Electronic Journal of Statistics
Vol. 16 (2022) 3861–3909
ISSN: 1935-7524
https://doi.org/10.1214/22-EJS2028

De-noising analysis of noisy data under mixed graphical models*

Li-Pang Chen[1] **and Grace Y. Yi**[†2]

[1] *Department of Statistics, National Chengchi University.*
e-mail: lchen723@nccu.edu.tw

[2] *Department Statistical and Actuarial Sciences, Department of Computer Science, University of Western Ontario.*
e-mail: gyi5@uwo.ca

Abstract: Graphical models are useful to characterize the dependence structure of variables and have been commonly used for analysis of complex structured data. While various estimation methods have been developed under different graphical models, those methods are, however, inadequate to handle noisy data with measurement error. The development of most existing approaches relies on the implicit yet stringent assumption that the associated variables must be measured precisely. This assumption is unrealistic for many applications because mismeasurement in variables is usually presented in the data collection process. In this paper, we consider analysis of error-prone data under graphical models. To understand the impact of measurement error, we first study the asymptotic bias of the naive analysis which disregards the feature of measurement error in the variables. Furthermore, we develop a de-noising estimation procedure to account for measurement error effects. Theoretical results are established for the proposed method and numerical studies are reported to assess the finite sample performance of our proposed method.

Keywords and phrases: Asymptotic bias, de-noise, error-prone variable, graphical model, simulation-extrapolation algorithm.

Received March 2021.

1. Introduction

Graphical models have been commonly used to describe data with complex structures arising in various fields, including genomic studies, cancer research, and medical health record frameworks. In particular, the *Ising model* and the *Gaussian graphical model* are popularly used to describe association structures for binary and continuous variables, respectively, and different inference methods have been proposed accordingly. For instance, [20] proposed an inferential procedure to estimate the graph for the Ising model. [29] considered the Gaussian

*This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and partially supported by a Collaborative Research Team Project of the Canadian Statistical Sciences Institute. Yi is Canada Research Chair in Data Science (Tier 1). Her research was undertaken, in part, thanks to funding from the Canada Research Chairs program.

†Corresponding author.

graphical model and adopted an interior point optimization method to obtain the network structure for the variables. [13] proposed the graphical lasso to select the variables and estimate the model parameters. The Gaussian graphical model with complex features was explored by many authors such as [11], [12], [21], [23], [24], and [37].

Extensions of the Gaussian graphical model and the Ising model have been explored in the literature. The exponential family graphical model was developed by [28]. *Mixed graphical models* were proposed to handle settings where the variables contain both continuous and discrete ones. For example, [18] discussed the pseudo-likelihood method to deal with the mixture of the Gaussian graphical model and the Ising model. [9] proposed the group lasso to conduct inferences under the mixed graphical models. [5] extended mixed graphical models by considering a mixture of Poisson distributions and exponential distributions.

Even though analysis of graphical models has been widely explored, research gaps still remain. A typical feature that is left unattended to is about handling noisy data with measurement error. Concerning regression analysis, research on measurement error has attracted extensive attention. It has been well understood that treating noisy data as if they were error-free often yields seriously biased and misleading results (e.g., [4]; [30]). A large body of research papers have been available to address measurement error effects for different settings (e.g., [6]; [7]; [8]; [32]; [33]; [35]). For detailed discussions, see the monographs including [2]; [3]; [4]; [14]; [15]; [30]; and [31].

While great attention on measurement error has been directed to regression analysis, there has been little work on investigating measurement error effects under graphical models. In this paper we discuss this important problem and explore graphical models with error-prone measurements. We consider graphical models which are described by the exponential family distribution. We investigate the asymptotic biases of analyzing noisy data using the naive method which disregards noise effects. Our development accommodates all the three scenarios of noisy data with mismeasurement: (1) all the error-prone variables are continuous, (2) all the error-prone variables are discrete, and (3) error-prone variables include both continuous and discrete components. Furthermore, we develop a de-noise inference procedure to address the effects due to noisy data. We establish theoretical results for the proposed method.

The remainder is organized as follows. In Section 2, we introduce the graphical model and discuss its estimation procedure. In Section 3, we describe measurement error models to characterize noisy data and explore the asymptotic biases induced from the naive analysis ignoring the feature of noisy data. We propose a simulation-based method to accommodate mismeasurement effects in Section 4 and establish theoretical results in Section 5. Empirical studies, including simulation results and real data analysis, are provided in Section 6. We conclude the article with discussion in Section 7.

2. Model framework

2.1. Mixed graphical model

Consider a p-dimensional random vector $X = (X_1, \cdots, X_p)^\top$, where the dimension p is a given positive integer. We use an undirected *graph*, denoted $G = (V, E)$, to describe the relationship among the components of X, where $V = \{1, \cdots, p\}$, called the *vertex* set, includes all the indices of the random variables, and $E \subset V \times V$, called the *edge* set, contains all the pairs of connected indices in V. Here $(r, s) \in E$ if X_r and X_s are conditionally dependent, given other variables; and (r, s) is called an edge.

To characterize the distribution of a random vector X, one often considers the graphical model with the exponential family distribution (e.g., [5]; [28])

$$P(X; \theta, \Theta) = \exp\left\{ \sum_{r \in V} \theta_r F(X_r) + \sum_{(s,t) \in V \times V} \theta_{st} F(X_s) F(X_t) + \sum_{r \in V} H(X_r) - A(\theta, \Theta) \right\}, \tag{1}$$

where the function $A(\theta, \Theta)$ is the normalizing constant, also called the *log-partition function*, which makes (1) be integrated as 1, $F(\cdot)$ and $H(\cdot)$ are given functions, $\theta = (\theta_1, \cdots \theta_p)^\top$ is a p-dimensional vector of parameters, and $\Theta = [\theta_{st}]$ is a $p \times p$ symmetric matrix with zero diagonal elements. Parameters θ_{st} facilitate the association of X_s and X_t for $s \neq t$; $\theta_{st} = 0$ shows that X_s and X_t are conditionally independent, given the remaining variables.

Formulation (1) covers a broad class of models whose distribution is of an exponential form. Model (1) has been used to facilitate the setting where the components in X are either all continuous or all discrete. For example, for $r \in V$, we set $F(X_r) = \frac{X_r}{\sigma_r}$ and $H(X_r) = -\frac{X_r^2}{2\sigma_r^2}$ with σ_r being a positive constant, then (1) is proportional to

$$\exp\left(\sum_{r \in V} \frac{1}{\sigma_r} \theta_r X_r + \sum_{(s,t) \in V \times V} \frac{1}{\sigma_r \sigma_t} \theta_{st} X_s X_t - \sum_{r \in V} \frac{X_r^2}{2\sigma_r^2} \right), \tag{2}$$

yielding the *Gaussian graphical model* ([13]; [16]; [18]). If we constrain θ_r to be 0 for all $r \in V$ and let $F(X_r) = X_r$ and $H(X_r) = 0$ with $X_r \in \{-1, 1\}$, then (1) reduces to

$$\exp\left\{ \sum_{(s,t) \in V \times V} \theta_{st} X_s X_t - A(\Theta) \right\}, \tag{3}$$

which is the *Ising model* without the singleton for the simplicity [20]. The structure (1) was discussed by [5] and [28] in detail.

In applications, one may have to deal with the settings where X contains both continuous and discrete variables with $X = \left(X^{\mathrm{C}\top}, X^{\mathrm{D}\top}\right)^\top$. Here $X^{\mathrm{C}} = (X_1^{\mathrm{C}}, \cdots, X_{p^{\mathrm{C}}}^{\mathrm{C}})^\top$ is a p^{C}-dimensional continuous random vector and $X^{\mathrm{D}} = (X_1^{\mathrm{D}}, \cdots, X_{p^{\mathrm{D}}}^{\mathrm{D}})^\top$ is a p^{D}-dimensional discrete random vector with $p^{\mathrm{C}} + p^{\mathrm{D}} = p$. For ease of referral, in the following discussions, we respectively use superscripts "C" and "D" to label continuous and discrete random variables or their associated quantities.

Now we extend (1) to a *mixed graphical model*. To show the nature of the variables in X, we write $V = V^{\mathrm{C}} \cup V^{\mathrm{D}}$, where V^{C} and V^{D} represent the set of all the indices of continuous and discrete random variables, respectively. Let E^{C} and E^{D} represent the set of the edges restricted to the pairs of the indices in V^{C} and V^{D}, respectively, and let E^{CD} denote the set of *heterogeneous* edges for the pairs of the indices in V^{C} and V^{D}, i.e.,

$$E^{\mathrm{CD}} = \{(r, t') : r \in V^{\mathrm{C}}, t' \in V^{\mathrm{D}}, \text{ and } X_r^{\mathrm{C}} \text{ and } X_{t'}^{\mathrm{D}} \text{ are conditionally dependent given the remaining variables}\}.$$

In the same spirit of graphical model (1), we formulate a mixed graphical model as

$$\begin{aligned} & P\left(X^{\mathrm{C}}, X^{\mathrm{D}}; \theta, \Theta\right) \\ = \; & \exp\Bigg\{ \sum_{r \in V^{\mathrm{C}}} \theta_r^{\mathrm{C}} X_r^{\mathrm{C}} + \sum_{(r,t) \in V^{\mathrm{C}} \times V^{\mathrm{C}}} \theta_{rt}^{\mathrm{C}} X_r^{\mathrm{C}} X_t^{\mathrm{C}} + \sum_{r \in V^{\mathrm{C}}} H^{\mathrm{C}}(X_r^{\mathrm{C}}) \\ & + \sum_{r' \in V^{\mathrm{D}}} \theta_{r'}^{\mathrm{D}} X_{r'}^{\mathrm{D}} + \sum_{(r',t') \in V^{\mathrm{D}} \times V^{\mathrm{D}}} \theta_{r't'}^{\mathrm{D}} X_{r'}^{\mathrm{D}} X_{t'}^{\mathrm{D}} + \sum_{r' \in V^{\mathrm{D}}} H^{\mathrm{D}}(X_{r'}^{\mathrm{D}}) \\ & + \sum_{(r,t') \in V^{\mathrm{C}} \times V^{\mathrm{D}}} \theta_{rt'}^{\mathrm{CD}} X_r^{\mathrm{C}} X_{t'}^{\mathrm{D}} - A_{\mathrm{mix}}(\theta, \Theta) \Bigg\}, \end{aligned} \tag{4}$$

where $\theta = \left(\theta^{\mathrm{C}\top}, \theta^{\mathrm{D}\top}\right)^\top$ with $\theta^{\mathrm{C}} = \left(\theta_1^{\mathrm{C}}, \cdots, \theta_{p^{\mathrm{C}}}^{\mathrm{C}}\right)^\top$ and $\theta^{\mathrm{D}} = \left(\theta_1^{\mathrm{D}}, \cdots, \theta_{p^{\mathrm{D}}}^{\mathrm{D}}\right)^\top$; $\Theta = \begin{bmatrix} \Theta^{\mathrm{C}} & \Theta^{\mathrm{CD}} \\ \Theta^{\mathrm{CD}\top} & \Theta^{\mathrm{D}} \end{bmatrix}$ is a block matrix with $p^{\mathrm{C}} \times p^{\mathrm{C}}$ matrix $\Theta^{\mathrm{C}} = [\theta_{rt}^{\mathrm{C}}]$, $p^{\mathrm{D}} \times p^{\mathrm{D}}$ matrix $\Theta^{\mathrm{D}} = [\theta_{r't'}^{\mathrm{D}}]$, and $p^{\mathrm{C}} \times p^{\mathrm{D}}$ matrix $\Theta^{\mathrm{CD}} = [\theta_{rt'}^{\mathrm{CD}}]$; $A_{\mathrm{mix}}(\theta, \Theta)$ is the normalizing constant of (4); $H^{\mathrm{C}}(\cdot)$ and $H^{\mathrm{D}}(\cdot)$ are given functions based on continuous and discrete random variables, respectively, playing the same role as $H(\cdot)$ in (1); θ_r^{C} and $\theta_{r'}^{\mathrm{D}}$ are the parameters corresponding to X_r^{C} and $X_{r'}^{\mathrm{D}}$ for $r \in V^{\mathrm{C}}$ and $r' \in V^{\mathrm{D}}$, respectively; θ_{rt}^{C} and $\theta_{r't'}^{\mathrm{D}}$ are the parameters indicating the pairwise dependence of the variables in E^{C} and E^{D}; and $\theta_{rt'}^{\mathrm{CD}}$ is the parameter showing the pairwise dependence of $(X_r^{\mathrm{C}}, X_{t'}^{\mathrm{D}})$ for $r \in V^{\mathrm{C}}$ and $t' \in V^{\mathrm{D}}$. Different from the setting in [5], our setup accommodates both homogeneous edges and heterogeneous edges.

2.2. Notation and definition

In this subsection, we first introduce basic notation and then present definitions for the mixed graphical model in Section 2.1.

For a p-dimensional vector $z = (z_1, \cdots, z_p)^\top$, let $\|z\|_\infty \triangleq \max_{1\le i\le p} |z_i|$ denote the infinity norm of z, let $\|z\|_1 \triangleq \sum_{j=1}^p |z_j|$ denote the L_1 norm of z, and let $\|z\|_2^2 \triangleq \sum_{j=1}^p |z_j|^2$ denote the square of the L_2 norm of z. For $r = 1, \cdots, p$, let $z_{\setminus r} = (z_1, \cdots, z_{r-1}, z_{r+1}, \cdots, z_p)^\top$ denote the $(p-1)\times 1$ subvector of z with the rth component removed.

For a $p \times p$ matrix $\Omega = [\omega_{ij}]$, its L_1-norm and the Frobenius norm are respectively defined as

$$\|\Omega\|_1 = \max_{j=1,\cdots,p} \sum_{i=1}^p |\omega_{ij}| \quad \text{and} \quad \|\Omega\|_F = \sqrt{\sum_{i=1}^p \sum_{j=1}^p |\omega_{ij}|^2},$$

and let $\Lambda_{\max}(\Omega)$ denote its largest eigenvalue.

For $r = 1, \cdots, p$, let $\Omega_{\setminus r}$ denote the $(p-1)\times(p-1)$ submatrix of Ω with its rth row and rth column deleted. For a positive integer ζ, let Ω^ζ denote the product that Ω multiplies itself ζ times.

For a univariate differentiable function $f(\beta)$, let $f'(\beta)$, $f''(\beta)$, and $f'''(\beta)$ denote its first, second, and third order derivative with respect to β, respectively. For a twice differentiable multivariate function $f(\beta)$, let $\nabla_\beta f(\beta) = \frac{\partial f(\beta)}{\partial \beta}$ and $\nabla^2_\beta f(\beta) = \frac{\partial^2 f(\beta)}{\partial \beta \partial \beta^\top}$ denote the gradient and Hessian of $f(\beta)$, respectively.

Next, for the mixed graphical model in Section 2.1, we describe basic notation or definitions. For $r \in V^{\mathrm{C}}$, let $\mathcal{N}^{\mathrm{C}}(r) = \{t \in V^{\mathrm{C}} : (r,t) \in E^{\mathrm{C}}\}$ denote the *homogeneous* neighbourhood of r containing all the indices of continuous random variables X_t^{C} that are conditionally dependent on X_r^{C}, given other variables; and let $\mathcal{N}^{\mathrm{CD}}(r) = \{t' \in V^{\mathrm{D}} : (r,t') \in E^{\mathrm{CD}}\}$ represent the *heterogeneous* neighbourhood of r containing all the indices of discrete random variables $X_{t'}^{\mathrm{D}}$ that are conditionally dependent on X_r^{C}, given the remaining variables. Similarly, for $r' \in V^{\mathrm{D}}$, define $\mathcal{N}^{\mathrm{D}}(r') = \{t' \in V^{\mathrm{D}} : (r',t') \in E^{\mathrm{D}}\}$ and $\mathcal{N}^{\mathrm{DC}}(r') = \{t \in V^{\mathrm{C}} : (t,r') \in E^{\mathrm{CD}}\}$.

For $r \in V^{\mathrm{C}}$ and $r' \in V^{\mathrm{D}}$, let

$$\theta_r^{\mathrm{CD}} = \left(\theta_{r1}^{\mathrm{CD}}, \cdots, \theta_{rp^{\mathrm{D}}}^{\mathrm{CD}}\right)^\top \quad \text{and} \quad \theta_{r'}^{\mathrm{DC}} = \left(\theta_{1r'}^{\mathrm{CD}}, \cdots, \theta_{p^{\mathrm{C}}r'}^{\mathrm{CD}}\right)^\top. \tag{5}$$

That is, θ_r^{CD} is the vector collecting all the parameters showing the conditional dependence of a discrete random variable on the continuous random variable X_r^{C}; $\theta_{r'}^{\mathrm{DC}}$ is the vector collecting all the parameters showing the conditional dependence of a continuous random variable on the discrete random variable $X_{r'}^{\mathrm{D}}$. Moreover, define

$$\theta^{\mathrm{C}}(r) = \left(\theta_r^{\mathrm{C}}, \theta_{\setminus r}^{\mathrm{C}\top}, \theta_r^{\mathrm{CD}\top}\right)^\top \text{ and } \theta^{\mathrm{D}}(r') = \left(\theta_{r'}^{\mathrm{D}}, \theta_{\setminus r'}^{\mathrm{D}\top}, \theta_{r'}^{\mathrm{DC}\top}\right)^\top, \tag{6}$$

where $\theta_{\setminus r}^{\mathrm{C}} = (\theta_{rt}^{\mathrm{C}} : t \in V^{\mathrm{C}} \setminus \{r\})^\top$ and $\theta_{\setminus r'}^{\mathrm{D}} = (\theta_{r't'}^{\mathrm{D}} : t' \in V^{\mathrm{D}} \setminus \{r'\})^\top$. Let $\theta_0^{\mathrm{C}}(r) = \left(\theta_{0;r}^{\mathrm{C}}, \theta_{0;\setminus r}^{\mathrm{C}\top}, \theta_{0;r}^{\mathrm{CD}\top}\right)^\top$ and $\theta_0^{\mathrm{D}}(r') = \left(\theta_{0;r'}^{\mathrm{D}}, \theta_{0;\setminus r'}^{\mathrm{D}\top}, \theta_{0;r'}^{\mathrm{DC}\top}\right)^\top$ denote the true

values of $\theta^{\mathrm{C}}(r)$ and $\theta^{\mathrm{D}}(r')$, respectively. Let $\theta^{\mathrm{C}}_{0;rt}$, $\theta^{\mathrm{D}}_{0;r't'}$, and $\theta^{\mathrm{CD}}_{0;rt'}$ denote the true value of θ^{C}_{rt}, $\theta^{\mathrm{D}}_{r't'}$ and $\theta^{\mathrm{CD}}_{rt'}$, respectively.

