\frac{1}{h})}{nh} + h^4 + 2h^2 \sqrt{\frac{\log(\frac{1}{h})}{nh}} \right] \right\} < \epsilon_a,$$

with $n > N_a$.

According to Assumption 1

$$\mathrm{P}\left\{ \left\| \frac{\mathbf{x}'^\top_j \left(\mathbf{I} - \mathbf{S}_j\right)\mathbf{M}_j}{n} \right\|_\infty > \delta_a C_1 \sqrt{\frac{\log p}{n}} \right\} < \epsilon_a. \tag{12}$$

Now, we study $\frac{\mathbf{x}'^\top_j (\mathbf{I}-\mathbf{S}_j)\boldsymbol{\epsilon}_j}{n}$. According to Lemma 9, we have

$$\frac{\mathbf{x}'^\top_j \left(\mathbf{I} - \mathbf{S}_j\right)\boldsymbol{\epsilon}_j}{n} = \sum_{i=1}^n \left\{ \mathbf{x}_{ij} - \mathbb{E}^\top \left[\mathbb{1}_{g_{i'} > g^*} \mathbf{Z}_{i,-j} \mathbf{Z}^\top_{i,-j} \mid g_i\right] \mathbb{E}^{-1}\left[\mathbb{1}^2_{g_{i'} > g^*} \mathbf{Z}_{i,-j} \mathbf{Z}^\top_{i,-j} \mid g_i\right] \tilde{\mathbf{x}}_{ij} \right\} \epsilon_{ij}(1 + o_p(1))/n,$$

uniformly for j. Note that $\frac{\mathbf{x}_j'^\top(\mathbf{I}-\mathbf{S}_j)\boldsymbol{\epsilon}_j}{n}$ is a $p \times 1$ vector. Therefore, for the j'^{th} component, we have

$$\begin{aligned}
&\left|\left[\frac{\mathbf{x}_j'^\top(\mathbf{I}-\mathbf{S}_j)\boldsymbol{\epsilon}_j}{n}\right]_{j'}\right| \\
\leq &\left|\sum_{i:g^i\leq g^*}\left(1-\mathbb{1}^2_{g^i>g^*}\right)z_{ij'}\epsilon_{ij}\right|(1+|o_p(1)|)/n \\
= &\frac{1}{2n}\left|\sum_{i=1}\left(1-\mathbb{1}^2_{g_i>g^*}\right)\left[\left(\frac{z_{ij'}+\epsilon_{ij}}{\sqrt{2}}\right)^2-1-\left(\frac{z_{ij'}-\epsilon_{ij}}{\sqrt{2}}\right)^2+1\right]\right|(1+|o_p(1)|)
\end{aligned} \tag{13}$$

It can be shown that $\left(\frac{z_{ij'}+\epsilon_{ij}}{\sqrt{2}}\right)^2$ and $\left(\frac{z_{ij'}-\epsilon_{ij}}{\sqrt{2}}\right)^2$ are independent and follow chi-squared distribution with degree equal to 1.

By Lemma 1 in (Laurent & Massart, 2000), the linear combinition of chi-squared random variables satisfies:

$$\mathrm{P}\left\{\sum_{i=1}\left(1-\mathbb{1}^2_{g_i>g^*}\right)\left[\left(\frac{z_{ij'}+\epsilon_{ij}}{\sqrt{2}}\right)^2-1\right]\geq 2\sqrt{nx}+2\epsilon_c\right\}\leq\exp(-\epsilon_c),$$

$$\mathrm{P}\left\{\sum_{i=1}\left(1-\mathbb{1}^2_{g_i>g^*}\right)\left[\left(\frac{z_{ij'}+\epsilon_{ij}}{\sqrt{2}}\right)^2-1\right]\leq -2\sqrt{nx}\right\}\leq\exp(-\epsilon_c),$$

$$\mathrm{P}\left\{-\sum_{i=1}\left(1-\mathbb{1}^2_{g_i>g^*}\right)\left[\left(\frac{z_{ij'}-\epsilon_{ij}}{\sqrt{2}}\right)^2-1\right]\leq -2\sqrt{nx}-2\epsilon_c\right\}\leq\exp(-\epsilon_c),$$

and

$$\mathrm{P}\left\{-\sum_{i=1}\left(1-\mathbb{1}^2_{g_i>g^*}\right)\left[\left(\frac{z_{ij'}-\epsilon_{ij}}{\sqrt{2}}\right)^2-1\right]\geq 2\sqrt{nx}\right\}\leq\exp(-\epsilon_c),$$

for any $\epsilon_c > 0$. Combing the previous four probabilistic bounds, we can derive

$$\begin{aligned}
\mathrm{P}\Bigg\{&\sum_{i=1}\left(1-\mathbb{1}^2_{g_i>g^*}\right)\left[\left(\frac{z_{ij'}+\epsilon_{ij}}{\sqrt{2}}\right)^2-1\right] \\
&-\sum_{i=1}\left(1-\mathbb{1}^2_{g_i>g^*}\right)\left[\left(\frac{z_{ij'}-\epsilon_{ij}}{\sqrt{2}}\right)^2-1\right]\geq 4\sqrt{n\epsilon}+2\epsilon_c\Bigg\}\leq\exp(-2\epsilon_c)
\end{aligned} \tag{14}$$

and

$$\begin{aligned}
\mathrm{P}\Bigg\{&\sum_{i=1}\left(1-\mathbb{1}^2_{g_i>g^*}\right)\left[\left(\frac{z_{ij'}+\epsilon_{ij}}{\sqrt{2}}\right)^2-1\right] \\
&-\sum_{i=1}\left(1-\mathbb{1}^2_{g_i>g^*}\right)\left[\left(\frac{z_{ij'}-\epsilon_{ij}}{\sqrt{2}}\right)^2-1\right]\leq -4\sqrt{n\epsilon}-2\epsilon_c\Bigg\}\leq\exp(-2\epsilon_c)
\end{aligned} \tag{15}$$