Now we re-express $\theta^{\mathrm{C}}_0(r)$ and $\theta^{\mathrm{D}}_0(r')$ by explicitly spelling out their nonzero and zero components. For $r \in V^{\mathrm{C}}$, let

$$\mathcal{S}^{\mathrm{C}}_r(V^{\mathrm{C}}) = \left\{t \in V^{\mathrm{C}} : \theta^{\mathrm{C}}_{0;rt} \neq 0\right\} \quad \text{and} \quad \mathcal{S}^{\mathrm{C}}_r(V^{\mathrm{D}}) = \left\{t' \in V^{\mathrm{D}} : \theta^{\mathrm{CD}}_{0;rt'} \neq 0\right\}.$$

Then let

$$\mathcal{S}^{\mathrm{C}}_r \triangleq \mathcal{S}^{\mathrm{C}}_r(V^{\mathrm{C}}) \cup \mathcal{S}^{\mathrm{C}}_r(V^{\mathrm{D}}) \tag{7}$$

denote the set of all indices having connections with continuous random variable X^{C}_r.

Similarly, for $r' \in V^{\mathrm{D}}$, let

$$\mathcal{S}^{\mathrm{D}}_{r'}(V^{\mathrm{D}}) = \left\{t' \in V^{\mathrm{D}} : \theta^{\mathrm{D}}_{0;r't'} \neq 0\right\} \quad \text{and} \quad \mathcal{S}^{\mathrm{D}}_{r'}(V^{\mathrm{C}}) = \left\{t \in V^{\mathrm{C}} : \theta^{\mathrm{CD}}_{0;tr'} \neq 0\right\}.$$

Then the set

$$\mathcal{S}^{\mathrm{D}}_{r'} \triangleq \mathcal{S}^{\mathrm{D}}_{r'}(V^{\mathrm{D}}) \cup \mathcal{S}^{\mathrm{D}}_{r'}(V^{\mathrm{C}}) \tag{8}$$

includes all indices having connections with discrete random variable $X^{\mathrm{D}}_{r'}$.

Let $\overline{\mathcal{S}}^{\mathrm{C}}_r$ and $\overline{\mathcal{S}}^{\mathrm{D}}_{r'}$ denote the complement sets of (7) and (8), respectively. Let $d^{\mathrm{C}}_r = |\mathcal{S}^{\mathrm{C}}_r|$ and $d^{\mathrm{D}}_{r'} = |\mathcal{S}^{\mathrm{D}}_{r'}|$ denote number of elements in $\mathcal{S}^{\mathrm{C}}_r$ and $\mathcal{S}^{\mathrm{D}}_{r'}$, respectively, and let $d^{\mathrm{C}} \triangleq \max\{d^{\mathrm{C}}_r : r \in V^{\mathrm{C}}\}$ and $d^{\mathrm{D}} \triangleq \max\{d^{\mathrm{D}}_{r'} : r' \in V^{\mathrm{D}}\}$. With (7) and (8), for $r \in V^{\mathrm{C}}$ and $r' \in V^{\mathrm{D}}$, we define

$$\theta_{0;\mathcal{S}^{\mathrm{C}}_r} \triangleq \left(\theta^{\mathrm{C}}_{0;rt}, \theta^{\mathrm{CD}}_{0;rt'} : t \in \mathcal{S}^{\mathrm{C}}_r(V^{\mathrm{C}}) \text{ and } t' \in \mathcal{S}^{\mathrm{C}}_r(V^{\mathrm{D}})\right)^\top \tag{9}$$

and

$$\theta_{0;\mathcal{S}^{\mathrm{D}}_{r'}} \triangleq \left(\theta^{\mathrm{D}}_{0;r't'}, \theta^{\mathrm{CD}}_{0;tr'} : t \in \mathcal{S}^{\mathrm{D}}_{r'}(V^{\mathrm{C}}) \text{ and } t' \in \mathcal{S}^{\mathrm{D}}_{r'}(V^{\mathrm{D}})\right)^\top, \tag{10}$$

and define

$$\theta_{0;\overline{\mathcal{S}}^{\mathrm{C}}_r} \triangleq \left(\theta^{\mathrm{C}}_{0;rt}, \theta^{\mathrm{CD}}_{0;rt'} : t, t' \in \overline{\mathcal{S}}^{\mathrm{C}}_r\right)^\top \quad \text{and} \quad \theta_{0;\overline{\mathcal{S}}^{\mathrm{D}}_{r'}} \triangleq \left(\theta^{\mathrm{D}}_{0;r't'}, \theta^{\mathrm{CD}}_{0;tr'} : t, t' \in \overline{\mathcal{S}}^{\mathrm{D}}_{r'}\right)^\top.$$

Consequently, $\theta^{\mathrm{C}}_0(r)$ and $\theta^{\mathrm{D}}_0(r')$ can be, respectively, rewritten as

$$\theta^{\mathrm{C}}_0(r) = \left(\theta^{\mathrm{C}}_{0;r}, \theta^\top_{0;\mathcal{S}^{\mathrm{C}}_r}, \theta^\top_{0;\overline{\mathcal{S}}^{\mathrm{C}}_r}\right)^\top \quad \text{and} \quad \theta^{\mathrm{D}}_0(r') = \left(\theta^{\mathrm{D}}_{0;r'}, \theta^\top_{0;\mathcal{S}^{\mathrm{D}}_{r'}}, \theta^\top_{0;\overline{\mathcal{S}}^{\mathrm{D}}_{r'}}\right)^\top. \tag{11}$$

2.3. Separate estimation

In this subsection, we describe a *neighbourhood selection* approach to identify the structure of mixed graphical models and estimate the associated parameters. The procedures are basically addressed to continuous and discrete random variables in a separate manner.

Following the same spirit of deriving Proposition 1 of [28], we obtain that the conditional distribution of the rth continuous random variable X_r^{C}, given $X_{\setminus r}^{\mathrm{C}}$ and X^{D}, is

$$P\left(X_r^{\mathrm{C}} \middle| X_{\setminus r}^{\mathrm{C}}, X^{\mathrm{D}}\right) = \exp\left\{X_r^{\mathrm{C}} \eta_{X,r}^{\mathrm{C}} + H^{\mathrm{C}}\left(X_r^{\mathrm{C}}\right) - K^{\mathrm{C}}(\eta_{X,r}^{\mathrm{C}})\right\}, \tag{12}$$

where $\eta_{X,r}^{\mathrm{C}} = \theta_r^{\mathrm{C}} + \sum_{t \in V^{\mathrm{C}} \setminus \{r\}} \theta_{rt}^{\mathrm{C}} X_t^{\mathrm{C}} + \sum_{t' \in V^{\mathrm{D}}} \theta_{rt'}^{\mathrm{CD}} X_{t'}^{\mathrm{D}}$ and $K^{\mathrm{C}}(\eta_{X,r}^{\mathrm{C}})$ is the normalizing constant ensuring the integration of the right-hand side (12) to be one. Similarly, the conditional distribution of the r'th discrete random variable $X_{r'}^{\mathrm{D}}$, given $X_{\setminus r'}^{\mathrm{D}}$ and X^{C}, is given by

$$P\left(X_{r'}^{\mathrm{D}} \middle| X_{\setminus r'}^{\mathrm{D}}, X^{\mathrm{C}}\right) = \exp\left\{X_{r'}^{\mathrm{D}} \eta_{X,r'}^{\mathrm{D}} + H^{\mathrm{D}}\left(X_{r'}^{\mathrm{D}}\right) - K^{\mathrm{D}}(\eta_{X,r'}^{\mathrm{D}})\right\} \tag{13}$$

with $\eta_{X,r'}^{\mathrm{D}} = \theta_{r'}^{\mathrm{D}} + \sum_{t' \in V^{\mathrm{D}} \setminus \{r'\}} \theta_{r't'}^{\mathrm{D}} X_{t'}^{\mathrm{D}} + \sum_{t \in V^{\mathrm{C}}} \theta_{tr'}^{\mathrm{CD}} X_t^{\mathrm{C}}$ and the normalizing constant $K^{\mathrm{D}}(\eta_{X,r'}^{\mathrm{D}})$.

To estimate parameters of model (4) or identify the graphic structure of (4), we assume the availability of a random sample $\left\{X^{(i)} : i = 1, \cdots, n\right\}$ of size n, where $X^{(i)} = \left(X^{\mathrm{C}(i)\top}, X^{\mathrm{D}(i)\top}\right)^{\top}$ represents the random vector for observation i which follows the same distribution as $X = \left(X^{\mathrm{C}\top}, X^{\mathrm{D}\top}\right)^{\top}$. By (12) and (13), we define

$$\ell^{\mathrm{C}}\left(\theta^{\mathrm{C}}(r)\right) = -\frac{1}{n}\sum_{i=1}^{n} \log\left\{P\left(X_r^{\mathrm{C}(i)} \middle| X_{\setminus r}^{\mathrm{C}(i)}, X^{\mathrm{D}(i)}\right)\right\} \tag{14}$$

and

$$\ell^{\mathrm{D}}\left(\theta^{\mathrm{D}}(r')\right) = -\frac{1}{n}\sum_{i=1}^{n} \log\left\{P\left(X_{r'}^{\mathrm{D}(i)} \middle| X_{\setminus r'}^{\mathrm{D}(i)}, X^{\mathrm{C}(i)}\right)\right\}. \tag{15}$$

To carry out selection of those nonzero elements associated with $\theta_{\setminus r}^{\mathrm{C}}$, $\theta_{\setminus r'}^{\mathrm{D}}$, θ_r^{CD} and $\theta_{r'}^{\mathrm{DC}}$, we implement the lasso penalty function [25] and obtain the estimators of $\theta^{\mathrm{C}}(r)$ and $\theta^{\mathrm{D}}(r')$, respectively, by

$$\widetilde{\theta}^{\mathrm{C}}(r) = \underset{\theta^{\mathrm{C}}(r)}{\operatorname{argmin}} \left\{\ell^{\mathrm{C}}\left(\theta^{\mathrm{C}}(r)\right) + \lambda_{n1}^{\mathrm{C}} \left\|\theta_{\setminus r}^{\mathrm{C}}\right\|_1 + \lambda_{n2}^{\mathrm{C}} \left\|\theta_r^{\mathrm{CD}}\right\|_1\right\} \tag{16}$$

and

$$\widetilde{\theta}^{\mathrm{D}}(r') = \underset{\theta^{\mathrm{D}}(r')}{\operatorname{argmin}} \left\{\ell^{\mathrm{D}}\left(\theta^{\mathrm{D}}(r')\right) + \lambda_{n1}^{\mathrm{D}} \left\|\theta_{\setminus r'}^{\mathrm{D}}\right\|_1 + \lambda_{n2}^{\mathrm{D}} \left\|\theta_{r'}^{\mathrm{DC}}\right\|_1\right\}, \tag{17}$$

where $\lambda_{n1}^{\mathrm{C}}$, $\lambda_{n2}^{\mathrm{C}}$, $\lambda_{n1}^{\mathrm{D}}$, and $\lambda_{n2}^{\mathrm{D}}$ are tuning parameters. For simplicity, we set $\lambda_{n1}^{\mathrm{C}} = \lambda_{n2}^{\mathrm{C}}$ and $\lambda_{n1}^{\mathrm{D}} = \lambda_{n2}^{\mathrm{D}}$, and let λ_n^{C} and λ_n^{D} denote them, respectively.

In the special case where X contains one type of variables only (e.g., either continuous or discrete but not both), then the preceding procedure is simpler with the last term in (16) or (17) becoming null.

3. Noisy data with measurement error and misclassification

3.1. Mismeasurement models

In applications, we often collect noisy data with measurement error. To feature this, let X^* denote the observed surrogate version of X. Similar to the presentation of $X = \left(X^{\mathrm{C}\top}, X^{\mathrm{D}\top}\right)^\top$ in Section 2.1, we write $X^* = (X^{*\mathrm{C}\top}, X^{*\mathrm{D}\top})^\top$ where $X^{*\mathrm{C}}$ is the surrogate measurement of X^{C}, and $X^{*\mathrm{D}}$ is the surrogate measurement of X^{D}.

To delineate the relationship between X^* and X, we employ the widely used classical additive measurement error model [30, Ch2] to describe the relationship between $X^{*\mathrm{C}}$ and X^{C}:

$$X^{*\mathrm{C}} = X^{\mathrm{C}} + \epsilon, \tag{18}$$

where ϵ is independent of X as well as $X^{*\mathrm{D}}$, and $\epsilon \sim N(0, \Sigma_\epsilon)$ with covariance matrix Σ_ϵ.

To describe the relationship between $X^{*\mathrm{D}}$ and X^{D}, we first write the vectors of all possible values of X^{D} as $x_{(1)}, x_{(2)}, \cdots, x_{(m)}$, where m is a positive integer. Assume that

$$P\left(X^{*\mathrm{D}} = x_{(k)} \middle| X^{\mathrm{D}} = x_{(l)}, X^{\mathrm{C}}\right) = P\left(X^{*\mathrm{D}} = x_{(k)} \middle| X^{\mathrm{D}} = x_{(l)}\right). \tag{19}$$

Let $p_{kl} = P\left(X^{*\mathrm{D}} = x_{(k)} | X^{\mathrm{D}} = x_{(l)}\right)$ be the (mis)classification probability for $k, l = 1, \cdots, m$, and define the $m \times m$ (mis)classification matrix $\mathcal{P} = [p_{kl}]_{m\times m}$.

Noting that $P\left(X^{*\mathrm{D}} = x_{(k)}\right) = \sum_{l=1}^m p_{kl} P\left(X^{\mathrm{D}} = x_{(l)}\right)$ for all $k = 1, \cdots, m$, we have the matrix expression

$$\begin{pmatrix} P\left(X^{*\mathrm{D}} = x_{(1)}\right) \\ \vdots \\ P\left(X^{*\mathrm{D}} = x_{(m)}\right) \end{pmatrix} = \mathcal{P} \begin{pmatrix} P\left(X^{\mathrm{D}} = x_{(1)}\right) \\ \vdots \\ P\left(X^{\mathrm{D}} = x_{(m)}\right) \end{pmatrix}. \tag{20}$$

To ease notation, let $MC\left[\mathcal{P}\right](\cdot)$ denote the misclassification operator, and we then write (20) as

$$X^{*\mathrm{D}} = MC\left[\mathcal{P}\right](X^{\mathrm{D}}) \tag{21}$$

as a short form; we may treat (21) as a link to connect $X^{*\mathrm{D}}$ and X^{D}, or broadly as a mapping from X^{D} to $X^{*\mathrm{D}}$ through the operation $\mathcal{P}$. Expression (20) extends the misclassification operator used by [4, p.125] and [17] who considered a misclassified binary random variable only. Consistent with [4, p.125], suppose that $\mathcal{P}$ has the decomposition $\mathcal{P} = \Omega\mathcal{D}\Omega^{-1}$, where $\mathcal{D}$ is the diagonal matrix of eigenvalues of $\mathcal{P}$ and Ω is the corresponding matrix of eigenvectors.

In applications, Σ_ϵ in (18) and $\mathcal{P}$ in (20) are usually unknown. To estimate Σ_ϵ and $\mathcal{P}$, one may need auxiliary information, such as repeated measurements or validation data; discussions about this are placed in Section 7. In the absence of auxiliary information, we can employ sensitivity analyses to specify values of Σ_ϵ and $\mathcal{P}$, and examine the measurement error effects (e.g., [6]; [7]; [8]), as shown in Section 6. In the following development, Σ_ϵ and $\mathcal{P}$ are taken as given.

3.2. Impact of naive analysis

In the presence of noisy data with mismeasurement, a naive approach is to disregard the feature of noise and directly employ an available inference procedure to the available data with X^*. This approach is tempting in its easiness of implementation. However, it may yield biased results. To understand the impact of ignoring the feature of noisy data, we study the asymptotic bias of the estimator derived from the naive analysis.

For $i = 1, \cdots, n$, let $X^{*(i)}$ denote the observed surrogate for the true random vector $X^{(i)}$ for observation i. To ease the discussion, we consider the case where all variables in the model are continuous. In this case, the distribution of the graphical model is given by (1), and the conditional distribution (12) of $X_r^{(i)}$ given $X_{V\setminus\{r\}}^{(i)}$ reduces to

$$P\left(X_r^{(i)}|X_{\setminus r}^{(i)}\right) = \exp\left\{\theta_r X_r^{(i)} + X_r^{(i)} \sum_{t\in V\setminus\{r\}} \theta_{rt} X_t^{(i)} + H(X_r^{(i)}) - K^{\mathrm{C}}\left(\theta_r + \sum_{t\in V\setminus\{r\}} \theta_{rt} X_t^{(i)}\right)\right\}. \tag{22}$$

The naive analysis is based on replacing $X^{(i)}$ in (22) with the surrogate random vector $X^{*(i)}$, yielding the *naive negative log likelihood function*

$$\ell_{nv}(\theta(r)) = -\frac{1}{n}\sum_{i=1}^{n} \log\left\{P\left(X_r^{*(i)}|X_{\setminus r}^{*(i)}\right)\right\}, \tag{23}$$

where $P\left(X_r^{*(i)}|X_{\setminus r}^{*(i)}\right)$ is determined by (22) with $X^{(i)}$ replaced by $X^{*(i)}$ and $\theta(r) = \left(\theta_r, \theta_{\setminus r}^\top\right)^\top$ for $r \in V$. Then the *naive estimator* of $\theta(r)$ is obtained as

$$\widehat{\theta}_{nv}(r) = \underset{\theta(r)}{\operatorname{argmin}} \left\{\ell_{nv}(\theta(r)) + \lambda_n \left\|\theta_{\setminus r}\right\|_1\right\}, \tag{24}$$

where λ_n is the tuning parameter.

To discuss the asymptotic bias of the naive estimator $\widehat{\theta}_{nv}(r)$ for $r \in V$, we introduce some notation. Let $\theta_0(r) = \left(\theta_{0;r}, \theta_{0;\setminus r}^\top\right)^\top$ denote the true value of $\theta(r)$. For $r \in V$, define $\mathcal{Q}_r = E\left\{\left(X_{\setminus r}^{(i)\top} X_{\setminus r}^{(i)^\top}\right) K^{\mathrm{C}''}\left(\theta_{0;r} + X_{\setminus r}^{(i)\top}\theta_{0;\setminus r}\right)\right\}$ and $\mathcal{D}_r = E\left\{K^{\mathrm{C}''}\left(\theta_{0;r} + X_{\setminus r}^{(i)\top}\theta_{0;\setminus r}\right)\right\}$, where $K^{\mathrm{C}''}(\cdot)$ represents the second derivative of $K^{\mathrm{C}}(\cdot)$. Let $\Sigma_{\epsilon;\setminus r}$ be the covariance matrix Σ_ϵ with the rth row and the rth column deleted.

Theorem 3.1. *Assume that regularity conditions in Section 3.1 of [28] hold. Then for any $r \in V$, there exist constants $\widetilde{\alpha} \in (0,1)$ and $\widetilde{\rho} > 0$ such that*

$$\left\|\widehat{\theta}_{nv}(r) - \theta_0(r)\right\|_\infty$$

$$\geq \left\{\|\mathcal{Q}_r\|_\infty + \left\|\Sigma_{\epsilon;\setminus r}\mathcal{D}_r\right\|_\infty\right\}^{-1}\left(\|\nabla_\theta \ell_{nv}(\theta_0)\|_\infty - \frac{\lambda_n\widetilde{\alpha}}{4(2-\widetilde{\alpha})} - 2\lambda_n\right)$$
$$- \left\{1 + \|\mathcal{Q}_r\|_\infty^{-1}\left\|\Sigma_{\epsilon;\setminus r}\mathcal{D}_r\right\|_\infty\right\}^{-1}\widetilde{\rho}\lambda_n. \tag{25}$$

With the range of λ_n specified in [28, p.3839], Theorem 3.1 shows that the naive estimator $\widehat{\theta}_{nv}(r)$ is not close to the true parameter $\theta_0(r)$ in the infinity norm if the right-hand-side of (25) is positive. This suggests that the estimated graph generally differs from the true graph, which is also confirmed numerically by the simulation studies to be reported in Section 6. Consequently, with noisy data involving mismeasurement, it is imperative to account for the induced error effects in inferential procedures.

4. De-nosing analysis with noisy data under mixed graphical models

To accommodate effects induced from noisy data, we develop a *simulation-based neighbourhood-set likelihood* method. The basic idea is to first depict how the bias is related to the degree of mismeasurement, and then use this relation to extrapolate it to the case without mismeasurement. Such an idea is motivated by the simulation-extrapolation (SIMEX) approach proposed by [10] and the misclassification SIMEX (MC-SIMEX) method considered by [17]; those methods were developed for regression models with mismeasured covariates which are either continuous or discrete. However, our setting is pertinent to mixed continuous and discrete variables with network structures. As both continuous and discrete variables are error-contaminated, the development here is more complex in technical details. Because the proposed method concerns addressing measurement error effects as well as identifying the association structure among the variables, the establishment of its theoretical results is more challenging than those focusing on one feature only.

4.1. Estimation procedure

The simulation-based neighbourhood-set likelihood method consists of the following four steps.

Step 1: Simulation

For observation i, let $X^{*\mathrm{C}(i)}$ and $X^{*\mathrm{D}(i)}$ denote the observed surrogates of $X^{\mathrm{C}(i)}$ and $X^{\mathrm{D}(i)}$, respectively. Let B be a given positive integer and let $\mathcal{Z} = \{\zeta_0, \zeta_1, \cdots, \zeta_M\}$ be a sequence of pre-specified values with $0 = \zeta_0 < \zeta_1 < \cdots < \zeta_M$, where M is a positive integer.

For $i = 1, \cdots, n$ and $b = 1, \cdots, B$, we generate $U_b^{(i)}$ from $N(0, \Sigma_\epsilon)$ and then define

$$W_b^{\mathrm{C}(i)}(\zeta) = X^{*\mathrm{C}(i)} + \sqrt{\zeta}U_b^{(i)} \tag{26}$$

for $\zeta \in \mathcal{Z}$. For the discrete random vector $X^{*\mathrm{D}(i)}$, we generate

$$W_b^{\mathrm{D}(i)}(\zeta) = MC[\mathcal{P}^\zeta]\left(X^{*\mathrm{D}(i)}\right) \tag{27}$$

for $\zeta \in \mathcal{Z}$, where $\mathcal{P}^{\zeta} = \Omega \mathcal{D}^{\zeta} \Omega^{-1}$ and $\mathcal{D}^{\zeta}$ represents the matrix derived from $\mathcal{D}$ by replacing its diagonal elements, say d_{jj}, with d_{jj}^{ζ}.

For $b = 1, \cdots, B$, $\zeta \in \mathcal{Z}$, and $i = 1, \cdots, n$, let $W_b^{(i)}(\zeta) = \left(W_b^{\mathrm{C}(i)\top}(\zeta), W_b^{\mathrm{D}(i)\top}(\zeta)\right)^{\top}$ and we call it *working data*.

Step 2: Estimation

For $r \in V^{\mathrm{C}}$ and $r' \in V^{\mathrm{D}}$, let $\ell_{b,\zeta}^{\mathrm{C}}\left(\theta^{\mathrm{C}}(r)\right)$ and $\ell_{b,\zeta}^{\mathrm{D}}\left(\theta^{\mathrm{D}}(r')\right)$ be determined by (14) and (15), respectively, with $X^{(i)}$ replaced by $W_b^{(i)}(\zeta)$. Then for each $\zeta \in \mathcal{Z}$ and $b = 1, \cdots, B$, compute

$$\widehat{\theta}^{\mathrm{C}}(r; \zeta, b) = \underset{\theta^{\mathrm{C}}(r)}{\operatorname{argmin}} \left\{ \ell_{b,\zeta}^{\mathrm{C}}\left(\theta^{\mathrm{C}}(r)\right) + \lambda_n^{\mathrm{C}} \left(\left\| \theta_{\backslash r}^{\mathrm{C}} \right\|_1 + \| \theta_r^{\mathrm{CD}} \|_1 \right) \right\} \tag{28}$$

and

$$\widehat{\theta}^{\mathrm{D}}(r'; \zeta, b) = \underset{\widehat{\theta}^{\mathrm{D}}(r')}{\operatorname{argmin}} \left\{ \ell_{b,\zeta}^{\mathrm{D}}\left(\theta^{\mathrm{D}}(r')\right) + \lambda_n^{\mathrm{D}} \left(\left\| \theta_{\backslash r'}^{\mathrm{D}} \right\|_1 + \| \theta_{r'}^{\mathrm{DC}} \|_1 \right) \right\}, \tag{29}$$

where λ_n^{C} and λ_n^{D} are tuning parameters. Next, we calculate

$$\widehat{\theta}^{\mathrm{C}}(r; \zeta) = \frac{1}{B} \sum_{b=1}^{B} \widehat{\theta}^{\mathrm{C}}(r; \zeta, b) \quad \text{and} \quad \widehat{\theta}^{\mathrm{D}}(r'; \zeta) = \frac{1}{B} \sum_{b=1}^{B} \widehat{\theta}^{\mathrm{D}}(r'; \zeta, b). \tag{30}$$

Step 3: Extrapolation

Grouping the estimates obtained from (30), we obtain two sequences $\mathbb{S}_r^{\mathrm{C}} = \left\{ \left(\zeta, \widehat{\theta}^{\mathrm{C}}\left(r; \zeta\right) \right) : \zeta \in \mathcal{Z} \right\}$ and $\mathbb{S}_{r'}^{\mathrm{D}} = \left\{ \left(\zeta, \widehat{\theta}^{\mathrm{D}}\left(r'; \zeta\right) \right) : \zeta \in \mathcal{Z} \right\}$ for $r \in V^{\mathrm{C}}$ and $r' \in V^{\mathrm{D}}$. Then we respectively regress $\widehat{\theta}^{\mathrm{C}}\left(r; \zeta\right)$ and $\widehat{\theta}^{\mathrm{D}}\left(r'; \zeta\right)$ on ζ by fitting models

$$\widehat{\theta}^{\mathrm{C}}\left(r; \zeta\right) = \mathcal{G}^{\mathrm{C}}\left(\zeta, \Gamma^{\mathrm{C}}\right) + \delta^{\mathrm{C}} \quad \text{and} \quad \widehat{\theta}^{\mathrm{D}}\left(r'; \zeta\right) = \mathcal{G}^{\mathrm{D}}\left(\zeta, \Gamma^{\mathrm{D}}\right) + \delta^{\mathrm{D}} \tag{31}$$

to the sequences $\mathbb{S}_r^{\mathrm{C}}$ and $\mathbb{S}_{r'}^{\mathrm{D}}$, where $\mathcal{G}^{\mathrm{C}}(\cdot, \cdot)$ and $\mathcal{G}^{\mathrm{D}}(\cdot, \cdot)$ are use-specified regression functions (such as linear or quadratic functions), Γ^{C} and Γ^{D} are the associated parameter vectors, and δ^{C} and δ^{D} represent the noise terms.

Parameters Γ^{C} and Γ^{D} can be estimated by applying the least squares method to the sequences $\mathbb{S}_r^{\mathrm{C}}$ and $\mathbb{S}_{r'}^{\mathrm{D}}$; let $\widehat{\Gamma}^{\mathrm{C}}$ and $\widehat{\Gamma}^{\mathrm{D}}$ denote the resulting estimates of Γ^{C} and Γ^{D}, respectively. Next, we extrapolate models in (31) by letting $\zeta = -1$ and calculate

$$\widehat{\theta}^{\mathrm{C}}(r) \triangleq \mathcal{G}^{\mathrm{C}}(-1, \widehat{\Gamma}^{\mathrm{C}}) \quad \text{and} \quad \widehat{\theta}^{\mathrm{D}}(r') \triangleq \mathcal{G}^{\mathrm{D}}(-1, \widehat{\Gamma}^{\mathrm{D}}). \tag{32}$$

Step 4: Graph Assembling

Corresponding to expression (6), we write $\widehat{\theta}^{\mathrm{C}}(r) = \left(\widehat{\theta}_r^{\mathrm{C}}, \widehat{\theta}_{\backslash r}^{\mathrm{C}\top}, \widehat{\theta}_r^{\mathrm{CD}\top}\right)^{\top}$ and $\widehat{\theta}^{\mathrm{D}}(r') = \left(\widehat{\theta}_{r'}^{\mathrm{D}}, \widehat{\theta}_{\backslash r'}^{\mathrm{D}\top}, \widehat{\theta}_{r'}^{\mathrm{DC}\top}\right)^{\top}$ as the estimators of $\theta^{\mathrm{C}}(r)$ and $\theta^{\mathrm{D}}(r')$ for $r \in$

V^{C} and $r' \in V^{\mathrm{D}}$, respectively. Then the homogeneous neighbourhoods $\mathcal{N}^{\mathrm{C}}(r)$ and $\mathcal{N}^{\mathrm{D}}(r')$ for $r \in V^{\mathrm{C}}$ and $r' \in V^{\mathrm{D}}$ are estimated by $\widehat{\mathcal{N}}^{\mathrm{C}}(r) = \left\{t \in V^{\mathrm{C}} \setminus \{r\} : \widehat{\theta}^{\mathrm{C}}_{rt} \neq 0\right\}$ and $\widehat{\mathcal{N}}^{\mathrm{D}}(r') = \left\{t' \in V^{\mathrm{D}} \setminus \{r'\} : \widehat{\theta}^{\mathrm{D}}_{r't'} \neq 0\right\}$, respectively, and the heterogeneous neighbourhoods $\mathcal{N}^{\mathrm{CD}}(r)$ and $\mathcal{N}^{\mathrm{DC}}(r')$ are estimated by $\widehat{\mathcal{N}}^{\mathrm{CD}}(r) = \left\{t' \in V^{\mathrm{D}} : \widehat{\theta}^{\mathrm{CD}}_{rt'} \neq 0\right\}$ and $\widehat{\mathcal{N}}_{\mathrm{DC}}(r') = \left\{t \in V^{\mathrm{C}} : \widehat{\theta}^{\mathrm{CD}}_{tr'} \neq 0\right\}$, respectively.

To obtain an estimator of Θ, one may consider to repeat the same procedures for $r \in V^{\mathrm{C}}$ and $r' \in V^{\mathrm{D}}$, which, however, is flawed since Θ is a symmetric matrix with $\theta^{\mathrm{C}}_{rt} = \theta^{\mathrm{C}}_{tr}$ and $\theta^{\mathrm{D}}_{rt} = \theta^{\mathrm{D}}_{tr}$ for $t \neq r$, but their estimators do not necessarily possess this symmetry property. As a remedy, we apply the AND rule proposed by [19]. That is, for the nodes related to continuous random variables, the AND rule declares that (r, t) belongs to the estimated edge set $\widehat{E}^{\mathrm{C}}$ if both $r \in \widehat{\mathcal{N}}^{\mathrm{C}}(t)$ and $t \in \widehat{\mathcal{N}}^{\mathrm{C}}(r)$ hold; the similar rule applies to the parameters $\theta^{\mathrm{D}}_{r't'}$ related to the discrete random variables and the parameters $\theta^{\mathrm{CD}}_{rt'}$ related to the mixture of random variables.

4.2. Remarks of the implementation procedure

The implementation procedure described in Section 4.1 applies for any given tuning parameters λ^{C}_n and λ^{D}_n in (28) and (29). One may wonder how to choose λ^{C}_n and λ^{D}_n to obtain sensible results. Here we use the Bayesian Information Criterion (BIC) (e.g., [7]; [27]; [34]) to select the tuning parameters. Specifically, for $r \in V^{\mathrm{C}}$ and $r' \in V^{\mathrm{D}}$, we let $\widehat{\theta}^{\mathrm{C}}(r; \zeta, b, \lambda^{\mathrm{C}}_n)$ and $\widehat{\theta}^{\mathrm{D}}(r'; \zeta, b, \lambda^{\mathrm{D}}_n)$ respectively denote the estimator obtained from (28) and (29) by spelling out their dependence on the tuning parameters. For given b and ζ, define

$$\mathrm{BIC}\left(\lambda^{\mathrm{C}}_n\right) = 2n\ell^{\mathrm{C}}_{b,\zeta}\left(\widehat{\theta}^{\mathrm{C}}(r; \zeta, b, \lambda^{\mathrm{C}}_n)\right) + (\log n) \times \mathrm{df}\left(\widehat{\theta}^{\mathrm{C}}(r; \zeta, b, \lambda^{\mathrm{C}}_n)\right) \tag{33}$$

and

$$\mathrm{BIC}\left(\lambda^{\mathrm{D}}_n\right) = 2n\ell^{\mathrm{D}}_{b,\zeta}\left(\widehat{\theta}^{\mathrm{D}}(r'; \zeta, b, \lambda^{\mathrm{D}}_n)\right) + (\log n) \times \mathrm{df}\left(\widehat{\theta}^{\mathrm{D}}(r'; \zeta, b, \lambda^{\mathrm{D}}_n)\right), \tag{34}$$

where notation $\mathrm{df}(a)$ represents the number of nonzero elements in a vector a.

Consider a grid for λ^{C}_n and λ^{D}_n, denoted Λ^{C} and Λ^{D}, respectively. The optimal tuning parameters λ^{C}_n and λ^{D}_n, denoted by $\widehat{\lambda}^{\mathrm{C}}_n$ and $\widehat{\lambda}^{\mathrm{D}}_n$, are determined by minimizing (33) and (34) over Λ^{C} and Λ^{D}, respectively. That is,

$$\widehat{\lambda}^{\mathrm{C}}_n = \underset{\lambda^{\mathrm{C}}_n \in \Lambda^{\mathrm{C}}}{\operatorname{argmin}} \mathrm{BIC}\left(\lambda^{\mathrm{C}}_n\right) \quad \text{and} \quad \widehat{\lambda}^{\mathrm{D}}_n = \underset{\lambda^{\mathrm{D}}_n \in \Lambda^{\mathrm{D}}}{\operatorname{argmin}} \mathrm{BIC}\left(\lambda^{\mathrm{D}}_n\right).$$

Consequently, with given b and ζ, the estimators of $\theta^{\mathrm{C}}(r)$ and $\theta^{\mathrm{D}}(r')$ are determined by $\widehat{\theta}^{\mathrm{C}}(r; \zeta, b) \triangleq \widehat{\theta}^{\mathrm{C}}(r; \zeta, b, \widehat{\lambda}^{\mathrm{C}}_n)$ and $\widehat{\theta}^{\mathrm{D}}(r'; \zeta, b) \triangleq \widehat{\theta}^{\mathrm{D}}(r'; \zeta, b, \widehat{\lambda}^{\mathrm{D}}_n)$ for $r \in V^{\mathrm{C}}$ and $r' \in V^{\mathrm{D}}$, respectively.

We conclude this subsection with remarks. Different from the conventional graphical model which focuses on either continuous or discrete random variables, we consider settings with both continuous and discrete random variables accommodated simultaneously and allow error-prone variables to be a mix of continuous and discrete variables.

To account for the effects of mismeasurement in the variables, we employ a simulation-based strategy by combining the simulation-extrapolation (SIMEX) method [10] and the MC-SIMEX method [17] which are separately developed for handling error-contaminated continuous and discrete variables. The basic idea of this strategy is to delineate how estimation biases for the model parameters may depend on varying magnitudes of mismeasurement by considering a series of artificial scenarios. Its implementation hings on three steps. The first simulation step creates artificial settings with different degrees of mismeasurement in the variables. The second estimation step quantifies induced biases for varying magnitudes of mismeasurement using a usual method developed for error-free settings. The third extrapolation step first traces the pattern of the estimation biases and then extrapolates it to the setting without measurement error which is reflected by setting $\zeta = -1$.

To see why $\zeta = -1$ represents the error-free setting, one may, for each i and b, re-write (26) and (27) as

$$W_b^{\mathrm{C}(i)}(\zeta) \sim N\left(X^{\mathrm{C}(i)}, (1+\zeta)\Sigma_\epsilon\right) \quad \text{and} \quad W_b^{\mathrm{D}(i)}(\zeta) = MC\left[\mathcal{P}^{1+\zeta}\right]\left(X^{\mathrm{D}(i)}\right) \tag{35}$$

for $\zeta \in \mathcal{Z}$. Different values of ζ reflect different degrees of mismeasurement in the artificially generated surrogates $W_b^{\mathrm{C}(i)}(\zeta)$ and $W_b^{\mathrm{D}(i)}(\zeta)$. With $\zeta = 0$, $W_b^{\mathrm{C}(i)}(\zeta)$ and $W_b^{\mathrm{D}(i)}(\zeta)$ recover the actually collected surrogates $X^{*\mathrm{C}(i)}$ and $X^{*\mathrm{D}(i)}$. With a positive and increasing ζ, $W_b^{\mathrm{C}(i)}(\zeta)$ and $W_b^{\mathrm{D}(i)}(\zeta)$ incur an increasing amount of mismeasurement. While (35) is defined for $\zeta \in \mathcal{Z}$ which are all nonnegative, if one extrapolates its value to be -1, then $\zeta = -1$ makes the right-hand side of the two expressions in (35) represent the true variables $X^{\mathrm{C}(i)}$ and $X^{\mathrm{D}(i)}$, respectively, the ideal situation without mismeasurement. Since the value $\zeta = -1$ is not in $\mathcal{Z}$, calculating the fitted values with $\zeta = -1$ in (31) is not called determining *predicted* values as usual but is termed as *extrapolating*. To have an intuitive illustration of the meaning corresponding to $\zeta = -1$, one may consider a simple linear regression response model with an additive measurement error model for the error-prone predictor. For details, see [30, pp.63-64].