Taking (14) and (15) into (13), we can derive

$$\mathrm{P}\left\{\left|\left[\frac{\mathbf{X}_j'^\top(\mathbf{I}-\mathbf{S}_j)\boldsymbol{\epsilon}_j}{n}\right]_{j'}\right|\geq\left(2\sqrt{\frac{\epsilon_c}{n}}+\frac{\epsilon_c}{n}\right)(1+|o_p(1)|)\right\}\leq 2\exp(-2\epsilon_c). \tag{16}$$

Then, by the definition of $o_p(1)$, for any $\epsilon_b > 0$, there exists N_b so that for $n > N_b$:

$$\mathrm{P}\left\{|o_p(1)|\geq 1\right\}\leq\epsilon_b. \tag{17}$$

Combining (16) and (17), we derive

$$\mathrm{P}\left\{\left\|\frac{\mathbf{X}_j'^{\top}\left(\mathbf{I}-\mathbf{S}_j\right)\mathbf{M}_j}{n}\right\|_\infty \geq \left(4\sqrt{\frac{\epsilon_c}{n}}+2\frac{\epsilon_c}{n}\right)\right\} \leq 2p\exp(-2\epsilon_c)+\epsilon_b, \tag{18}$$

by a union bound. Eventually, according to (12) and (18), and by setting $\epsilon_c = 2\log p$ we prove:

$$\left\|\frac{\mathbf{x}_j'^{\top}\mathbf{x}_j'\mathbf{\Omega}_{0\cdot j}^* - \mathbf{x}_j'^{\top}\mathbf{y}_j'}{n}\right\|_\infty \leq (6+\delta_a C_1)\sqrt{\frac{2\log p}{n}},$$

with probability larger than $1-\epsilon_b-\epsilon_a-2p^{-1}$. Thus, for any $\epsilon_d>0$, there exists $C_4>0$ and $N_d>0$

$$\|\nabla F(\mathbf{\Theta}^*)\|_\infty \leq C_4\sqrt{\frac{\log p}{n}},$$

with probability larger than $1-\epsilon_d$, for $n>N_d$.

PROOF OF (9)

To prove (9), we use the fixed point method by defining a map $G(\mathbf{\Delta}_S) := -\mathbf{H}_{SS}^{-1}\left[\nabla_S F(\mathbf{\Omega}_{0S}^*+\mathbf{\Delta}_S)+\lambda\hat{\mathbf{Z}}_S\right]+\mathbf{\Delta}_S$. If $\|\mathbf{\Delta}\|_\infty \leq r$, by Taylor expansion of $\nabla_S F(\mathbf{\Omega}_0^*+\mathbf{\Delta})$ centered at $\nabla_S F(\mathbf{\Omega}_0^*)$,

$$\begin{aligned}\|G(\mathbf{\Delta}_S)\|_\infty =& \left\|-\mathbf{H}_{SS}^{-1}\left[\nabla_S F(\mathbf{\Omega}_{0S}^*)+\mathbf{H}_{SS}\mathbf{\Delta}_S+\boldsymbol{R}_S(\mathbf{\Delta})+\lambda\hat{\mathbf{Z}}_S\right]+\mathbf{\Delta}_S\right\|_\infty\\ =& \left\|-\mathbf{H}_{SS}^{-1}\left(\nabla_S F(\mathbf{\Omega}_{0S}^*)+\boldsymbol{R}_S(\mathbf{\Delta})+\lambda\hat{\mathbf{Z}}_S\right)\right\|_\infty\\ \leq& \left\|\mathbf{H}_{SS}^{-1}\right\|_\infty\left(\|\nabla_S F(\mathbf{\Omega}_{0S}^*)\|_\infty+\|R_S(\mathbf{\Delta})\|_\infty+\lambda\left\|\hat{\mathbf{Z}}_S\right\|_\infty\right)\\ \leq& C_2(\lambda+C_3r^2+\lambda)=C_2C_3r^2+2C_2\lambda,\end{aligned}$$

where the inequality is due to Assumption 4 and Assumption 5, and $\|\nabla_S F(\mathbf{\Theta}^*)\|_\infty \leq \lambda$ with a high probability, according to (8). Then, based on the definition of r, we can derive the upper bound of $\|G(\mathbf{\Delta}_S)\|_\infty$ as $\|G(\mathbf{\Delta}_S)\|_\infty \leq r/2+r/2=r$.

Therefore, according to the fixed point theorem (Ortega & Rheinboldt, 2000; Yang & Ravikumar, 2011), there exists $\mathbf{\Delta}_S$ satisfying $G(\mathbf{\Delta}_S)=\mathbf{\Delta}_S$, which indicates $\nabla_S F(\mathbf{\Omega}_0^*+\mathbf{\Delta})+\lambda\hat{\mathbf{Z}}_S=\mathbf{0}$. The optimal solution to (20) is unique, and thus $\tilde{\mathbf{\Delta}}_S=\mathbf{\Delta}_S$. Therefore, $\left\|\tilde{\mathbf{\Delta}}_S\right\|_\infty \leq r$, with probability larger than $1-\epsilon$. □

PDW

By Lemma 3, we can prove the sparsistency by building an optimal solution to (5) satisfying the strict dual feasibility (SDF) defined as $\left\|\hat{\mathbf{Z}}_N\right\|_\infty < 1$, which is summarized. Therefore, we now build a solution by solving a restricted problem.