The idea of the estimation steps is conceptually straightforward, yet the implementation can be computationally demanding, which typically depends on the specification of B and $\mathcal{Z}$ in Step 1. Although taking a larger value of B and more points in $\mathcal{Z}$ may help improve estimation results, it will greatly increase computation time. The choice of B and $\mathcal{Z}$ is driven by the balance of the affordability of computation resources and accuracy of results. While B and $\mathcal{Z}$ are not uniquely specified, commonly, B is set as a value between 100 and 500, $\mathcal{Z}$ is taken as a collection of M points that equally cut the interval $[0, 1]$ or $[0, 2]$ with M set as 5 or 10. Further, the implementation requires the specification

of the extrapolation function in Step 3, and numerical experience suggests that the quadratic extrapolation function tends to perform well in various settings. Detailed comments on these are available in [4, p.106] and [30, p.64].

5. Theory

In this section we establish theoretical results for the proposed estimators. First, we introduce additional notation based on those defined in Section 2.2 and the negative log-likelihood functions in Section 4.1. For $r \in V^{\mathrm{C}}$ and $r' \in V^{\mathrm{D}}$, define $Q^{\mathrm{C}}_{\zeta} = \nabla^2_{\theta^{\mathrm{C}}(r)} \ell^{\mathrm{C}}_{b,\zeta}(\theta^{\mathrm{C}}_0(r))$ and $Q^{\mathrm{D}}_{\zeta} = \nabla^2_{\theta^{\mathrm{D}}(r')} \ell^{\mathrm{D}}_{b,\zeta}(\theta^{\mathrm{D}}_0(r'))$, and based on (7) and (8), we write them as the block matrices

$$Q^{\mathrm{C}}_{\zeta} = \begin{pmatrix} Q_{\zeta,\mathcal{S}^{\mathrm{C}}_r \mathcal{S}^{\mathrm{C}}_r} & Q_{\zeta,\mathcal{S}^{\mathrm{C}}_r \overline{\mathcal{S}}^{\mathrm{C}}_r} \\ Q_{\zeta,\overline{\mathcal{S}}^{\mathrm{C}}_r \mathcal{S}^{\mathrm{C}}_r} & Q_{\zeta,\overline{\mathcal{S}}^{\mathrm{C}}_r \overline{\mathcal{S}}^{\mathrm{C}}_r} \end{pmatrix} \quad \text{and} \quad Q^{\mathrm{D}}_{\zeta} = \begin{pmatrix} Q_{\zeta,\mathcal{S}^{\mathrm{D}}_{r'} \mathcal{S}^{\mathrm{D}}_{r'}} & Q_{\zeta,\mathcal{S}^{\mathrm{D}}_{r'} \overline{\mathcal{S}}^{\mathrm{D}}_{r'}} \\ Q_{\zeta,\overline{\mathcal{S}}^{\mathrm{D}}_{r'} \mathcal{S}^{\mathrm{D}}_{r'}} & Q_{\zeta,\overline{\mathcal{S}}^{\mathrm{D}}_{r'} \overline{\mathcal{S}}^{\mathrm{D}}_{r'}} \end{pmatrix},$$

where

$$\begin{aligned} Q_{\zeta,\mathcal{S}^{\mathrm{C}}_r \overline{\mathcal{S}}^{\mathrm{C}}_r} &= \nabla_{\theta_{\overline{\mathcal{S}}^{\mathrm{C}}_r}} \left\{ \nabla_{\theta_{\mathcal{S}^{\mathrm{C}}_r}(r)} \ell^{\mathrm{C}}_{b,\zeta}(\theta^{\mathrm{C}}_0(r)) \right\}, \quad Q_{\zeta,\overline{\mathcal{S}}^{\mathrm{C}}_r \mathcal{S}^{\mathrm{C}}_r} = Q^{\top}_{\zeta,\mathcal{S}^{\mathrm{C}}_r \overline{\mathcal{S}}^{\mathrm{C}}_r}, \\ Q_{\zeta,\mathcal{S}^{\mathrm{C}}_r \mathcal{S}^{\mathrm{C}}_r} &= \nabla^2_{\theta_{\mathcal{S}^{\mathrm{C}}_r}} \ell^{\mathrm{C}}_{b,\zeta}(\theta^{\mathrm{C}}_0(r)), \quad \text{and} \quad Q_{\zeta,\overline{\mathcal{S}}^{\mathrm{C}}_r \overline{\mathcal{S}}^{\mathrm{C}}_r} = \nabla^2_{\theta_{\overline{\mathcal{S}}^{\mathrm{C}}_r}} \ell^{\mathrm{C}}_{b,\zeta}(\theta^{\mathrm{C}}_0(r)); \end{aligned}$$

and

$$\begin{aligned} Q_{\zeta,\mathcal{S}^{\mathrm{D}}_{r'} \mathcal{S}^{\mathrm{D}}_{r'}} &= \nabla^2_{\theta_{\mathcal{S}^{\mathrm{D}}_{r'}}} \ell^{\mathrm{D}}_{b,\zeta}(\theta^{\mathrm{D}}_0(r)), \quad Q_{\zeta,\overline{\mathcal{S}}^{\mathrm{D}}_{r'} \overline{\mathcal{S}}^{\mathrm{D}}_{r'}} = \nabla^2_{\theta_{\overline{\mathcal{S}}^{\mathrm{D}}_{r'}}} \ell^{\mathrm{D}}_{b,\zeta}(\theta^{\mathrm{D}}_0(r')) \\ Q_{\zeta,\mathcal{S}^{\mathrm{D}}_{r'} \overline{\mathcal{S}}^{\mathrm{D}}_{r'}} &= \nabla_{\theta_{\overline{\mathcal{S}}^{\mathrm{D}}_{r'}}} \left\{ \nabla_{\theta_{\mathcal{S}^{\mathrm{D}}_{r'}}(r')} \ell^{\mathrm{D}}_{b,\zeta}(\theta^{\mathrm{D}}_0(r')) \right\}, \quad \text{and} \quad Q_{\zeta,\overline{\mathcal{S}}^{\mathrm{D}}_{r'} \mathcal{S}^{\mathrm{D}}_{r'}} = Q^{\top}_{\zeta,\mathcal{S}^{\mathrm{D}}_{r'} \overline{\mathcal{S}}^{\mathrm{D}}_{r'}}. \end{aligned}$$

Corresponding to (11), for $r \in V^{\mathrm{C}}$ and $r' \in V^{\mathrm{D}}$, we write the estimators $\widehat{\theta}^{\mathrm{C}}(r)$ and $\widehat{\theta}^{\mathrm{D}}(r')$ obtained from Step 4 of Section 4.1 as

$$\widehat{\theta}^{\mathrm{C}}(r) = \left(\widehat{\theta}^{\mathrm{C}}_r, \widehat{\theta}^{\top}_{\mathcal{S}^{\mathrm{C}}_r}, \widehat{\theta}^{\top}_{\overline{\mathcal{S}}^{\mathrm{C}}_r} \right)^{\top} \quad \text{and} \quad \widehat{\theta}^{\mathrm{D}}(r') = \left(\widehat{\theta}^{\mathrm{D}}_{r'}, \widehat{\theta}^{\top}_{\mathcal{S}^{\mathrm{D}}_{r'}}, \widehat{\theta}^{\top}_{\overline{\mathcal{S}}^{\mathrm{D}}_{r'}} \right)^{\top},$$

respectively, where

$$\widehat{\theta}_{\mathcal{S}^{\mathrm{C}}_r} = \left(\widehat{\theta}^{\mathrm{C}}_{rt}, \widehat{\theta}^{\mathrm{CD}}_{rt'} : t \in \mathcal{S}^{\mathrm{C}}_r(V^{\mathrm{C}}) \text{ and } t' \in \mathcal{S}^{\mathrm{C}}_r(V^{\mathrm{D}}) \right)^{\top}$$

and

$$\widehat{\theta}_{\mathcal{S}^{\mathrm{D}}_{r'}} = \left(\widehat{\theta}^{\mathrm{D}}_{r't'}, \widehat{\theta}^{\mathrm{DC}}_{tr'} : t \in \mathcal{S}^{\mathrm{D}}_{r'}(V^{\mathrm{C}}) \text{ and } t' \in \mathcal{S}^{\mathrm{D}}_{r'}(V^{\mathrm{D}}) \right)^{\top}$$

are the estimators for (9) and (10), respectively.

5.1. Regularity conditions

To establish the asymptotic results of the developed method, the following conditions are required.

(A1) There exists a positive value $\alpha \in (0,1)$ such that

$$\left\| Q_{\zeta,\overline{\mathcal{S}}_r^{\mathrm{C}}\mathcal{S}_r^{\mathrm{C}}} Q_{\zeta,\mathcal{S}_r^{\mathrm{C}}\mathcal{S}_r^{\mathrm{C}}}^{-1} \right\|_\infty \leq 1-\alpha$$

and

$$\left\| Q_{\zeta,\overline{\mathcal{S}}_{r'}^{\mathrm{D}}\mathcal{S}_{r'}^{\mathrm{D}}} Q_{\zeta,\mathcal{S}_{r'}^{\mathrm{D}}\mathcal{S}_{r'}^{\mathrm{D}}}^{-1} \right\|_\infty \leq 1-\alpha$$

for $\zeta \in \mathcal{Z}$, $r \in V^{\mathrm{C}}$, and $r' \in V^{\mathrm{D}}$.

(A2) There exists $\rho_2 < \infty$ such that

$$\Lambda_{\max}\left(\sum_{i=1}^{n} W_b^{(i)}(\zeta) W_b^{(i)\top}(\zeta)\right) < \rho_2$$

for all $b = 1, \cdots, B$ and $\zeta \in \mathcal{Z}$.

(A3) Functions $K^{\mathrm{C}}(\cdot)$ and $K^{\mathrm{D}}(\cdot)$ are third-order differentiable, and there exist positive η_1 and η_2 such that

$$\left|K^{\mathrm{C}''}(y)\right| < \eta_1,\ \left|K^{\mathrm{D}''}(y)\right| < \eta_1,\ \left|K^{\mathrm{C}'''}(y)\right| < \eta_2,\ \text{and}\ \left|K^{\mathrm{D}'''}(y)\right| < \eta_2$$

for every y.

(A4) The exact extrapolation functions in Step 3 of Section 4.1 are assumed to be known.

(A5) There exist constants κ_1 and $\kappa_2 > 0$ such that

$$E\left(X_r^{\mathrm{C}}\right) < \kappa_1 \quad \text{and} \quad E\left\{\left(X_r^{\mathrm{C}}\right)^2\right\} < \kappa_2$$

for all $r = 1, \cdots, p^{\mathrm{C}}$.

Assumptions (A1) and (A2), called *mutual incoherence* and *dependency condition*, respectively, (e.g., [19]), are frequently assumed in the literature (e.g., [5]; [19]; [20]; [28]). Assumption (A3) describes boundness and differentiation conditions for $K^{\mathrm{C}}(\cdot)$ and $K^{\mathrm{D}}(\cdot)$. (A4) is a regular condition required to establish the consistency for the SIMEX estimators (e.g., [4]). (A5) requires the first and second moments of the continuous random variables to be bounded (e.g., [5]; [28]), and this condition also implies the boundness of the moments for working data. To see this, let $\Sigma_{\epsilon;r,r}$ denote entry (r,r) of Σ_ϵ. For any $b = 1, \cdots, B$ and $\zeta \in \mathcal{Z}$, $E\left\{W_{b,r}^{\mathrm{C}}(\zeta)\right\} = E\left\{E\left(W_{b,r}^{\mathrm{C}}(\zeta)|X_r^{\mathrm{C}}\right)\right\} = E\left(X_r^{\mathrm{C}}\right)$ and $E\left[\left\{W_{b,r}^{\mathrm{C}}(\zeta)\right\}^2\right] = E\left(E\left[\left\{W_{b,r}^{\mathrm{C}}(\zeta)\right\}^2 \middle| X_r^{\mathrm{C}}\right]\right) = E\left\{\left(X_r^{\mathrm{C}}\right)^2\right\} + (1+\zeta)\Sigma_{\epsilon;r,r}$, and thus, $E\left[\left\{W_{b,r}^{\mathrm{C}}(\zeta)\right\}^2\right] < \kappa_2 + (1+\zeta)\Sigma_{\epsilon;r,r}$.

5.2. Theoretical results

We now establish theoretical results for the estimators proposed in Section 4.1. For $\zeta \in \mathcal{Z}$, let $\rho_\zeta > 0$ denote the smallest eigenvalue for all $Q_{\zeta,\mathcal{S}_r^{\mathrm{C}}\mathcal{S}_r^{\mathrm{C}}}$ and $Q_{\zeta,\mathcal{S}_{r'}^{\mathrm{D}}\mathcal{S}_{r'}^{\mathrm{D}}}$ with $r \in V^{\mathrm{C}}$ and $r' \in V^{\mathrm{D}}$. Define $\rho \triangleq \min\{\rho_\zeta : \zeta \in \mathcal{Z}\}$.

Theorem 5.1 (Sparsity recovery)**.** *Assume regularity conditions in Section 5.1. Suppose that tuning parameters λ_n^C and λ_n^D satisfy*

$$\sqrt{\frac{\eta_1 \kappa_2^* \log p}{n}} \left(\frac{2-\alpha}{\alpha}\right) < \lambda_n^C < \frac{\rho^2}{\eta_1 \rho_2^* d^C} \tag{36}$$

and

$$\sqrt{\frac{\eta_1 \kappa_2^* \log p}{n}} \left(\frac{2-\alpha}{\alpha}\right) < \lambda_n^D < \frac{\rho^2}{\eta_1 \rho_2^* d^D}, \tag{37}$$

where $\rho_2^* = 288\rho_2$ and $\kappa_2^* = 64\kappa_2$, with ρ_2 and κ_2 being defined in Conditions (A2) and (A5), respectively. Then with probability greater than $1 - \tau_n$ where $\tau_n = \left[c_1\{\max(n,p)\}^{-2} + \exp(-c_2 n)\right]$ for some positive constants c_1 and c_2,

$$\widehat{\mathcal{N}}^C(r) = \mathcal{N}^C(r) \quad \text{and} \quad \widehat{\mathcal{N}}^{CD}(r) = \mathcal{N}^{CD}(r) \tag{38}$$

for $r \in V^C$, and

$$\widehat{\mathcal{N}}^D(r') = \mathcal{N}^D(r') \quad \text{and} \quad \widehat{\mathcal{N}}^{DC}(r') = \mathcal{N}^{DC}(r') \tag{39}$$

for $r' \in V^D$.

This theorem shows that with a given sample of size n, the estimated graph can be ensured to equal the true graphical structure with a probability greater than 1 minus a value related to $\{\max(n,p)\}^{-2}$ and $\exp(-n)$ if tuning parameters λ_n^{C} and λ_n^{D} are properly chosen; when the sample size n approaches infinity, the probability that the estimated graph equals the true graph approaches 1. Both tuning parameters λ_n^{C} and λ_n^{D} are required to be lower bounded by the same value proportional to $n^{-1/2}$. Such a lower bound enables the probability of having the desired result (i.e., (38) and (39)) to be high, as indicated by the form of τ_n, and shown in Lemma A.2 and (C.11) in Step 1 appearing in Appendix C. Thereby, the desired result holds with probability approaching one as n goes to infinity (e.g., [20, p.1299]). Meanwhile, both λ_n^{C} and λ_n^{D} need to be upper bounded by functions of the eigenvalues of $Q_{\zeta,\mathcal{S}_r^{\mathrm{C}}\mathcal{S}_r^{\mathrm{C}}}$ and $Q_{\zeta,\mathcal{S}_{r'}^{\mathrm{D}}\mathcal{S}_{r'}^{\mathrm{D}}}$, which ensures the derivations for (38) and (39), as shown in (C.10) and (C.11) in Appendix C.

Theorem 5.2. *For $r \in V^C$, let $\theta^C_{0;\mathcal{S}_r^C}(r) = \left(\theta^C_{0;r}, \theta^\top_{0;\mathcal{S}_r^C}\right)^\top$ denote the true value for the subvector of nonzero parameters associated with r, and let $\widehat{\theta}^C_{\mathcal{S}_r^C}(r) = \left(\widehat{\theta}^C_r, \widehat{\theta}^\top_{\mathcal{S}_r^C}\right)^\top$ denote its estimator. For $r' \in V^D$, let $\theta^D_{0;\mathcal{S}_{r'}^D}(r') = \left(\theta^D_{0;r'}, \theta^\top_{0;\mathcal{S}_{r'}^D}\right)^\top$*

denote the true value for the subvector of nonzero parameters associated with r'*, and let* $\widehat{\theta}^{D}_{\mathcal{S}^{D}_{r'}}(r') = \left(\widehat{\theta}^{D}_{r'}, \widehat{\theta}^{\top}_{\mathcal{S}^{D}_{r'}}\right)^{\top}$ *denote its estimator. Assume regularity conditions in Theorem 5.1. Then with probability greater than* $1 - \{\max(n,p)\}^{-2}$*, the following results hold for* $r \in V^{C}$ *and* $r' \in V^{D}$*:*

(a) Sign recovery:

$$sign\left(\widehat{\theta}^{C}_{\mathcal{S}^{C}_{r}}(r)\right) = sign\left(\theta^{C}_{0;\mathcal{S}^{C}_{r}}(r)\right) \quad and \quad sign\left(\widehat{\theta}^{D}_{\mathcal{S}^{D}_{r'}}(r')\right) = sign\left(\theta^{D}_{0;\mathcal{S}^{D}_{r'}}(r')\right).$$

(b) Boundness of the estimators:

$$\left\|\widehat{\theta}^{C}_{\mathcal{S}^{C}_{r}}(r) - \theta^{C}_{0;\mathcal{S}^{C}_{r}}(r)\right\|_{\infty} \leq \frac{6\sqrt{d^{C}_{r}}\lambda^{C}_{n}}{\rho}$$

and

$$\left\|\widehat{\theta}^{D}_{\mathcal{S}^{D}_{r'}}(r') - \theta^{D}_{0;\mathcal{S}^{D}_{r'}}(r')\right\|_{\infty} \leq \frac{6\sqrt{d^{D}_{r'}}\lambda^{D}_{n}}{\rho}.$$

Theorem 5.2 describes the finite sample performance of the proposed method with probability higher than $1-\{\max(n,p)\}^{-2}$. As n goes to infinity, the results hold with probability approaching one. Theorem 5.2 (a) says that the sign of the estimators is identical to that of the true parameter values. For $r \in V^{\mathrm{C}}$ and $r' \in V^{\mathrm{D}}$, regarding the subvectors of nonzero parameters $\theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_{r}}(r)$ and $\theta^{\mathrm{D}}_{0;\mathcal{S}^{\mathrm{D}}_{r'}}(r')$ in $\theta^{\mathrm{C}}_{0}(r)$ and $\theta^{\mathrm{D}}_{0}(r')$, respectively, Theorem 5.2 (b) offers upper bounds for the differences between their estimators and their true values. These bounds suggest that the discrepancies of the proposed estimators from their target parameters are not unlimited but are bounded. We stress that the dependence of the upper bounds on the magnitude of mismeasurement in X^{C} and X^{D} is tacitly reflected by ρ. Being the smallest eigenvalue of $Q_{\zeta,\mathcal{S}^{\mathrm{C}}_{r}\mathcal{S}^{\mathrm{C}}_{r}}$ and $Q_{\zeta,\mathcal{S}^{\mathrm{D}}_{r'}\mathcal{S}^{\mathrm{D}}_{r'}}$ for all $\zeta \in \mathcal{Z}$, ρ implicitly depends on both the original degrees of mismeasurement in (18) and (21) and those artificially generated magnitudes of mismeasurements in (26) and (27). As the dependence of ρ involves the latter ones, the identified upper bounds in Theorem 5.2 (b) are not sharp. This latter dependence also reflects the use of working data (26) and (27) when implementing the estimation procedure. While the inequalities in Theorem 5.2 (b) do not show analytical forms in terms of Σ_{ϵ} or $\mathcal{P}$, the impact of measurement error can be evaluated using numerical studies, as shown in Section 6.

When components of X are all continuous (or discrete), i.e., $X = X^{\mathrm{C}}$ (or $X = X^{\mathrm{D}}$), results similar to Theorems 5.1 and 5.2 still hold for the corresponding estimators.

6. Numerical studies

In this section, we conduct numerical studies to assess the performance of the proposed estimators for a variety of settings. We first design the simulation

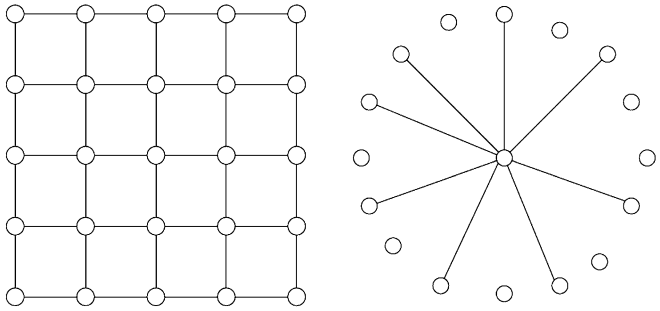

FIG 1. *The left-hand-side structure is a Lattice and the right-hand-side structure is a Hub.*

settings and then present the simulation results. Finally, the proposed method is implemented to analyze a real dataset, where we set $B = 500$ and $\mathcal{Z} = \{0, 0.25, 0.50, 0.75, 1, 1.25, 1.50, 1.75, 2\}$.

6.1. Model settings

Let Θ_0 be the $p \times p$ matrix which is specified to have the network structure, a lattice or a hub structure, as shown in Figure 1. Let X denote the p-dimensional random vector which follows the exponential family distribution (4), where the continuous random vector X^{C} assumes the structure (2), and the discrete random vector X^{D} assumes the form (3).

For the measurement error process, we consider the following three scenarios:

Scenario 1: *Only continuous variables are subject to measurement error*
In this scenario, all error-prone random variables are continuous, i.e., $X = X^{\mathrm{C}}$, and they assume the classical additive measurement error model (18), where ϵ is independent of X^{C}, $\epsilon \sim N(0, \Sigma_\epsilon)$, and Σ_ϵ is a $p \times p$ diagonal matrix with entries σ_ϵ^2, with σ_ϵ^2 set as 0.15, 0.5 or 0.75 to reflect increasing degrees of measurement error.

Scenario 2: *Only binary variables are subject to misclassification*
In this scenario, all the error-contaminated random variables are considered to be binary, taking value 1 or -1, i.e., $X = X^{\mathrm{D}}$. In contrast to the misclassification probabilities defined in Section 3.1, we consider that

$$p_{ll} = P\left(X^{*\mathrm{D}} = x_{(l)} | X^{\mathrm{D}} = x_{(l)}\right) \tag{40}$$

assumes a common value, say π, for $l = 1, \cdots, m$, where $m = 2^p$, representing the cardinality of the set $\{-1, 1\}^p$. Thus, the misclassification

matrix is the $m \times m$ matrix

$$\mathcal{P} = \begin{bmatrix} \pi & 1-\pi & 0 & 0 & \cdots & 0 & 0 & 0 \\ \frac{1}{2}(1-\pi) & \pi & \frac{1}{2}(1-\pi) & 0 & \cdots & 0 & 0 & 0 \\ 0 & \frac{1}{2}(1-\pi) & \pi & \frac{1}{2}(1-\pi) & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1-\pi & \pi \end{bmatrix},$$

where we set $\pi = 0.7, 0.8$, or 0.9 to reflect different degrees of misclassification.

Scenario 3: *Both measurement error and misclassification exist*

In this scenario, we examine the case where both continuous and discrete random variables are subject to mismeasurement by combining Scenarios 1 and 2 with additional assumptions that ϵ in (18) is independent of $X^{*\mathrm{D}}$. Consistent with the notation in Section 2.1, let $X = \left(X^{\mathrm{C}\top}, X^{\mathrm{D}\top}\right)^\top$ be the vector of the true random variables and let $X^* = \left(X^{*\mathrm{C}\top}, X^{*\mathrm{D}\top}\right)^\top$ denote the surrogate random vector generated from setting (σ_ϵ^2, π) to be (0.15, 0.9), (0.50, 0.8), or (0.75, 0.7).

We take the regression functions $\mathcal{G}^{\mathrm{C}}(\cdot,\cdot)$ and $\mathcal{G}^{\mathrm{D}}(\cdot,\cdot)$ in (31) to be the quadratic or linear function, as considered by [4, p.126]. In each setting, we consider different combinations of the sample size n and the dimension of X. In Scenario 1, we set $(n, p^{\mathrm{C}}) = (400, 20)$, $(400, 100)$, or $(200, 400)$; in Scenario 2 we examine $(n, p^{\mathrm{D}}) = (400, 20)$, $(400, 15)$, or $(15, 20)$; and in Scenario 3 we set $(n, p) = (400, 20)$, $(400, 100)$, or $(200, 300)$ with $(p^{\mathrm{C}}, p^{\mathrm{D}}) = (10, 10)$, $(90, 10)$, or $(280, 20)$. We perform 500 simulations for each setting.

6.2. Simulation results

We examine the *accuracy of the estimator* of Θ by reporting its L_1-norm $\|\Delta_\Theta\|_1$ and the Frobenius norm $\|\Delta_\Theta\|_F$, respectively, where $\Delta_\Theta = \widehat{\Theta} - \Theta_0$. To report the *accuracy of variable selection* for the graphical structure, we examine the *specificity* (Spe) and the *sensitivity* (Sen) for the estimator $\widehat{\Theta}$. The specificity is defined as the proportion of zero coefficients that are correctly estimated to be zero, and the sensitivity is defined as the proportion of nonzero coefficients that are correctly estimated to be nonzero. The simulation results of the naive and proposed methods are reported in Tables 1-3, where "naive" refers to the naive method which ignores the mismeasurement feature, "corrected-Q" represents the proposed method using the quadratic extrapolation function at Step 3 in Section 4.1, and 'corrected-L" indicates the proposed method using the linear extrapolation function at Step 3 in Section 4.1. As a reference for comparisons, we also use the true values of X for the estimation, and denote this method as "true".

It is apparent that the naive method yields seriously biased results. The values of the L_1-norm and the Frobenius norm are noticeably large whereas the

specificities are small for various settings. Although the sensitivities are all good for Scenario 1, they tend to be far off 1 in Scenarios 2 and 3. As the degree of mismeasurement increases, the bias incurred in the naive method becomes more substantial.

On the contrary, with measurement error effects taken into account, the proposed method obviously outperforms the naive method with smaller values of the L_1-norm and the Frobenius norm as well as higher specificities and sensitivities, and thus, yielding more accurate recovery of graphical structures. As noted in the comments of Theorem 5.2 (b), the measurement error degree has effects on the performance of the proposed estimators. Simulation results show that the performance of the proposed method deteriorates as mismeasurement becomes more severe. In addition, the choice of the extrapolation function affects the performance of the proposed method, as expected. Our numerical studies suggest that using the quadratic extrapolation function slightly outperforms using the linear extrapolation function.

To see computational complexity of the proposed method, we apply the R function `proc.time()` to record the CPU time (in seconds) taken to implement the preceding methods. Runtimes based on using Intel(R) Core(TM) i7-6700HQ CPU2.60GHz are displayed in the last column of Tables 1-3. Unsurprisingly, the runtime of all the methods increases as dimension p becomes larger. It is interesting that the runtime of all the methods increases with the increase of the mismeasurement degree. This may be partly attributed to that data involve more variability when increasing the mismeasurement magnitude. As expected, the proposed method is a lot more time-consuming than the naive method, and using the quadratic extrapolation function takes a longer time than using the linear extrapolation function.

6.3. Analysis of cell-signalling data

We implement the proposed method to analyze the cell-signalling data [22]. This dataset contains $p = 11$ proteins and $n = 7466$ cells. According to [22], the cell signaling is a communication process that controls cell activities. When an external signal (e.g., growth factor) binds to its specific cell surface receptor, the activated receptor will interact with signaling proteins inside cell, which triggers a cascade of information flow or signalling pathway. The signaling pathway involves chemical, physical or locational modifications of protein-protein interaction, which leads to a specific cell response such as inducing the transcription and translation to produce certain proteins. It is important to understand the relationship among various signaling proteins/molecules by investigating signaling pathways and the dependence structure of proteins.

To this end, several authors analyzed the data with different approaches. [22] fitted a directed acyclic graph to the data, and [13] implemented the graphical lasso method to estimate the network structure of the proteins. However, those methods do not address the effects due to mismeasurement, a common phenomenon associated with measurements of cell signaling, as pointed out by [1] and [36].

In our analysis here, we address the feature of mismeasurement by applying the proposed method. Since the dataset has no additional information such as repeated measurements or validation data to quantify the degree of measurement error, we conduct sensitivity analyses to investigate how the analysis results are affected by different magnitudes of measurement error. Let Σ_X and Σ_{X^*} denote covariance matrices of X and X^*, respectively. Let $\sigma_{X;i,j}$, $\sigma_{X^*;i,j}$, and $\sigma_{\epsilon;i,j}$ denote entry (i,j) of Σ_X, Σ_{X^*}, and Σ_ϵ, respectively. Since the dataset only contains continuous variables, we adopt measurement error model (18), which gives that $\Sigma_{X^*} = \Sigma_X + \Sigma_\epsilon$, suggesting that $\sigma_{X^*;i,j}$ is bigger than $\sigma_{X;i,j}$ for all i and j. As Σ_{X^*} is unknown, we use the empirical estimate, denoted by $\widehat{\Sigma}_{X^*}$, to estimate Σ_{X^*}, and take $\sigma_{X;i,j}$ as $\sigma_{X;i,j} = 0.9\widehat{\sigma}_{X^*;i,j}$ where $\widehat{\sigma}_{X^*;i,j}$ is entry (i,j) of $\widehat{\Sigma}_{X^*}$. Regarding the specification of $\sigma_{\epsilon;i,j}$, we use the reliability ratio $R_{ij} = \frac{\sigma_{X;i,j}}{\sigma_{X^*;i,j}}$ to guide us:

$$\sigma_{\epsilon;i,j} = (R_{ij}^{-1} - 1)\widehat{\sigma}_{X^*;i,j}. \tag{41}$$

Taking R_{ij} as a common value, say, R, for all i and j, then (41) gives that

$$\Sigma_\epsilon = (R^{-1} - 1)\widehat{\Sigma}_{X^*}.$$

Here we consider $R = 0.65, 0.75$ and 0.85, and report in Figure 2 the estimated networks that are determined by the proposed method using both the quadratic and linear extrapolation functions in Step 3 described in Section 4.1. In comparison, we also examine the naive analysis discussed in Section 3.2, and display the result in Figure 3.

Figure 2 demonstrates that the estimation of the network structure is clearly influenced by the degree of measurement error and the choice of extrapolation functions. When the quadratic extrapolation function is used, more connected edges are identified as R increases; comparing the case with $R = 0.75$ to that with $R = 0.65$ shows two additionally identified pairs (`PIP3` and `praf`; `pjnk` and `praf`); and another two pairs (`pakts473` and `pjnk`; `pakts473` and `praf`) are revealed by increasing $R = 0.75$ to $R = 0.85$.

On the contrary, when the linear extrapolation function is used, we do not observe the pattern of identifying more connected edges with an increase of R. Relative to the case with $R = 0.65$, under the setting with $R = 0.75$, we find evidence of four additional edges (`pjnk` and `P38`; `PKC` and `P38`; `PKA` and `PIP3`; `pjnk` and `pakts473`) but no evidence of the edge connecting `pjnk` and `praf`; under $R = 0.85$, we detect evidence of two extra edges (`praf` and `PIP2`; `pjnk` and `pakts473`).

In contrast, the naive method produces a more complex network structure and the result is clearly different from the proposed method which accounts for the measurement error effects. The naive method indicates more connected variables than the method which accommodates different magnitudes of measurement error. These studies demonstrate that in the presence of measurement error in the variables, ignoring the feature of mismeasurement may produce spurious correlation structures among the variables.

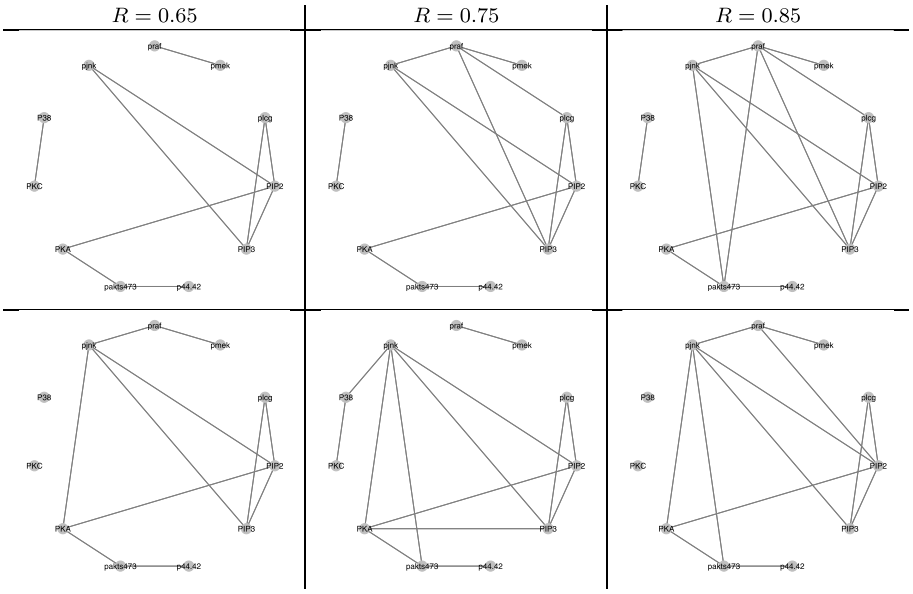

FIG 2. *Identification of graphical structures of 11 proteins using the proposed method by assuming different degrees of mismeasurement in cell-signalling data: the first row records the results obtained from the proposed method using the quadratic extrapolation function; the second row displays the results obtained from the proposed method using the linear extrapolation function.*

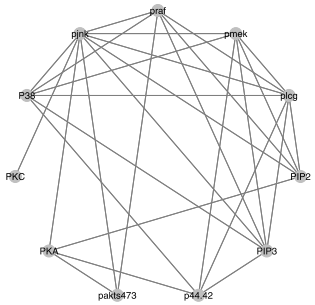

FIG 3. *Identification of graphical structures of 11 proteins using the naive method which ignores the feature of mismeasurement in cell-signalling data.*

7. Discussion

While graphical models have been extensively studied, existing work is mainly applied to mismeasurement-free data, though [26, Section 11.4.1] considered Gaussian graphical models with measurement error model (18). It is useful to develop methods to handle graphical models with error-prone variables, which is our objective here. Our framework covers a broad range of problems in the

sense that (1) both continuous and discrete variables are allowed, (2) graphical models do not have to be Gaussian but can be featured by the exponential family distribution, and (3) the variables can be error-contaminated. Our contributions are multiple. First, to demonstrate the mismeasurement effects, we derive a lower bound of the asymptotic bias due to the naive analysis which ignores measurement error. Secondly, to address the mismeasurement effects, we propose a simulation-based method for graph estimation. Thirdly, we establish theoretical results to justify the validity of the proposed method. Our development offers a complement to the available research on graphical models applicable to error-free settings.

To highlight the key ideas, we focus our attention on estimation of the network structure and assume the parameters for the mismeasurement models (18) and (20) to be known. Such an assumption is typically feasible in two circumstances: (i) prior studies provide the information on the degree of mismeasurement, and (ii) we are interested in conducting sensitivity analyses to understand how mismeasurement effects may affect inference results.

In situations where the parameters for the mismeasurement models (18) and (20) must be estimated, we may utilize the information carried by additional data sources such as repeated measurements or validation subsamples. For instance, with the availability of repeated measurements, estimation of misclassification probabilities can proceed in the same manner as that discussed by [32, Section 4], and Σ_ϵ for the measurement error model (18) can be estimated, using the method of moments, by

$$\widehat{\Sigma}_\epsilon = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left(X^{*\mathrm{C}(ij)} - \bar{X}^{*\mathrm{C}(i)}\right)\left(X^{*\mathrm{C}(ij)} - \bar{X}^{*\mathrm{C}(i)}\right)^\top}{\sum_{i=1}^{n}(n_i - 1)},$$

where $X^{*\mathrm{C}(ij)}$ denotes the jth replicate of $X^{\mathrm{C}(i)}$ with $j = 1, \cdots, n_i$, n_i is the number of the replicates for subject i, and $\bar{X}^{*\mathrm{C}(i)} = n_i^{-1}\sum_{j=1}^{n_i} X^{*\mathrm{C}(ij)}$ for $i = 1, \cdots, n$. When validation data are available, one may adapt the discussion of [33] and [35] to incorporate estimation of the parameters for the mismeasurement models (18) and (20) into inferential procedures.