SOLVE A RESTRICTED PROBLEM

First of all, we derive the KKT condition of (5):

$$\nabla F(\hat{\mathbf{\Omega}}_0)+\lambda\hat{\mathbf{Z}}=\mathbf{0}. \tag{19}$$

To construct an optimal primal-dual pair solution, we define $\tilde{\mathbf{\Omega}}_0$ as an optimal solution to the restricted problem:

$$\tilde{\mathbf{\Omega}}_0 := \min_{\mathbf{\Omega}_0} F(\mathbf{\Omega}_0)+\lambda\|\mathbf{\Omega}_0\|_1, \tag{20}$$

with $\mathbf{\Omega}_{0N}=\mathbf{0}$. $\tilde{\mathbf{\Omega}}_0$ is unique due to Lemma 3. Then, we define the subgradient corresponding to $\tilde{\mathbf{\Omega}}_0$ as $\tilde{\mathbf{Z}}$. Therefore, $(\tilde{\mathbf{\Omega}}_0,\tilde{\mathbf{Z}})$ is a pair of optimal solutions to the restricted problem (20). $\tilde{\mathbf{Z}}_S$ is determined according to the values of $\tilde{\mathbf{\Omega}}_{0S}$ via the KKT conditions of (20). Thus we have

$$\nabla_S F(\tilde{\mathbf{\Theta}})+\lambda\tilde{\mathbf{Z}}_S=\mathbf{0}, \tag{21}$$

where ∇_S represents the gradient components with respect to S. Letting $\hat{\mathbf{\Omega}}_0=\tilde{\mathbf{\Omega}}_0$, we determine $\tilde{\mathbf{Z}}_N$ according to (19). It now remains to show that $\tilde{\mathbf{Z}}_N$ satisfies SDF.

SDF

Now, we demonstrate that $\tilde{\mathbf{\Theta}}$ and $\tilde{\mathbf{Z}}$ satisfy SDF. We define $\tilde{\mathbf{\Delta}} := \tilde{\mathbf{\Theta}} - \mathbf{\Theta}^*$. By (21), and by the Taylor expansion of $\nabla_S F(\tilde{\mathbf{\Omega}}_0)$, we have that

$$\mathbf{H}_{SS}\tilde{\mathbf{\Delta}}_S + \nabla_S F(\mathbf{\Omega}_0^*) + \boldsymbol{R}_S(\tilde{\mathbf{\Delta}}) + \lambda\tilde{\mathbf{Z}}_S = \mathbf{0},$$

which means

$$\tilde{\mathbf{\Delta}}_S = \mathbf{H}_{SS}^{-1}\left[-\nabla_S F(\mathbf{\Omega}_0^*) - \boldsymbol{R}_S(\tilde{\mathbf{\Delta}}) - \lambda\tilde{\mathbf{Z}}_S\right], \tag{22}$$

where $\mathbf{H}_{SS}$ is positive definite and hence invertible.

By the definition of $\tilde{\mathbf{\Omega}}_0$ and $\tilde{\mathbf{Z}}$,

$$\nabla F(\tilde{\mathbf{\Omega}}_0) + \lambda\tilde{\mathbf{Z}} = \mathbf{0} \Rightarrow \nabla F(\mathbf{\Omega}_0^*) + \mathbf{H}\tilde{\mathbf{\Delta}} + \boldsymbol{R}(\tilde{\mathbf{\Omega}}_0) + \lambda\tilde{\mathbf{Z}} = \mathbf{0} \Rightarrow \nabla_N F(\tilde{\mathbf{\Theta}}) + \mathbf{H}_{NS}\tilde{\mathbf{\Delta}}_S + \boldsymbol{R}_N(\tilde{\mathbf{\Delta}}) + \lambda\tilde{\mathbf{Z}}_N = \mathbf{0}. \tag{23}$$

Due to (22),

$$\begin{aligned}
\lambda\left\|\tilde{\mathbf{Z}}_N\right\|_\infty =& \left\|-\mathbf{H}_{NS}\tilde{\mathbf{\Delta}}_S - \nabla_N F(\mathbf{\Omega}_0^*) - \boldsymbol{R}_N(\tilde{\mathbf{\Delta}})\right\|_\infty \\
\leq& \left\|\mathbf{H}_{NS}\mathbf{H}_{SS}^{-1}\left[-\nabla_S F(\mathbf{\Omega}_0^*) - \boldsymbol{R}_S(\tilde{\mathbf{\Delta}}) - \lambda\tilde{\mathbf{Z}}_S\right]\right\|_\infty + \left\|\nabla_N F(\mathbf{\Omega}_0^*) + \boldsymbol{R}_N(\tilde{\mathbf{\Delta}})\right\|_\infty \\
\leq& \left\|\mathbf{H}_{NS}\mathbf{H}_{SS}^{-1}\right\|_\infty \left\|\nabla_S F(\mathbf{\Omega}_0^*) + \boldsymbol{R}_S(\tilde{\mathbf{\Delta}})\right\|_\infty + \left\|\mathbf{H}_{NS}\mathbf{H}_{SS}^{-1}\right\|_\infty \left\|\lambda\tilde{\mathbf{Z}}_S\right\|_\infty + \left\|\nabla_N F(\mathbf{\Omega}_0^*) + \boldsymbol{R}_N(\tilde{\mathbf{\Delta}})\right\|_\infty
\end{aligned}.$$