To describe the relationship between $\left(X^{*\mathrm{C}\top}, X^{*\mathrm{D}\top}\right)^\top$ and $\left(X^{\mathrm{C}\top}, X^{\mathrm{D}\top}\right)^\top$, we use the factorization

$$\left[X^{*\mathrm{C}}, X^{*\mathrm{D}} \middle| X^{\mathrm{C}}, X^{\mathrm{D}}\right] = \left[X^{*\mathrm{C}} \middle| X^{*\mathrm{D}}, X^{\mathrm{C}}, X^{\mathrm{D}}\right] \times \left[X^{*\mathrm{D}} \middle| X^{\mathrm{C}}, X^{\mathrm{D}}\right], \tag{42}$$

where $[\,\bullet \mid \bullet\,]$ represents the conditional distribution for the random variables indicated by the arguments. Making the independence assumptions in (18) and (19) allows (42) to be written as

$$\left[X^{*\mathrm{C}}, X^{*\mathrm{D}} \middle| X^{\mathrm{C}}, X^{\mathrm{D}}\right] = \left[X^{*\mathrm{C}} \middle| X^{\mathrm{C}}\right] \times \left[X^{*\mathrm{D}} \middle| X^{\mathrm{D}}\right],$$

and thus enabling us to use model (18) to characterize $\left[X^{*\mathrm{C}} \middle| X^{\mathrm{C}}\right]$ and the misclassification matrix $\mathcal{P}$ in (20) to facilitate $\left[X^{*\mathrm{D}} \middle| X^{\mathrm{D}}\right]$. Consequently, the SIMEX

[10] and MC-SIMEX [17] algorithms, developed based on (18) and (20), respectively, can be employed to address the effects of mismeasurement in X^{C} and X^{D}. It is interesting to extend our development here by relaxing the independence assumptions in (18) and (19) or considering other mismeasurement models; valid inference procedures under these settings warrant a careful study.

The development here takes the dimension p as fixed and allows p to be bigger than the sample size n, as shown in the simulation studies. The order of p is basically of the same order as n. In situations where p depends on n such as being of a polynomial or exponential order of n, it is not feasible to apply the proposed method because of the prohibitive computation time. In addition, our theoretical results in Section 5 are established only for the case with p fixed. It is interesting to extend the development to accommodating settings with a diverging p, such as $p \gg n$.

Appendix A: Technical Lemmas

In this section, we present some lemmas which will be used in the proof of the main theorems.

Lemma A.1. *Let $Z_1, \cdots, Z_n$ be independent and identically distributed (i.i.d.) random variables. Define $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$. Suppose that $E\{\exp(aZ_i)\}$ exists for $a > 0$, then for any $\delta > 0$,*

$$P\left(\bar{Z} > \delta\right) \leq \frac{\exp\left(n \log\left[E\left\{\exp(aZ_i)\right\}\right]\right)}{\exp\left(na\delta\right)}. \tag{A.1}$$

Proof:

Let $Z = \sum_{i=1}^{n} Z_i$, then by the Markov's inequality, for any $a > 0$, we have

$$\begin{aligned} P\left(\bar{Z} > \delta\right) &= P\left(Z > n\delta\right) \\ &\leq \frac{E\left\{\exp\left(aZ\right)\right\}}{\exp\left(na\delta\right)} \\ &= \frac{E\left\{\prod_{i=1}^{n} \exp\left(aZ_i\right)\right\}}{\exp\left(na\delta\right)}. \end{aligned} \tag{A.2}$$

Noting that $E\left\{\prod_{i=1}^{n} \exp\left(aZ_i\right)\right\}$ in (A.2) can be written as

$$\begin{aligned} E\left\{\prod_{i=1}^{n} \exp\left(aZ_i\right)\right\} &= \prod_{i=1}^{n} E\left\{\exp\left(aZ_i\right)\right\} \\ &= \exp\left(\log\left[\prod_{i=1}^{n} E\left\{\exp\left(aZ_i\right)\right\}\right]\right) \end{aligned}$$

$$= \exp\left(\sum_{i=1}^{n} \log\left[E\left\{\exp\left(aZ_i\right)\right\}\right]\right)$$
$$= \exp\left(n \log\left[E\left\{\exp\left(aZ_i\right)\right\}\right]\right). \quad \text{(A.3)}$$

As a result, combining (A.2) and (A.3) gives (A.1). □

Lemma A.2. *Assume regularity conditions in Section 5.1. Consider the case with $n \geq 2\kappa_2 \log(p)$. Then for all $r \in V^C$ and $r' \in V^D$, there exist positive constants c_1 and c_2 such that*

$$P\left(\left\|\nabla_{\theta^C(r)} \ell^C_{b,\zeta}(\theta^C(r))\right\|_\infty > \frac{\alpha}{2-\alpha}\frac{\lambda^C_n}{4}\right) < c_1\{\max(n,p)\}^{-2} + \exp(-c_2 n) \quad \text{(A.4)}$$

and

$$P\left(\left\|\nabla_{\theta^D(r')} \ell^D_{b,\zeta}(\theta^D(r'))\right\|_\infty > \frac{\alpha}{2-\alpha}\frac{\lambda^D_n}{4}\right) < c_1\{\max(n,p)\}^{-2} + \exp(-c_2 n).$$

Proof:

Here we show only the first inequality; the second inequality can be proved similarly. Taking partial derivative of $\ell^{\mathrm{C}}_{b,\zeta}(\theta^{\mathrm{C}}(r))$ defined in (28) with respect to $\theta^{\mathrm{C}}(r)$, we have that for $r \in V^{\mathrm{C}}$,

$$\nabla_{\theta^{\mathrm{C}}(r)} \ell^{\mathrm{C}}_{b,\zeta}(\theta^{\mathrm{C}}(r)) = \left(\nabla_{\theta^{\mathrm{C}}_r} \ell^{\mathrm{C}}_{b,\zeta}(\theta^{\mathrm{C}}(r)), \nabla_{\theta^{\mathrm{C}}_{\setminus r}} \ell^{\mathrm{C}\top}_{b,\zeta}(\theta^{\mathrm{C}}(r)), \nabla_{\theta^{\mathrm{CD}}_r} \ell^{\mathrm{C}\top}_{b,\zeta}(\theta^{\mathrm{C}}(r))\right)^\top,$$

where

$$\nabla_{\theta^{\mathrm{C}}_{\setminus r}} \ell^{\mathrm{C}}_{b,\zeta}(\theta^{\mathrm{C}}(r)) = \left(-\frac{1}{n}\sum_{i=1}^{n} U^{\mathrm{C}(i)}_{r,t}(b,\zeta) : t \in V^{\mathrm{C}} \setminus \{r\}\right), \quad \text{(A.5)}$$

$$\nabla_{\theta^{\mathrm{CD}}_r} \ell^{\mathrm{C}}_{b,\zeta}(\theta^{\mathrm{C}}(r)) = \left(-\frac{1}{n}\sum_{i=1}^{n} U^{\mathrm{D}(i)}_{r,t'}(b,\zeta) : t' \in V^{\mathrm{D}}\right), \quad \text{(A.6)}$$

and

$$\nabla_{\theta^{\mathrm{C}}_r} \ell^{\mathrm{C}}_{b,\zeta}(\theta^{\mathrm{C}}(r)) = -\frac{1}{n}\sum_{i=1}^{n} U^{(i)}_r(b,\zeta) \quad \text{(A.7)}$$

with

$$U^{\mathrm{C}(i)}_{r,t}(b,\zeta) = W^{\mathrm{C}(i)}_{b,r}(\zeta) W^{\mathrm{C}(i)}_{b,t}(\zeta) - W^{\mathrm{C}(i)}_{b,t}(\zeta) K^{\mathrm{C}'}\left(\eta^{\mathrm{C}(i)}_r(b,\zeta)\right),$$

$$U^{\mathrm{D}(i)}_{r,t'}(b,\zeta) = W^{\mathrm{C}(i)}_{b,r}(\zeta) W^{\mathrm{D}(i)}_{b,t'}(\zeta) - W^{\mathrm{D}(i)}_{b,t'}(\zeta) K^{\mathrm{C}'}\left(\eta^{\mathrm{C}(i)}_r(b,\zeta)\right),$$

$$U_r^{(i)}(b,\zeta) = W_{b,r}^{\mathrm{C}(i)}(\zeta) - K^{\mathrm{C}'}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)\right),$$

and $\eta_r^{\mathrm{C}(i)}(b,\zeta) = \theta_r^{\mathrm{C}} + \sum_{t\in V^{\mathrm{C}}\setminus\{r\}} \theta_{rt}^{\mathrm{C}} W_{b,t}^{\mathrm{C}(i)}(\zeta) + \sum_{t'\in V^{\mathrm{D}}} \theta_{rt'}^{\mathrm{CD}} W_{b,t'}^{\mathrm{D}(i)}(\zeta)$ for $t \in V^{\mathrm{C}}\setminus\{r\}$ and $t' \in V^{\mathrm{D}}$. In the sequel, we examine (A.5), (A.6) and (A.7) separately.

Step 1: *Examine $U_{r,t}^{C(i)}(b,\zeta)$ in (A.5), and show that for $r \in V^C$ and $t \in V^C\setminus\{r\}$ and any constant $a>0$, there exists some $v_t \in (0,1)$ such that*

$$\begin{aligned} & E\left\{\exp\left(aU_{r,t}^{C(i)}(b,\zeta)\right)\middle| W_{b;\setminus r}^{C(i)}(\zeta), W_b^{D(i)}(\zeta)\right\} \\ = \;& \exp\left\{\frac{a^2}{2}\left(W_{b,t}^{C(i)}(\zeta)\right)^2 K^{C''}\left(v_t a W_{b,t}^{C(i)}(\zeta) + \eta_r^{C(i)}(b,\zeta)\right)\right\} \end{aligned}$$

for $b=1,\cdots,B$ and $\zeta\in\mathcal{Z}$, where $W_{b;\setminus r}^{C(i)}(\zeta)$ is the (p^C-1)-dimensional vector of $W_b^{C(i)}(\zeta)$ with the rth component deleted.

First, applying the one-to-one transformation, the conditional distribution (12) based on $X^{(i)}$ can be transferred to

$$\begin{aligned} & P\left(W_{b,r}^{\mathrm{C}(i)}(\zeta)\middle| W_{b;\setminus r}^{\mathrm{C}(i)}(\zeta), W_b^{\mathrm{D}(i)}(\zeta)\right) \\ = \;& \exp\left\{W_{b,r}^{\mathrm{C}(i)}(\zeta)\eta_r^{\mathrm{C}}(b,r) + H^{\mathrm{C}}\left(W_{b,r}^{\mathrm{C}(i)}(\zeta)\right) - K^{\mathrm{C}}(\eta_r^{\mathrm{C}}(b,r))\right\}. \end{aligned} \tag{A.8}$$

By the definition of $U_{r,t}^{\mathrm{C}(i)}(b,\zeta)$ in (A.7), we have that for any constant $a>0$,

$$\begin{aligned} & E\left\{\exp\left(aU_{r,t}^{\mathrm{C}(i)}(b,\zeta)\right)\middle| W_{b;\setminus r}^{\mathrm{C}(i)}(\zeta), W_b^{\mathrm{D}(i)}(\zeta)\right\} \\ = \;& E\Big(\exp\left[a\left\{W_{b,r}^{\mathrm{C}(i)}(\zeta)W_{b,t}^{\mathrm{C}(i)}(\zeta) - W_{b,t}^{\mathrm{C}(i)}(\zeta)K^{\mathrm{C}'}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\}\right]\Big| W_{b;\setminus r}^{\mathrm{C}(i)}(\zeta), W_b^{\mathrm{D}(i)}(\zeta)\Big) \\ = \;& \int\Big[\exp\left\{aW_{b,r}^{\mathrm{C}(i)}(\zeta)W_{b,t}^{\mathrm{C}(i)}(\zeta)\right\} \times \exp\left\{-aW_{b,t}^{\mathrm{C}(i)}(\zeta)K^{\mathrm{C}'}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\} \\ & \times P\left(W_{b,r}^{\mathrm{C}(i)}(\zeta)\middle| W_{b;\setminus r}^{\mathrm{C}(i)}(\zeta), W_b^{\mathrm{D}(i)}(\zeta)\right)\Big] dW_{b,r}^{\mathrm{C}(i)}(\zeta) \\ = \;& \int\Big[\exp\left\{aW_{b,r}^{\mathrm{C}(i)}(\zeta)W_{b,t}^{\mathrm{C}(i)}(\zeta)\right\}\exp\left\{W_{b,r}^{\mathrm{C}(i)}(\zeta)\eta_r^{\mathrm{C}(i)}(b,\zeta) + H^{\mathrm{C}}\left(W_{b,r}^{\mathrm{C}(i)}(\zeta)\right)\right. \\ & \left. -K^{\mathrm{C}}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\} \times \exp\left\{-aW_{b,t}^{\mathrm{C}(i)}(\zeta)K^{\mathrm{C}'}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\}\Big] dW_{b,r}^{\mathrm{C}(i)}(\zeta) \\ = \;& \int\Big[\exp\left\{W_{b,r}^{\mathrm{C}(i)}(\zeta)\left(\eta_r^{\mathrm{C}(i)}(b,\zeta) + aW_{b,t}^{\mathrm{C}(i)}(\zeta)\right) + H^{\mathrm{C}}\left(W_{b,r}^{\mathrm{C}(i)}(\zeta)\right)\right. \\ & \left. -K^{\mathrm{C}}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\} \times \exp\left\{-aW_{b,t}^{\mathrm{C}(i)}(\zeta)K^{\mathrm{C}'}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\}\Big] dW_{b,r}^{\mathrm{C}(i)}(\zeta), \end{aligned} \tag{A.9}$$

where the second step is due to the definition of the conditional expectation, the third step comes from (A.8), and the last step combines the terms with $W_{b,r}^{\mathrm{C}(i)}(\zeta)$.

Furthermore, by adding and subtracting an additional term $K^{\mathrm{C}}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)+aW_{b,t}^{\mathrm{C}(i)}(\zeta)\right)$, (A.9) can be written as

$$\begin{aligned}
& E\left\{\exp\left(aU_{r,t}^{\mathrm{C}(i)}(b,\zeta)\right)\middle| W_{b;\backslash r}^{\mathrm{C}(i)}(\zeta), W_b^{\mathrm{D}(i)}(\zeta)\right\} \\
= & \int\Big[\exp\left\{W_{b,r}^{\mathrm{C}(i)}(\zeta)\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)+aW_{b,t}^{\mathrm{C}(i)}(\zeta)\right)+H^{\mathrm{C}}\left(W_{b,r}^{\mathrm{C}(i)}(\zeta)\right)\right. \\
& \left.-K^{\mathrm{C}}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)+aW_{b,t}^{\mathrm{C}(i)}(\zeta)\right)\right\}\Big] dW_{b,r}^{\mathrm{C}(i)}(\zeta) \\
& \times\exp\left\{K^{\mathrm{C}}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)+aW_{b,t}^{\mathrm{C}(i)}(\zeta)\right)-K^{\mathrm{C}}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\} \\
& \times\exp\left\{-aW_{b,t}^{\mathrm{C}(i)}(\zeta)K^{\mathrm{C}'}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\} \\
= & \exp\left\{K^{\mathrm{C}}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)+aW_{b,t}^{\mathrm{C}(i)}(\zeta)\right)-K^{\mathrm{C}}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\} \\
& \times\exp\left\{-aW_{b,t}^{\mathrm{C}(i)}(\zeta)K^{\mathrm{C}'}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\} \\
= & \exp\left\{\frac{a^2}{2}\left(W_{b,r}^{\mathrm{C}(i)}(\zeta)\right)^2 K^{\mathrm{C}''}\left(v_t aW_{b,t}^{\mathrm{C}(i)}(\zeta)+\eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\}, \qquad \text{(A.10)}
\end{aligned}$$

where the second step holds since the integration is one by (A.8), and the third step is due to the second order Taylor series expansion on $K^{\mathrm{C}}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)+aW_{b,t}^{\mathrm{C}(i)}(\zeta)\right)$ around $a=0$ for some $v_t\in(0,1)$.

Step 2: *Examine (A.5) and show that*

$$\begin{aligned}
& P\left(\left|\frac{1}{n}\sum_{i=1}^{n}U_{r,t}^{\mathrm{C}(i)}(b,\zeta)\right|>\frac{\alpha}{2-\alpha}\frac{\lambda_n^C}{4}\,\middle|\,\mathcal{E}_1,\mathcal{E}_2\right) \\
& <\exp\left\{-\frac{n}{2\eta_1\kappa_2}\left(\frac{\alpha}{2-\alpha}\frac{\lambda_n^C}{4}\right)^2\right\} \qquad \text{(A.11)}
\end{aligned}$$

for $b=1,\cdots,B$ *and* $\zeta\in\mathcal{Z}$ *defined in Section 4.1, where* $\mathcal{E}_1=\left\{\max_{i,r}W_{b,r}^{C(i)}(\zeta)\le 4\log n\right\}$, $\mathcal{E}_2=\left\{\max_{t\in V^C}\frac{1}{n}\sum_{i=1}^{n}\left\{W_{b,t}^{C(i)}(\zeta)\right\}^2\le\kappa_2\right\}$, λ_n^C *is the tuning parameter defined in (28);* α, η_1, *and* κ_2 *are constants defined in Section 5.1.*

By the derivations of Proposition 3 and Lemma 9 in [28], we have $P\left(\overline{\mathcal{E}}_1\right)\le c_1\{\max(n,p)\}^{-2}$ and $P\left(\overline{\mathcal{E}}_2\right)\le\exp\left(-b_1 n\right)$ for some positive constants c_1 and b_1, where $\overline{\mathcal{E}}_1$ and $\overline{\mathcal{E}}_2$ are the complement sets of $\mathcal{E}_1$ and $\mathcal{E}_2$, respectively. Therefore, by Assumption (A3), the upper bound of (A.10) is given by

$$E\left\{\exp\left(aU_{r,t}^{\mathrm{C}(i)}(b,\zeta)\right)\middle| W_{b;\backslash r}^{\mathrm{C}(i)}(\zeta), W_b^{\mathrm{D}(i)}(\zeta)\right\}\le\exp\left(\frac{a^2}{2}\eta_1\kappa_2\right). \qquad \text{(A.12)}$$

By (A.12) and Lemma A.1 with Z_i treated as $U_{r,t}^{\mathrm{C}(i)}(b,\zeta)$, we have that for any

$\delta > 0$,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} U_{r,t}^{\mathrm{C}(i)}(b,\zeta)\right| > \delta \,\middle|\, \mathcal{E}_1, \mathcal{E}_2\right) < \exp\left\{n\left(\frac{\eta_1\kappa_2 a^2}{2} - \delta a\right)\right\},$$

and specifying $a = \frac{\delta}{\eta_1\kappa_2}$ yields

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} U_{r,t}^{\mathrm{C}(i)}(b,\zeta)\right| > \delta \,\middle|\, \mathcal{E}_1, \mathcal{E}_2\right) < \exp\left(-\frac{n\delta^2}{2\eta_1\kappa_2}\right).$$

Finally, specifying $\delta = \frac{\alpha}{2-\alpha}\frac{\lambda_n^{\mathrm{C}}}{4}$ gives (A.11).