Further, we use the Assumption 4,

$$\begin{aligned}
\lambda\left\|\tilde{\mathbf{Z}}_N\right\|_\infty \leq& (1-\alpha)\left(\|\nabla_S F(\mathbf{\Omega}_0^*)\|_\infty + \left\|\boldsymbol{R}_S(\tilde{\mathbf{\Delta}})\right\|_\infty\right) + (1-\alpha)\lambda + \left(\|\nabla_N F(\mathbf{\Omega}_0^*)\|_\infty + \left\|\boldsymbol{R}_N(\tilde{\mathbf{\Delta}})\right\|_\infty\right) \\
\leq& (2-\alpha)\left(\|\nabla F(\mathbf{\Omega}_0^*)\|_\infty + \left\|\boldsymbol{R}(\tilde{\mathbf{\Delta}})\right\|_\infty\right) + (1-\alpha)\lambda,
\end{aligned} \tag{24}$$

where we have used in the first inequality, and the third inequality is due to Assumption 4.

Now, we study $\|\nabla F(\mathbf{\Omega}_0^*)\|_\infty$.By Lemma 4 and the assumption on λ in Theorem 1, $\|\nabla F(\mathbf{\Theta}^*)\|_\infty \leq \frac{\alpha C_4}{4}\sqrt{\frac{\log p}{n}} \leq \frac{\alpha\lambda}{4}$, with probability larger than $1 - \epsilon_d$.

It remains to control $\left\|\boldsymbol{R}(\tilde{\mathbf{\Delta}})\right\|_\infty$. According to Assumption 5 and Lemma 4,

$$\left\|\boldsymbol{R}(\tilde{\mathbf{\Delta}})\right\|_\infty \leq C_3\|\mathbf{\Delta}\|_\infty^2 \leq C_3 r^2 \leq C_3(4C_2\lambda)^2 = \lambda\frac{64C_2^2C_3}{\alpha}\frac{\alpha\lambda}{4} \leq \left(C_5\sqrt{\frac{\log p}{n}}\right)\frac{64C_2^2C_3}{\alpha}\frac{\alpha\lambda}{4}, \tag{25}$$

where in the last inequality we have used the assumption $\lambda \leq C_5\sqrt{\frac{\log p}{n}}$ in Theorem 1. Therefore, when we choose $n \geq \left(64C_5C_2^2C_3/\alpha\right)^2 \log p$ in Theorem 1, from (25), we can conclude that $\left\|\boldsymbol{R}(\tilde{\mathbf{\Delta}})\right\|_\infty \leq \frac{\alpha\lambda}{4}$. As a result, $\lambda\left\|\hat{\mathbf{Z}}_N\right\|_\infty$ can be bounded by $\lambda\left\|\tilde{\mathbf{Z}}_N\right\|_\infty < \alpha\lambda/2 + \alpha\lambda/2 + (1-\alpha)\lambda = \lambda$. Combined with Lemma 3, we demonstrate that any optimal solution of (5) satisfies $\hat{\mathbf{\Theta}}_N = \mathbf{0}$. Furthermore, (9) controls the difference between the optimal solution of (5) and the real parameter by $\left\|\tilde{\mathbf{\Delta}}_S\right\|_\infty \leq r$, by the fact that $r \leq \|\mathbf{\Theta}_S^*\|_\infty$ in Theorem 1, $\hat{\mathbf{\Theta}}_S$ shares the same sign with $\mathbf{\Theta}_S^*$.

Auxiliary Lemmas

In this section, we provide and prove the used auxiliary lemmas.

Lemma 5. *For the graphical model defined in Section 2 parameterized by $\mathbf{\Omega}_0^*$, the conditional distribution of Z_{ij} follows*

$$(Z_{ij} \mid G_i = g_i) \sim \mathbf{Z}_{i,-j}^\top \mathbf{\Omega}_{0\cdot j} + M_{ij} + \epsilon_{ij},$$

where

$$[\mathbf{Z}_{i,-j}]_{j'} = \begin{cases} Z_{ij'} & j' \neq j \\ 1 & j' = j \end{cases}.$$

ϵ_{ij}'s follow the standard normal distribution, and ϵ_{ij} is independent with $\epsilon_{i'j}$ for $j \neq j' \in [p]$.

Proof. According to Lemma 1, the node-wise conditional distribution of a PLA-GGM follows a Gaussian distribution. Then, Lemma 5 can be proved. □

Lemma 6. *For a kernel regression on* $\{x_i, y_i\}_{i=1}^n$ *as the IID samples of* (X, Y)*. Assume that* $\mathbb{E}|Y|^s < \infty$ *and* $\sup_X \in |Y|^s f(X,Y)dY \leq \infty$*. Given that* $n^{2\epsilon-1}h \to \infty$ *for* $\epsilon < 1 - s^{-1}$*, we have*

$$\sum_x \left| \frac{1}{n} \sum_{i=1}^n \left[K_h(x_i - x) - \mathbb{E}\left\{ K_h(x_i - x) y_i \right\} \right] \right| = O_p\left(\left\{ \frac{\log(1/h)}{nh} \right\}^{1/2} \right).$$

Proof. Lemma 6 follows (Mack & Silverman, 1982). □

Lemma 7. *Suppose* $\mathbf{Y} = \{Y_1, Y_2 \cdots, Y_n\}$ *follows a multivariate Gaussian distribution, then* $\max|Y_i|$ *follows a sub-Gaussian distribution with variance* $\max var(Y_i)$*. Further, for any* $t > 0$*, the tail probability can be controlled via*

$$\mathrm{P}\left\{ \max|\epsilon_{ij}| \geq t \right\} \leq \exp\left(\frac{-t^2}{2} \right).$$

Lemma 8. *For any* $\epsilon > 0$*, there exists* $\delta > 0$ *and* $N > 0$*, so that when* $n > N$*, we have*

$$\mathrm{P}\left\{ \left\| \frac{\mathbf{X}_j'(\mathbf{I} - \mathbf{S}_j)\mathbf{M}_j}{n} \right\|_\infty \geq \delta c_n^2 \right\} \leq \epsilon,$$

uniformly for $j \in [p]$.

Proof. To start with, we review the definition of $\mathbf{S}_{ij}$

$$\mathbf{S}_{ij} = \begin{bmatrix} \mathbb{1}_{g_{i'} > g^*} \mathbf{z}_{i,-j}^\top & 0 \end{bmatrix} \left(\mathbf{D}_{ij}^\top \mathbf{W}_i \mathbf{D}_{ij} \right)^{-1} \mathbf{D}_{ij}^\top \mathbf{W}_i.$$

We first study $\mathbf{D}_{ij}^\top \mathbf{W}_i \mathbf{D}_{ij}$:

$$\mathbf{D}_{ij}^\top \mathbf{W}_i \mathbf{D}_{ij} = \begin{bmatrix} \sum_{i'=1}^n \mathbb{1}_{g_{i'} > g^*}^2 \mathbf{z}_{i',-j} \mathbf{z}_{i',-j}^\top \psi\left(|g_{i'} - g_i|/h \right) & \sum_{i'=1}^n \mathbb{1}_{g_{i'} > g^*}^2 \mathbf{z}_{i',-j} \mathbf{z}_{i',-j}^\top \frac{g_{i'} - g_i}{h} \psi\left(|g_{i'} - g_i|/h \right) \\ \sum_{i'=1}^n \mathbb{1}_{g_{i'} > g^*}^2 \mathbf{z}_{i',-j} \mathbf{z}_{i'-j}^\top \frac{g_{i'} - g_i}{h} \psi\left(|g_{i'} - g_i|/h \right) & \sum_{i'=1}^n \mathbb{1}_{g_{i'} > g^*}^2 \mathbf{z}_{i',-j} \mathbf{z}_{i',-j}^\top \left(\frac{g_{i'} - g_i}{h} \right)^2 \psi\left(|g_{i'} - g_i|/h \right) \end{bmatrix}.$$

To bound $\mathbf{D}_{ij}^\top \mathbf{W}_i \mathbf{D}_{ij}$ uniformly over j, we consider a random vector $\mathbf{B}_i = [\mathbb{1}_{g_{i'} > g^*} \mathbf{Z}_i^\top, 1]^\top$, with observations

$$\begin{bmatrix} \mathbf{b}_1 = \left[\mathbb{1}_{g_{i'} > g^*} \mathbf{z}_1^\top, 1 \right] \\ \vdots \\ \mathbf{b}_n = \left[\mathbb{1}_{g_{i'} > g^*} \mathbf{z}_n^\top, 1 \right] \end{bmatrix}.$$

Then, we study an auxiliary matrix

$$\mathbf{O}_i = \begin{bmatrix} \sum_{i'=1}^n \mathbb{1}_{g_{i'} > g^*}^2 \mathbf{b}_{i'} \mathbf{b}_{i'}^\top \psi\left(|g_{i'} - g_i|/h \right) & \sum_{i'=1}^n \mathbb{1}_{g_{i'} > g^*}^2 \mathbf{b}_{i'} \mathbf{b}_{i'}^\top \frac{g_{i'} - g_i}{h} \psi\left(|g_{i'} - g_i|/h \right) \\ \sum_{i'=1}^n \mathbb{1}_{g_{i'} > g^*}^2 \mathbf{b}_{i'} \mathbf{b}_{i'}^\top \frac{g_{i'} - g_i}{h} \psi\left(|g_{i'} - g_i|/h \right) & \sum_{i'=1}^n \mathbb{1}_{g_{i'} > g^*}^2 \mathbf{b}_{i'} \mathbf{b}_{i'}^\top \left(\frac{g_{i'} - g_i}{h} \right)^2 \psi\left(|g_{i'} - g_i|/h \right) \end{bmatrix}.$$

Therefore, the components of $\mathbf{D}_{ij}^\top \mathbf{W}_i \mathbf{D}_{ij}$ belong to $\mathbf{O}_i$, and each part of $\mathbf{O}_i$ is in the form of a kernel regression. By Lemma 6, we have

$$\mathbf{O}_i = nf(g_i) \mathbb{E}\left[\mathbf{B}_i \mathbf{B}_i^\top \mid g_i \right] \otimes \begin{bmatrix} 1 & 0 \\ 0 & \mu_2 \end{bmatrix} \left\{ 1 + O_p(c_n) \right\},$$

which holds uniformly for i. Therefore,

$$\mathbf{D}_{ij}^\top \mathbf{W}_i \mathbf{D}_{ij} = nf(g_i) \mathbb{E}\left[\mathbb{1}_{g_{i'} > g^*}^2 \mathbf{Z}_{i,-j} \mathbf{Z}_{i,-j}^\top \mid g_i \right] \otimes \begin{bmatrix} 1 & 0 \\ 0 & \mu_2 \end{bmatrix} \left\{ 1 + O_p(c_n) \right\} \tag{26}$$

holds uniformly for i with the same $O_p(c_n)$ for every j. Define

$$\boldsymbol{\alpha}_j(g_i) = \begin{bmatrix}\boldsymbol{\Omega}_{1\cdot j} & \cdots & \boldsymbol{\Omega}_{n\cdot j}\end{bmatrix}.$$