Step 3: *Examine (A.6) and show that for* $b = 1, \cdots, B$ *and* $\zeta \in \mathcal{Z}$,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} U_{r,t'}^{D(i)}(b,\zeta)\right| > \frac{\alpha}{2-\alpha}\frac{\lambda_n^{C}}{4} \,\middle|\, \mathcal{E}_1, \mathcal{E}_2\right) < \exp\left\{-\frac{n}{2\eta_1\kappa_2}\left(\frac{\alpha}{2-\alpha}\frac{\lambda_n^{C}}{4}\right)^2\right\}. \tag{A.13}$$

By the similar derivations of (A.9) and (A.10), we have that for any $a > 0$,

$$\begin{aligned}
& E\left\{\exp\left(aU_{r,t'}^{\mathrm{D}(i)}(b,\zeta)\right) \middle| W_{b;\setminus r}^{\mathrm{C}(i)}(\zeta), W_b^{\mathrm{D}(i)}(\zeta)\right\} \\
= \;& E\left(\exp\left[a\left\{W_{b,r}^{\mathrm{C}(i)}(\zeta)W_{b,t'}^{\mathrm{D}(i)}(\zeta) - W_{b,t'}^{\mathrm{D}(i)}(\zeta)K^{\mathrm{C}'}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\}\right] \middle| W_{b;\setminus r}^{\mathrm{C}(i)}(\zeta), W_b^{\mathrm{D}(i)}(\zeta)\right) \\
= \;& \int \left[\exp\left\{W_{b,r}^{\mathrm{C}(i)}(\zeta)\left(\eta_r^{\mathrm{C}(i)}(b,\zeta) + aW_{b,t'}^{\mathrm{D}(i)}(\zeta)\right) + H^{\mathrm{C}}\left(W_{b,r}^{\mathrm{C}(i)}(\zeta)\right)\right. \right. \\
& \left.\left. -K^{\mathrm{C}}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\} \times \exp\left\{-aW_{b,t'}^{\mathrm{D}(i)}(\zeta)K^{\mathrm{C}'}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\}\right] dW_{b,r}^{\mathrm{C}(i)}(\zeta) \\
= \;& \int \left[\exp\left\{W_{b,r}^{\mathrm{C}(i)}(\zeta)\left(\eta_r^{\mathrm{C}(i)}(b,\zeta) + aW_{b,t'}^{\mathrm{D}(i)}(\zeta)\right) + H^{\mathrm{C}}\left(W_{b,r}^{\mathrm{C}(i)}(\zeta)\right)\right. \right. \\
& \left.\left. -K^{\mathrm{C}}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta) + aW_{b,t'}^{\mathrm{D}(i)}(\zeta)\right)\right\}\right] dW_{b,r}^{\mathrm{C}(i)}(\zeta) \\
& \times \exp\left\{K^{\mathrm{C}}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta) + aW_{b,t'}^{\mathrm{D}(i)}(\zeta)\right) - K^{\mathrm{C}}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\} \\
& \times \exp\left\{-aW_{b,t'}^{\mathrm{D}(i)}(\zeta)K^{\mathrm{C}'}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\} \\
= \;& \exp\left\{\frac{a^2}{2}\left(W_{b,t'}^{\mathrm{D}(i)}(\zeta)\right)^2 K^{\mathrm{C}''}\left(v_{t'}aW_{b,t'}^{\mathrm{D}(i)}(\zeta) + \eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\},
\end{aligned}$$

where the third step is due to adding and subtracting $\exp\left\{K^{\mathrm{C}}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta) + aW_{b,t'}^{\mathrm{D}(i)}(\zeta)\right)\right\}$ and $v_{t'}$ in the fourth step is between 0 and 1. After that, by the similar derivation in Step 2, we have (A.13).

Step 4: *Examine (A.7) and show that for* $b = 1, \cdots, B$ *and* $\zeta \in \mathcal{Z}$,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} U_r^{(i)}(b,\zeta)\right| > \frac{\alpha}{2-\alpha}\frac{\lambda_n^{C}}{4} \,\middle|\, \mathcal{E}_1, \mathcal{E}_2\right) < \exp\left\{-\frac{n}{\eta_1}\left(\frac{\alpha}{2-\alpha}\frac{\lambda_n^{C}}{4}\right)^2\right\}. \tag{A.14}$$

Indeed, by the derivations similar to Step 1, for any constant $\widetilde{a} > 0$, there exists $\widetilde{v} \in (0,1)$ such that

$$E\left\{\exp\left(\widetilde{a}U_r^{(i)}(b,\zeta)\right)\middle| W_{b;\setminus r}^{\mathrm{C}(i)}(\zeta), W_b^{\mathrm{D}(i)}(\zeta)\right\} = \exp\left\{\frac{\widetilde{a}^2}{2}K^{\mathrm{C}''}\left(\widetilde{v}_t\widetilde{a} + \eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\}.$$

By the derivations similar to Step 2 with Assumption (A3) and $\widetilde{a}$ replaced by $\frac{\widetilde{\delta}}{\eta_1}$, we have that for any constant $\widetilde{\delta}$,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} U_r^{(i)}(b,\zeta)\right| > \widetilde{\delta}\,\middle|\,\mathcal{E}_1,\mathcal{E}_2\right) < \exp\left\{-\frac{n\widetilde{\delta}^2}{\eta_1}\right\}. \tag{A.15}$$

Finally, replacing $\widetilde{\delta}$ in (A.15) by $\frac{\alpha}{2-\alpha}\frac{\lambda_n^{\mathrm{C}}}{4}$ gives (A.14).

Step 5: *Examine* $\nabla_{\theta^{C}(r)}\ell_{b,\zeta}^{C}(\theta^{C}(r))$ *and show the final result.*

Recall that

$$\nabla_{\theta^{\mathrm{C}}(r)}\ell_{b,\zeta}^{\mathrm{C}}(\theta^{\mathrm{C}}(r)) = \left(\nabla_{\theta_r^{\mathrm{C}}}\ell_{b,\zeta}^{\mathrm{C}}(\theta(r)), \nabla_{\theta_{\setminus r}^{\mathrm{C}}}\ell_{b,\zeta}^{\mathrm{C}\top}(\theta^{\mathrm{C}}(r)), \nabla_{\theta_r^{\mathrm{CD}}}\ell_{b,\zeta}^{\mathrm{C}\top}(\theta^{\mathrm{C}}(r))\right)^{\top}.$$

Then by (A.11), (A.13) and (A.14), we have

$$\begin{aligned} & P\left(\left\|\nabla_{\theta^{\mathrm{C}}(r)}\ell_{b,\zeta}^{\mathrm{C}}(\theta^{\mathrm{C}}(r))\right\|_\infty > \frac{\alpha}{2-\alpha}\frac{\lambda_n^{\mathrm{C}}}{4}\,\middle|\,\mathcal{E}_1,\mathcal{E}_2\right) \\ < \;& 2\exp\left\{-\frac{n}{2\eta_1\kappa_2}\left(\frac{\alpha}{2-\alpha}\frac{\lambda_n^{\mathrm{C}}}{4}\right)^2 + \log p\right\} + \exp\left\{-\frac{n}{\eta_1}\left(\frac{\alpha}{2-\alpha}\frac{\lambda_n^{\mathrm{C}}}{4}\right)^2\right\}. \end{aligned}$$

As a result, provided that $\lambda_n^{\mathrm{C}} > \sqrt{\frac{\eta_1\kappa_2^*\log p}{n}}\left(\frac{2-\alpha}{\alpha}\right)$ with $\kappa_2^* \triangleq 64\kappa_2$ and inequality $P(A) \le P(\overline{\mathcal{E}}_1) + P(\overline{\mathcal{E}}_2) + P(A|\,\mathcal{E}_1,\mathcal{E}_2)$ for an event A [28, pp.29], we have

$$\begin{aligned} & P\left(\left\|\nabla_{\theta^{\mathrm{C}}(r)}\ell_{b,\zeta}^{\mathrm{C}}(\theta^{\mathrm{C}}(r))\right\|_\infty > \frac{\alpha}{2-\alpha}\frac{\lambda_n^{\mathrm{C}}}{4}\right) \\ < \;& c_1\{\max(n,p)\}^{-2} + 2\exp(-b_1 n) + \exp(-c_3 n) \\ < \;& c_1\{\max(n,p)\}^{-2} + \exp(-c_2 n) \end{aligned}$$

for some constants c_3 and $c_2 < \min\{b_1, c_3\} - \frac{\log 3}{n}$. □

Lemma A.3. *For* $r \in V^{C}$ *and* $r' \in V^{D}$, *let*

$$\widehat{\theta}^{C}(r;\zeta,b) = \left(\widehat{\theta}_{\mathcal{S}_r^C}^{C\top}(r;\zeta,b), \widehat{\theta}_{\overline{\mathcal{S}}_r^C}^{C\top}(\zeta,b)\right)^{\top}$$

with $\widehat{\theta}_{\mathcal{S}_r^C}^{C}(r;\zeta,b) = \left(\widehat{\theta}_r^{C}(\zeta,b), \widehat{\theta}_{\mathcal{S}_r^C}^{C\top}(\zeta,b)\right)^{\top}$ *and let*

$$\widehat{\theta}^{D}(r';\zeta,b) = \left(\widehat{\theta}_{\mathcal{S}_{r'}^D}^{D\top}(r';\zeta,b), \widehat{\theta}_{\overline{\mathcal{S}}_{r'}^D}^{D\top}(\zeta,b)\right)^{\top}$$

with $\widehat{\theta}^{D}_{\mathcal{S}^{D}_{r'}}(r';\zeta,b)=\left(\widehat{\theta}^{D}_{r'}(\zeta,b),\widehat{\theta}^{D\top}_{\mathcal{S}^{D}_{r'}}(\zeta,b)\right)^{\top}$. *Under regularity conditions in Section 5.1, we have*

$$\left\|\widehat{\theta}^{C}_{\mathcal{S}^{C}_{r}}(r;\zeta,b)-\theta^{C}_{0;\mathcal{S}^{C}_{r}}(r)\right\|_2\leq\frac{6\sqrt{d^{C}_{r}}\lambda^{C}_{n}}{\rho} \tag{A.16}$$

and

$$\left\|\widehat{\theta}_{\mathcal{S}^{D}_{r'}}(r';\zeta,b)-\theta^{D}_{0;\mathcal{S}^{D}_{r'}}(r')\right\|_2\leq\frac{6\sqrt{d^{D}_{r'}}\lambda^{D}_{n}}{\rho}. \tag{A.17}$$

Proof:

In the following proof, we show only (A.16); the second inequality (A.17) can be proved similarly.

For $r\in V^{\mathrm{C}}$, by the definition of $\overline{\mathcal{S}}^{\mathrm{C}}_{r}$, we have

$$\widehat{\theta}^{\mathrm{C}}(r;\zeta,b)=\left(\widehat{\theta}^{\mathrm{C}\top}_{\mathcal{S}^{\mathrm{C}}_{r}}(r;\zeta,b),\widehat{\theta}^{\mathrm{C}\top}_{\overline{\mathcal{S}}^{\mathrm{C}}_{r}}(\zeta,b)\right)^{\top}=\left(\widehat{\theta}^{\mathrm{C}\top}_{\mathcal{S}^{\mathrm{C}}_{r}}(r;\zeta,b),0^{\top}_{(p-d^{\mathrm{C}}_{r}-1)}\right)^{\top},$$

where 0_d stands for the d-dimensional zero vector. We write the true value of $\theta^{\mathrm{C}}(r)$ as $\theta^{\mathrm{C}}_{0}(r)=\left(\theta^{\mathrm{C}\top}_{0;\mathcal{S}^{\mathrm{C}}_{r}}(r),\theta^{\top}_{0;\overline{\mathcal{S}}^{\mathrm{C}}_{r}}\right)^{\top}=\left(\theta^{\mathrm{C}\top}_{0;\mathcal{S}^{\mathrm{C}}_{r}}(r),0^{\top}_{(p-d^{\mathrm{C}}_{r}-1)}\right)^{\top}$ with $\theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_{r}}(r)=\left(\theta^{\mathrm{C}}_{0;r},\theta^{\top}_{0;\mathcal{S}^{\mathrm{C}}_{r}}\right)^{\top}$ for $r\in V^{\mathrm{C}}$.

Claim: *For* $r\in V^{C}$, $\zeta\in\mathcal{Z}$ *and* $b=1,\cdots,B$, *let* $\widehat{u}_{\mathcal{S}^{C}_{r}}=\widehat{\theta}^{C}_{\mathcal{S}^{C}_{r}}(r;\zeta,b)-\theta^{C}_{0;\mathcal{S}^{C}_{r}}(r)$. *Show that*

$$\left\|\widehat{u}_{\mathcal{S}^{C}_{r}}\right\|_2\leq\frac{6\sqrt{d^{C}_{r}}\lambda^{C}_{n}}{\rho}. \tag{A.18}$$

We define the function $\Phi:\mathbb{R}^{d^{\mathrm{C}}_{r}+1}\to\mathbb{R}$ by

$$\begin{aligned}\Phi(u) &= \ell^{\mathrm{C}}_{b,\zeta}\left(\theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_{r}}(r)+u\right)-\ell^{\mathrm{C}}_{b,\zeta}\left(\theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_{r}}(r)\right)\\ &\quad+\lambda^{\mathrm{C}}_{n}\left(\left\|\theta_{0;\mathcal{S}^{\mathrm{C}}_{r}}+u\right\|_1-\left\|\theta_{0;\mathcal{S}^{\mathrm{C}}_{r}}\right\|_1\right),\end{aligned} \tag{A.19}$$

where we express any parameter value $\theta^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_{r}}(r)$ by $u+\theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_{r}}(r)$.

Note that $\Phi(u)$ is a convex function since $\ell^{\mathrm{C}}_{b,\zeta}(\cdot)$ defined in Section 4.1 and the L_1-norm $\|\cdot\|_1$ are both convex functions. Similar to the derivations for Lemma 3 of [20], to show (A.18), it suffices to show that

$$\Phi(u)>0\ \text{ for any }\ u\ \text{ with }\ \|u\|_2=\mathcal{B}, \tag{A.20}$$

where $\mathcal{B}=\frac{6\sqrt{d^{\mathrm{C}}_{r}}\lambda^{\mathrm{C}}_{n}}{\rho}$.

By the second order Taylor series expansion on

$$\ell^{\mathrm{C}}_{b,\zeta}\left(\theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_{r}}(r)+u\right)-\ell^{\mathrm{C}}_{b,\zeta}\left(\theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_{r}}(r)\right)$$

around $u = 0$, (A.19) becomes

$$\Phi(u) \quad = \quad T_1 + T_2 + T_3, \tag{A.21}$$

where

$$T_1 \quad = \quad \nabla_{\theta^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}(r)} \ell^{\mathrm{C}}_{b,\zeta} \left(\theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_r}(r) \right) u; \tag{A.22a}$$

$$T_2 \quad = \quad \frac{1}{2} u^\top \nabla^2_{\theta^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}(r)} \ell^{\mathrm{C}}_{b,\zeta} \left(\theta_{0;\mathrm{C};\mathcal{S}^{\mathrm{C}}_r}(r) + vu \right) u; \tag{A.22b}$$

$$T_3 \quad = \quad \lambda^{\mathrm{C}}_n \left(\left\| \theta_{0;\mathcal{S}^{\mathrm{C}}_r} + u \right\|_1 - \left\| \theta_{0;\mathcal{S}^{\mathrm{C}}_r} \right\|_1 \right), \tag{A.22c}$$

and v is some constant in $(0, 1)$.

We first specify $\mathcal{B}$ in (A.20) by $\mathcal{M}\lambda^{\mathrm{C}}_n \sqrt{d^{\mathrm{C}}_r}$ for some $\mathcal{M} > 0$. The remaining task is to individually examine T_1, T_2 and T_3 for their bound when $\|u\|_2 = \mathcal{M}\lambda^{\mathrm{C}}_n \sqrt{d^{\mathrm{C}}_r}$. We proceed with the following four steps.

Step 1: *For $r \in V^{\mathrm{C}}$, show that*

$$\|T_1\|_1 \quad < \quad \frac{(\lambda^{\mathrm{C}}_n \sqrt{d^{\mathrm{C}}_r})^2}{4} \mathcal{M} \quad \textit{for} \quad \|u\|_2 = \mathcal{M}\lambda^{\mathrm{C}}_n \sqrt{d^{\mathrm{C}}_r}. \tag{A.23}$$

For the first term T_1 in (A.21), by the result in Lemma A.2, we have

$$\begin{aligned} \|T_1\|_1 \quad &= \quad \left\| \nabla_{\theta^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}(r)} \ell^{\mathrm{C}}_{b,\zeta} \left(\theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_r}(r) \right) u \right\|_1 \\ &\leq \quad \left\| \nabla_{\theta^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}(r)} \ell^{\mathrm{C}}_{b,\zeta} \left(\theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_r}(r) \right) \right\|_\infty \|u\|_1 \\ &\leq \quad \left\| \nabla_{\theta^{\mathrm{C}}(r)} \ell^{\mathrm{C}}_{b,\zeta} \left(\theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_r}(r) \right) \right\|_\infty \sqrt{d^{\mathrm{C}}_r} \|u\|_2 \\ &< \quad \frac{(\lambda^{\mathrm{C}}_n \sqrt{d^{\mathrm{C}}_r})^2}{4} \mathcal{M}. \end{aligned}$$

Step 2: *For $r \in V^{\mathrm{C}}$, show that*

$$T_2 \quad \geq \quad \frac{\rho(\lambda^{\mathrm{C}}_n \sqrt{d^{\mathrm{C}}_r})^2 \mathcal{M}^2}{2} \quad \textit{for} \quad \|u\|_2 = \mathcal{M}\lambda^{\mathrm{C}}_n \sqrt{d^{\mathrm{C}}_r}. \tag{A.24}$$

Note that $\nabla^2_{\theta^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}(r)} \ell^{\mathrm{C}}_{b,\zeta} \left(\theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_r}(r) + vu \right)$ can be expressed as

$$\begin{aligned} &\nabla^2_{\theta^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}(r)} \ell^{\mathrm{C}}_{b,\zeta} \left(\theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_r}(r) + vu \right) \\ = \quad &\sum_{i=1}^{n} W^{(i)}_{b,\backslash r}(\zeta) \, W^{(i)\top}_{b,\backslash r}(\zeta) \, K^{\mathrm{C}''} \left(\theta^{\mathrm{C}}_{0;r} + W^{(i)\top}_{b,\backslash r}(\zeta) \left(\theta_{0;\mathcal{S}^{\mathrm{C}}_r} + vu \right) \right). \end{aligned} \tag{A.25}$$