By the same technique, uniformly for i and with the same $O_p(c_n)$ for every j, we can show

$$\mathbf{D}_{ij}^\top \mathbf{W}_i \mathbf{M}_j = nf(g_i)\mathbb{E}\left[\mathbb{1}^2_{g_{i'}>g^*}\mathbf{Z}_{i,-j}\mathbf{Z}_{i,-j}^\top \mid g_i\right] \otimes \begin{bmatrix}1 & 0\end{bmatrix}^\top \boldsymbol{\alpha}_j(g_i)\left\{1+O_p(c_n)\right\}, \tag{27}$$

and

$$\mathbf{D}_{ij}^\top \mathbf{W}_i \mathbf{x}_j = nf(g_i)\mathbb{E}\left[\mathbb{1}_{g_{i'}>g^*}\mathbf{Z}_{i,-j}\mathbf{Z}_{i,-j}^\top \mid g_i\right] \otimes \begin{bmatrix}1 & 0\end{bmatrix}^\top \left\{1+O_p(c_n)\right\}. \tag{28}$$

Combining (26) and (27) we have

$$\begin{bmatrix}\tilde{\mathbf{x}}_j^\top & 0\end{bmatrix}\left(\mathbf{D}_{ij}^\top \mathbf{W}_i \mathbf{D}_{ij}\right)^{-1}\mathbf{D}_{ij}^\top \mathbf{W}_i \mathbf{M}_j = \tilde{\mathbf{x}}_j^\top \boldsymbol{\alpha}_j(g_i)\left\{1+O_p(c_n)\right\}. \tag{29}$$

Similarly, combining (26) and (28), we have

$$\mathbf{x}'_{ij} = \mathbf{x}_{ij} - \tilde{\mathbf{x}}_{ij}\mathbb{E}^{-1}\left[\mathbb{1}^2_{g_{i'}>g^*}\mathbf{Z}_{i,-j}\mathbf{Z}_i^\top \mid g_i\right]\mathbb{E}\left[\mathbb{1}_{g_{i'}>g^*}\mathbf{Z}_{i,-j}\mathbf{Z}_{i,-j}^\top \mid g_i\right]. \tag{30}$$

Next, we follow the rationale of the Lemma A.4 in (Fan et al., 2005), and combine (29) and (30). Finally, we have

$$\frac{\mathbf{x}'_j(\mathbf{I}-\mathbf{S}_j)\mathbf{M}_j}{n} = O_p(c_n^2)$$

uniformly for j. □

Lemma 9. *For any $\epsilon > 0$, there exists $N > 0$, so that when $n > N$, we have*

$$\left\|\mathbf{x}_j'^\top(\mathbf{I}-\mathbf{S}_j)\boldsymbol{\epsilon}_j\right\|_\infty \geq 2\sum_{i=1}^n\left\{\mathbf{x}_{ij} - \mathbb{E}^\top\left[\mathbb{1}_{g_{i'}>g^*}\mathbf{Z}_{i,-j}\mathbf{Z}_{i,-j}^\top \mid g_i\right]\mathbb{E}^{-1}\left[\mathbb{1}^2_{g_{i'}>g^*}\mathbf{Z}_{i,-j}\mathbf{Z}_i^\top \mid g_i\right]\tilde{\mathbf{x}}_{ij}\right\}\epsilon_{ij},$$

uniformly for $j \in [p]$ with probability less than ϵ.

Proof. By definition, we have

$$\mathbf{x}_j'^\top(\mathbf{I}-\mathbf{S}_j)\boldsymbol{\epsilon}_j = \sum_{i=1}^n \mathbf{x}'_{ij}\left\{\epsilon_{ij} - \begin{bmatrix}\tilde{\mathbf{x}}_{ij}^\top & 0\end{bmatrix}\left(\mathbf{D}_{ij}^\top \mathbf{W}_i \mathbf{D}_{ij}\right)^{-1}\mathbf{D}_{ij}^\top \mathbf{W}_i \boldsymbol{\epsilon}_j\right\}.$$

Using the technique in (26), we have

$$\begin{bmatrix}\tilde{\mathbf{x}}_{ij}^\top & 0\end{bmatrix}\left(\mathbf{D}_{ij}^\top \mathbf{W}_i \mathbf{D}_{ij}\right)^{-1}\mathbf{D}_{ij}^\top \mathbf{W}_i \boldsymbol{\epsilon}_j = \tilde{\mathbf{x}}_{ij}^\top \mathbb{E}^{-1}\left[\mathbb{1}^2_{g_{i'}>g^*}\mathbf{Z}_{i,-j}\mathbf{Z}_i^\top \mid g_i\right]\mathbb{E}\left[\tilde{\mathbf{x}}_{ij}^\top \mid g_i\right]O_p(c_n).$$

Therefore,

$$\mathbf{x}_j'^\top(\mathbf{I}-\mathbf{S}_j)\boldsymbol{\epsilon}_j = \sum_{i=1}^n\left\{\mathbf{x}_{ij} - \mathbb{E}^\top\left[\mathbb{1}_{g_{i'}>g^*}\mathbf{Z}_{i,-j}\mathbf{Z}_{i,-j}^\top \mid g_i\right]\mathbb{E}^{-1}\left[\mathbb{1}^2_{g_{i'}>g^*}\mathbf{Z}_{i,-j}\mathbf{Z}_i^\top \mid g_i\right]\tilde{\mathbf{x}}_{ij}\right\}\epsilon_{ij}[1+o_p(1)],$$

uniformly for j. □

D. Proof of Theorem 2

We first study CON-GGMs. According to (6) and (Eaton, 1983), we have

$$[\text{cov}\,(\mathbf{Z} \mid G = g)]^{-1} = \left[\mathbf{\Sigma}_{\mathbf{ZZ}} - \mathbf{\Sigma}_{\mathbf{Z}G}\mathbf{\Sigma}_{GG}^{-1}\mathbf{\Sigma}_{G\mathbf{Z}}\right]^{-1},$$

whose right-hand side has nothing to do with g. Therefore, the conditional distribution of $\mathbf{Z} \mid G = g$ follows a GGM with parameter $\left[\mathbf{\Sigma}_{\mathbf{ZZ}} - \mathbf{\Sigma}_{\mathbf{Z}G}\mathbf{\Sigma}_{GG}^{-1}\mathbf{\Sigma}_{G\mathbf{Z}}\right]^{-1}$ irrelevant to g. In other words CON-GGM is equivalent to assuming that G follows a normal distribution and $\mathbf{R}(g) = 0$ on the basis of the proposed PLA-GGM.

Then, we study LR-GGMs. Again, given $G = g$ for any g, we have

$$[\text{cov}\,(\mathbf{Z} \mid G = g)]^{-1} = \mathbf{\Omega}_0,$$

which has nothing to do with G either. Given $G = g$, the conditional distribution of $\mathbf{Z} \mid G = g$ follows a GGM with the parameter $\mathbf{\Omega}$. Therefore, LR-GGM is a special case of the proposed PLA-GGM by assuming $\mathbf{R}(g) = 0$.

E. Experiments

Data Simulation

To simulate the samples from PLA-GGMs, we first define

$$f(g) = \begin{cases} g - 10 & g > 12 \\ x + \frac{(x-12)^2}{4} - 11 & 10 < g \leq 12 \\ 0 & -10 < g \leq 10 \\ x + \frac{(x+12)^2}{4} + 11 & -12 < g \leq -10 \\ g + 10 & g \leq -12 \end{cases}$$

We provide the following procedure:

1. We consider $p = 10, 20, 50, 100$, and implement the following steps separately.
2. We randomly generate a sparse precision matrix as Ω_0 Specifically, each element of Ω_0 is drawn randomly to be non-zero with probability 0.3.
3. A dense precision matrix $\mathbf{W}$ is generated to build the confounding.
4. We take $\{-400, \cdots, 0, \cdots, 399\}$ as the confounders. For each $g \in \{-400, \cdots, 0, \cdots, 399\}$, the precision matrix is selected to be $\mathbf{\Omega}(g) = \mathbf{\Omega}_0 + f(g)\mathbf{W}$, and a sample is generated by a GGM with parameter $\mathbf{\Omega}(g)$. Thus, we get 800 samples.

Note that the procedure is equivalent to selecting $g^* = 10$.

Glass Brains for Brain Function Connectivity Estimation

We report the glass brains from other angles for the brain function connectivity estimation experiment in Section 6.2.

Schizophrenia Diagnosis using Different $\mathbb{1}_{\{|g| \geq g^*\}}$'s

We conduct the analysis in Section 6.2 using different $\mathbb{1}_{\{|g| \geq g^*\}}$'s. Specifically, we consider the function $1 - \exp\left(-kx^2\right)/2$ using $k = 144, 150$. The achieved accuracy using the parameter selected by the 10-fold cross validation and AIC are reported in Figure 9. The performance of PLA-GGMs is not hugely affected when selecting $\mathbb{1}_{\{|g| \geq g^*\}}$ in a reasonable range, which is consistent with our analysis in Theorem 1. Note that, if we select k too large, the PPL method will be not applicable. The reason is that a large k corresponds to a small g^*, and will induce few non-confounded samples observed. As a result, $\left(\mathbf{D}_{ij}^{\top}\mathbf{W}_i\mathbf{D}_{ij}\right)$ will be singular. In practice, if we use a relative large g^* corresponding to a small k, (2) will tend to be like $\mathbf{R}(g) = 0$ used in CON-GGMs and LR-GGMs.

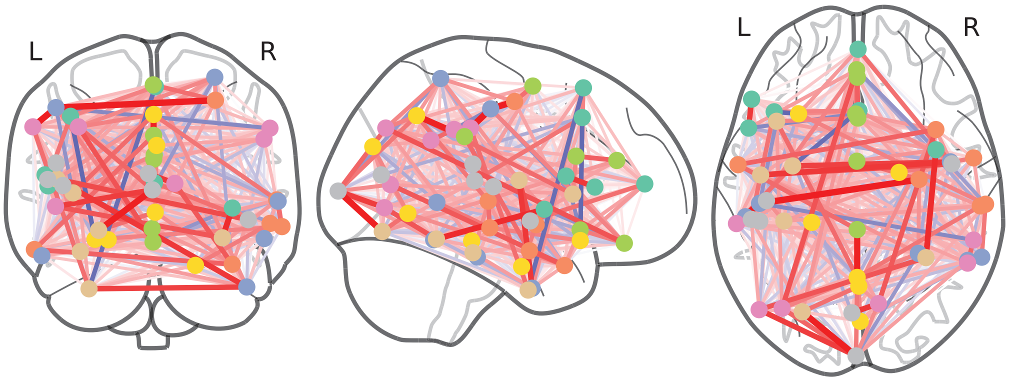


Figure 5: Controls using PLA-GGMs

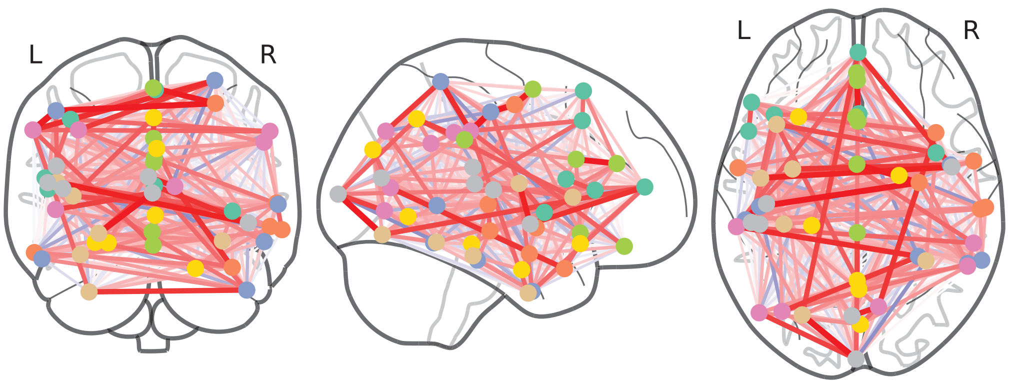


Figure 6: Patients using PLA-GGMs

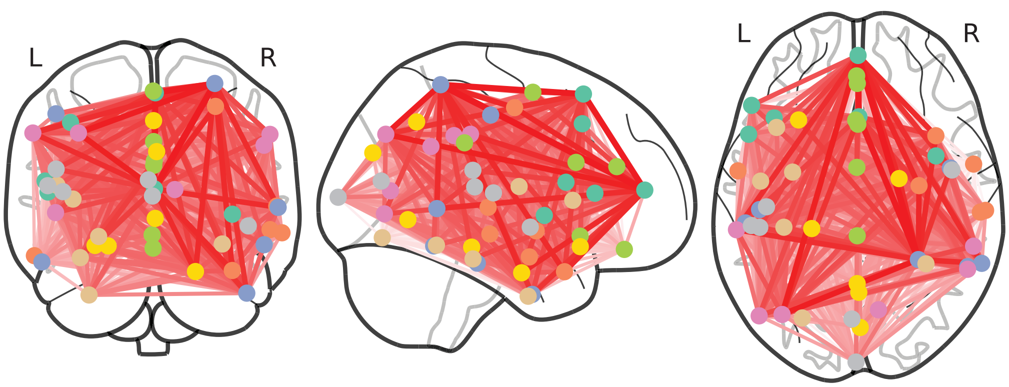


Figure 7: Controls using LR-GGMs

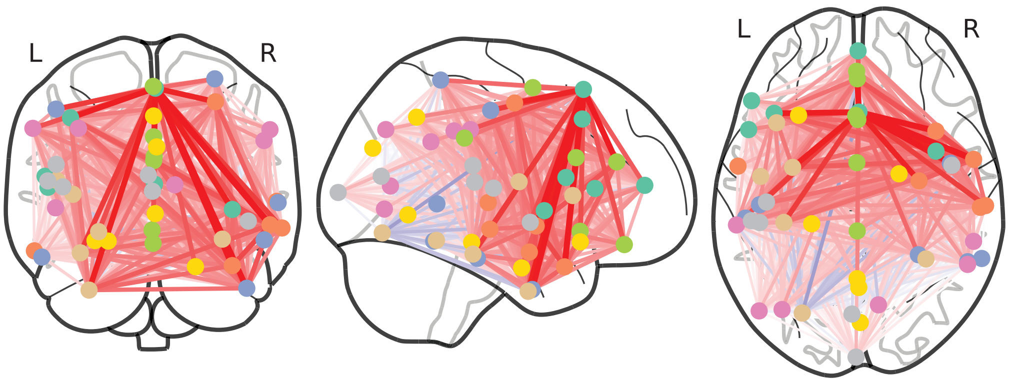
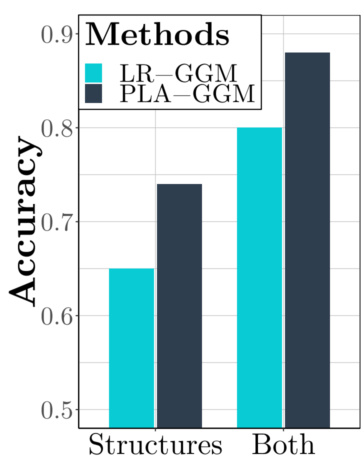
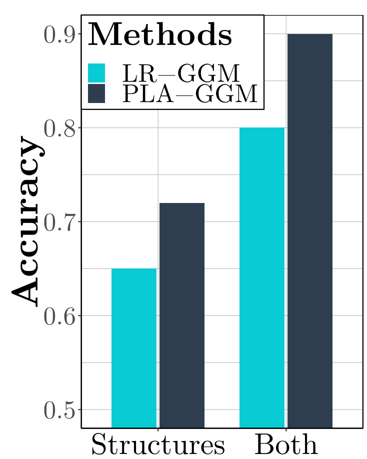


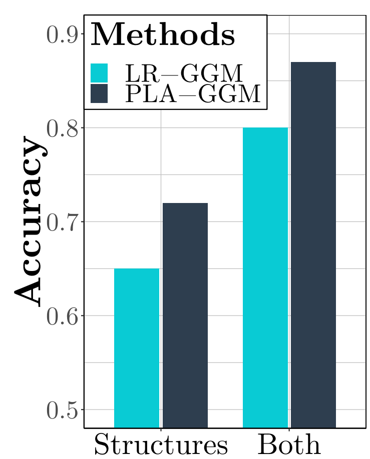
Figure 8: Patients using LR-GGMs



(a) $k = 100$



(b) $k = 144$



(c) $k = 150$

Figure 9: Diagnosis using different $\mathbb{1}_{\{|g|\geq g^*\}}$'s.