Then applying the Taylor series expansion on $K^{C''}(\cdot)$ around $\theta_{0;\mathcal{S}_r^C} = 0$, then (A.25) can be re-written as

$$\begin{aligned}
&\nabla^2_{\theta^C_{\mathcal{S}_r^C}(r)} \ell^C_{b,\zeta}\left(\theta^C_{0;\mathcal{S}_r^C}(r) + vu\right) \\
= &\sum_{i=1}^{n} W^{(i)}_{b,\backslash r}(\zeta) W^{(i)^\top}_{b,\backslash r}(\zeta) K^{C''}\left(\theta^C_{0;r} + W^{(i)\top}_{b,\backslash r}(\zeta)\theta_{0;\mathcal{S}_r^C}\right) \\
&+ \sum_{i=1}^{n} W^{(i)}_{b,\backslash r}(\zeta) W^{(i)^\top}_{b,\backslash r}(\zeta) K^{C'''}(\bar{\eta}_C)\left(vuW^{(i)}_{b,\backslash r}(\zeta)\right),
\end{aligned}$$

where $\bar{\eta}_C$ lies on the "line segment" between $\theta_{0;\mathcal{S}_r^C}$ and $\theta_{0;\mathcal{S}_r^C} + vu$. Recall that ρ_ζ is the smallest eigenvalue of $Q_{\zeta,\mathcal{S}_r^C\mathcal{S}_r^C}$. Then by Conditions (A2), (A3), and (A5), we have

$$\begin{aligned}
T_2 &= u^\top \nabla^2_{\theta^C_{\mathcal{S}_r^C}(r)} \ell^C_{b,\zeta}\left(\theta^C_{0;\mathcal{S}_r^C}(r) + vu\right) u \\
&\geq \min_{u:\|u\|_2=\mathcal{B}} \left[u^\top \left\{\sum_{i=1}^{n} W^{(i)}_{b,\backslash r}(\zeta) W^{(i)^\top}_{b,\backslash r}(\zeta) K^{C''}\left(\theta^C_{0;r} + W^{(i)}_{b,\backslash r}(\zeta)\theta_{0;\mathcal{S}_r^C}\right)\right\} u\right] \\
&\quad + \min_{u:\|u\|_2=\mathcal{B}} \left[u^\top \left\{\sum_{i=1}^{n} W^{(i)}_{b,\backslash r}(\zeta) W^{(i)^\top}_{b,\backslash r}(\zeta) K^{C'''}(\bar{\eta}_C)\left(vuW^{(i)\top}_{b,\backslash r}(\zeta)\right)\right\} u\right] \\
&\geq \mathcal{B}^2\rho_\zeta \\
&\quad - \max_{u:\|u\|_2=\mathcal{B}} \left[u^\top \left\{\sum_{i=1}^{n} W^{(i)}_{b,\backslash r}(\zeta) W^{(i)^\top}_{b,\backslash r}(\zeta) K^{C'''}(\bar{\eta}_C)\left(vuW^{(i)}_{b,\backslash r}(\zeta)\right)\right\} u\right] \\
&\geq \mathcal{B}^2\rho - \mathcal{B}^3\rho_2\eta_2\kappa_1 \\
&\geq \frac{\rho(\lambda_n^C\sqrt{d_r^C})^2\mathcal{M}^2}{2}.
\end{aligned}$$

Step 3: *For $r \in V^C$, show that*

$$T_3 \geq -(\lambda_n^C\sqrt{d_r^C})^2\mathcal{M}^2 \quad \textit{for} \quad \|u\|_2 = \mathcal{M}\lambda_n^C\sqrt{d_r^C}. \tag{A.26}$$

Finally, for the last term T_3 in (A.21), applying the triangle inequality gives

$$\|\theta_{0;\mathcal{S}_r^C}\|_1 = \|\theta_{0;\mathcal{S}_r^C} + u - u\|_1 \leq \|\theta_{0;\mathcal{S}_r^C} + u\|_1 + \|u\|_1,$$

which implies

$$\|\theta_{0;\mathcal{S}_r^C} + u\|_1 - \|\theta_{0;\mathcal{S}_r^C}\|_1 \geq -\|u\|_1.$$

Therefore, we have

$$T_3 = \lambda_n^C\left(\|\theta_{0;\mathcal{S}_r^C} + u\|_1 - \|\theta_{0;\mathcal{S}_r^C}\|_1\right) \geq -(\lambda_n^C\sqrt{d_r^C})^2\mathcal{M}^2.$$

Step 4: *Establish (A.18).*

Therefore, for $r \in V^{\mathrm{C}}$, combining (A.23), (A.24), and (A.26) with (A.21) gives

$$\Phi(u) \geq (\lambda_n^{\mathrm{C}}\sqrt{d_r^{\mathrm{C}}})^2 \mathcal{M}\left(-\frac{1}{4}+\frac{\rho}{4}\mathcal{M}-1\right). \tag{A.27}$$

To ensure the right-hand-side of (A.27) be bounded below by zero, we must have

$$-\frac{1}{4}+\frac{\rho}{4}\mathcal{M}-1>0,$$

which is equivalent to $\mathcal{M} > \frac{5}{\rho}$. We take $\mathcal{M}^* = \frac{6}{\rho}$, and thus, $\mathcal{B} = \mathcal{M}^*\lambda_n^{\mathrm{C}}\sqrt{d_r^{\mathrm{C}}} = \frac{6\sqrt{d_r^{\mathrm{C}}}\lambda_n^{\mathrm{C}}}{\rho}$ and (A.20) holds. As a result, (A.18) is shown. □

Lemma A.4. *For $r \in V^C$ and $r' \in V^D$, let*

$$R_n^C = \left\{\nabla^2_{\theta^C(r)}\ell^C_{b,\zeta}\left(\bar{\theta}_C\right) - \nabla^2_{\theta^C(r)}\ell^C_{b,\zeta}\left(\theta_0^C(r)\right)\right\}\left\{\widehat{\theta}^C(r;\zeta,b) - \theta_0^C(r)\right\} \tag{A.28}$$

and

$$R_n^D = \left\{\nabla^2_{\theta^D(r')}\ell^D_{b,\zeta}\left(\bar{\theta}_D\right) - \nabla^2_{\theta^D(r')}\ell^D_{b,\zeta}\left(\theta_0^D(r')\right)\right\}\left\{\widehat{\theta}^D(r';\zeta,b) - \theta_0^D(r')\right\}, \tag{A.29}$$

where $\bar{\theta}_C$ lies on the "line segment" between $\widehat{\theta}^C(r;\zeta,b)$ and $\theta_0^C(r)$, and $\bar{\theta}_D$ lies on the "line segment" between $\widehat{\theta}^D(r';\zeta,b)$ and $\theta_0^D(r')$. Then under regularity conditions in Section 5.1, we have

$$\|R_n^C\|_\infty \leq \frac{72\eta_1\rho_2 d_r^C \lambda_n^{C2}}{\rho^2} \quad \textit{and} \quad \|R_n^D\|_\infty \leq \frac{72\eta_1\rho_2 d_{r'}^D \lambda_n^{D2}}{\rho^2}.$$

Proof:

Note that $\|R_n^{\mathrm{C}}\|_\infty$ and $\|R_n^{\mathrm{D}}\|_\infty$ have the similar derivations, so we only demonstrate the derivations for $\|R_n^{\mathrm{C}}\|_\infty$ in the following proof. Similar proofs give the result of $\|R_n^{\mathrm{D}}\|_\infty$.

Since $\nabla^2_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\theta^{\mathrm{C}}(r)\right) = \sum_{i=1}^n W^{(i)}_{b,\backslash r}(\zeta)\, W^{(i)^\top}_{b,\backslash r}(\zeta)\, K^{\mathrm{C}''}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)\right)$, then

$$\begin{aligned}
&\nabla^2_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\bar{\theta}_{\mathrm{C}}\right) - \nabla^2_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\theta_0^{\mathrm{C}}(r)\right)\\
=\ &\sum_{i=1}^n W^{(i)}_{b,\backslash r}(\zeta)\, W^{(i)^\top}_{b,\backslash r}(\zeta)\left\{K^{\mathrm{C}''}\left(\bar{\eta}_r^{\mathrm{C}(i)}(b,\zeta)\right) - K^{\mathrm{C}''}\left(\eta_r^{\mathrm{C}(i)}(b,\zeta)\right)\right\}
\end{aligned}$$

for $r \in V^{\mathrm{C}}$, where $\bar{\eta}_r^{\mathrm{C}(i)}(b,\zeta)$ is determined by $\eta_r^{\mathrm{C}(i)}(b,\zeta)$ with $\theta^{\mathrm{C}}(r)$ replaced by $\bar{\theta}_{\mathrm{C}}$. By Assumptions (A2) and (A3), the maximum eigenvalue of $\nabla^2_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\bar{\theta}_{\mathrm{C}}\right) - \nabla^2_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\theta_0^{\mathrm{C}}(r)\right)$ is

$$\Lambda_{\max}\left\{\nabla^2_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\bar{\theta}_{\mathrm{C}}\right) - \nabla^2_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\theta_0^{\mathrm{C}}(r)\right)\right\}$$

$$
\begin{aligned}
&= \max_{\xi:\|\xi\|_2=1} \xi^\top \left\{ \nabla^2_{\theta^{\mathrm{C}}(r)} \ell^{\mathrm{C}}_{b,\zeta} \left(\bar{\theta}_{\mathrm{C}}\right) - \nabla^2_{\theta^{\mathrm{C}}(r)} \ell^{\mathrm{C}}_{b,\zeta} \left(\theta^{\mathrm{C}}_0(r)\right) \right\} \xi \\
&\leq \max_{\xi:\|\xi\|_2=1} \xi^\top \left(\sum_{i=1}^{n} W^{(i)}_{b,\setminus r}(\zeta) \, W^{(i)^\top}_{b,\setminus r}(\zeta) \right) \xi \\
&\quad \times \xi^\top \left| K^{\mathrm{C}''} \left(\bar{\eta}^{\mathrm{C}(i)}_r(b,\zeta) \right) - K^{\mathrm{C}''} \left(\eta^{\mathrm{C}(i)}_r(b,\zeta) \right) \right| \xi \\
&\leq 2\eta_1 \rho_2. \qquad \text{(A.30)}
\end{aligned}
$$

As a result, by Lemma A.3 and (A.30), we have

$$
\begin{aligned}
\|R^{\mathrm{C}}_n\|_1 &\leq \|R^{\mathrm{C}}_n\|_2^2 \\
&\leq \Lambda_{\max} \left\{ \nabla^2_{\theta^{\mathrm{C}}(r)} \ell^{\mathrm{C}}_{b,\zeta} \left(\bar{\theta}_{\mathrm{C}}\right) - \nabla^2_{\theta^{\mathrm{C}}(r)} \ell^{\mathrm{C}}_{b,\zeta} \left(\theta^{\mathrm{C}}_0(r)\right) \right\} \times \left\| \widehat{\theta}^{\mathrm{C}}(r;\zeta,b) - \theta^{\mathrm{C}}_0(r) \right\|_2^2 \\
&\leq 2\eta_1\rho_2 \left(\frac{6\sqrt{d^{\mathrm{C}}_r}\lambda^{\mathrm{C}}_n}{\rho} \right)^2 \\
&= \frac{72\eta_1\rho_2 d^{\mathrm{C}}_r \lambda^{\mathrm{C}2}_n}{\rho^2},
\end{aligned}
$$

and thus the proof completes. □

Appendix B: Proof of Theorem 3.1

In contrast to the naive negative log likelihood function, we first consider the negative log likelihood function based on true random variables for $r \in V^{\mathrm{C}}$:

$$
\ell(\theta(r)) = -\frac{1}{n} \sum_{i=1}^{n} \log \left\{ P\left(X^{(i)}_r | X^{(i)}_{\setminus r} \right) \right\},
$$

where $P\left(X^{(i)}_r | X^{(i)}_{\setminus r} \right)$ is defined in (22). Similar to (24), the estimator based on the true random vector $X^{(i)}$ is given by

$$
\widetilde{\theta}(r) = \underset{\theta(r)}{\operatorname{argmin}} \left\{ \ell(\theta(r)) + \lambda_n \left\| \theta_{\setminus r} \right\|_1 \right\} \qquad \text{(B.1)}
$$

with $\widetilde{\theta}(r) = \left(\widetilde{\theta}_r, \widetilde{\theta}^\top_{\setminus r} \right)^\top$ for $r \in V^{\mathrm{C}}$. To ease the notation, let $\widetilde{\theta}$, $\widehat{\theta}_{nv}$, θ and θ_0 denote $\widetilde{\theta}(r)$, $\widehat{\theta}_{nv}(r)$, $\theta(r)$ and $\theta_0(r)$, respectively.

Let $\widetilde{\theta}_{rt}$ denote the tth component in $\widetilde{\theta}_{\setminus r}$. Let $\widetilde{z} = \left(\widetilde{z}_r, \widetilde{z}^\top_{\setminus r} \right)^\top$ be a p-dimensional vector with the tth component in $\widetilde{z}_{\setminus r}$ being $\widetilde{z}_t = \operatorname{sign}\left(\widetilde{\theta}_{rt} \right)$ if $\widetilde{\theta}_{rt} \neq 0$ and $|\widetilde{z}_t| \leq 1$ otherwise, while $\widetilde{z}_r$, corresponding to θ_r, is set to zero since the nodewise term θ_r is not penalized in (B.1). In addition, let $\widehat{z}_{nv}$ denote a p-dimensional vector which is defined similar to $\widetilde{z}$ but corresponds to $\widehat{\theta}_{nv}$. Then by the Karush–Kuhn–Tucker (KKT) conditions, we have

$$
\nabla_\theta \ell_{nv} \left(\widehat{\theta}_{nv} \right) + \lambda_n \widehat{z}_{nv} = 0 \qquad \text{(B.2)}
$$

and

$$\nabla_\theta \ell\left(\widetilde{\theta}\right) + \lambda_n \widetilde{z} = 0. \tag{B.3}$$

By the first order Taylor series expansion on $\nabla_\theta \ell_{nv}\left(\widehat{\theta}_{nv}\right)$ and $\nabla_\theta \ell\left(\widetilde{\theta}\right)$ around θ_0, we have

$$\nabla_\theta \ell_{nv}\left(\widehat{\theta}_{nv}\right) \approx \nabla_\theta \ell_{nv}\left(\theta_0\right) + \nabla_\theta^2 \ell_{nv}\left(\theta_0\right)\left(\widehat{\theta}_{nv} - \theta_0\right) \tag{B.4}$$

and

$$\nabla_\theta \ell\left(\widetilde{\theta}\right) \approx \nabla_\theta \ell\left(\theta_0\right) + \nabla_\theta^2 \ell\left(\theta_0\right)\left(\widetilde{\theta} - \theta_0\right). \tag{B.5}$$

Combining (B.4) and (B.5) yields

$$\begin{aligned} & \nabla_\theta \ell_{nv}\left(\widehat{\theta}_{nv}\right) - \nabla_\theta \ell\left(\widetilde{\theta}\right) \\ \approx\ & \left\{\nabla_\theta \ell_{nv}\left(\theta_0\right) - \nabla_\theta \ell\left(\theta_0\right)\right\} + \nabla_\theta^2 \ell_{nv}\left(\theta_0\right)\widehat{\theta}_{nv} - \nabla_\theta^2 \ell\left(\theta_0\right)\widetilde{\theta} \\ & - \left\{\nabla_\theta^2 \ell_{nv}\left(\theta_0\right) - \nabla_\theta^2 \ell\left(\theta_0\right)\right\}\theta_0. \end{aligned} \tag{B.6}$$

The second order derivative of $\ell_{nv}(\theta_0)$ and $\ell(\theta_0)$ can be, respectively, expressed as

$$\nabla_\theta^2 \ell_{nv}\left(\theta_0\right) = \frac{1}{n}\sum_{i=1}^{n} X_{\setminus r}^{*(i)} X_{\setminus r}^{*(i)\top} K^{c''}\left(\theta_{0;r} + X_{\setminus r}^{*(i)\top}\theta_{0;\setminus r}\right)$$

and

$$\nabla_\theta^2 \ell\left(\theta_0\right) = \frac{1}{n}\sum_{i=1}^{n} X_{\setminus r}^{(i)} X_{\setminus r}^{(i)\top} K^{c''}\left(\theta_{0;r} + X_{\setminus r}^{(i)\top}\theta_{0;\setminus r}\right).$$

Since $X_{\setminus r}^{*(i)} \left| X^{(i)} \right. \sim N\left(X_{\setminus r}^{(i)}, \Sigma_{\epsilon;\setminus r}\right)$, we have $E\left(X_{\setminus r}^{*(i)} X_{\setminus r}^{*(i)\top} \middle| X^{(i)}\right) = \Sigma_{\epsilon;\setminus r} + X_{\setminus r}^{(i)} X_{\setminus r}^{(i)\top}$. Hence, we have

$$\begin{aligned} & E\left\{\nabla_\theta^2 \ell_{nv}\left(\theta_0\right) \middle| X_{\setminus r}\right\} \\ =\ & \frac{1}{n}\sum_{i=1}^{n}\left\{\left(X_{\setminus r}^{(i)} X_{\setminus r}^{(i)\top} + \Sigma_{\epsilon;\setminus r}\right) K^{c''}\left(\theta_{0;r} + X_{\setminus r}^{(i)\top}\theta_{0;\setminus r}\right)\right\} \end{aligned} \tag{B.7}$$

and

$$E\left\{\nabla_\theta^2 \ell\left(\theta_0\right) \middle| X_{\setminus r}\right\} = \frac{1}{n}\sum_{i=1}^{n}\left\{\left(X_{\setminus r}^{(i)} X_{\setminus r}^{(i)\top}\right) K^{c''}\left(\theta_{0;r} + X_{\setminus r}^{(i)\top}\theta_{0;\setminus r}\right)\right\}. \tag{B.8}$$

Therefore, by the Law of Large Numbers with (B.7) and (B.8), we have that as $n \to \infty$,

$$\nabla_\theta^2 \ell_{nv}\left(\theta_0\right) \xrightarrow{p} \mathcal{Q}_{nv} \quad \text{and} \quad \nabla_\theta^2 \ell\left(\theta_0\right) \xrightarrow{p} \mathcal{Q}_r, \tag{B.9}$$

where

$$\mathcal{Q}_{nv} = E\left\{\left(X_{\setminus r}^{(i)} X_{\setminus r}^{(i)\top} + \Sigma_{\epsilon;\setminus r}\right) K^{\mathrm{C}''}\left(\theta_{0;r} + X_{\setminus r}^{(i)\top}\theta_{0;\setminus r}\right)\right\} \tag{B.10}$$

and

$$\mathcal{Q}_{r} = E\left\{\left(X_{\setminus r}^{(i)} X_{\setminus r}^{(i)\top}\right) K^{\mathrm{C}''}\left(\theta_{0;r} + X_{\setminus r}^{(i)\top}\theta_{0;\setminus r}\right)\right\}. \tag{B.11}$$

Then the relationship between (B.10) and (B.11) is determined by

$$\mathcal{Q}_{nv} = \mathcal{Q}_{r} + \Sigma_{\epsilon;\setminus r}\mathcal{D}_{r}, \tag{B.12}$$

where $\mathcal{D}_r = E\left\{K^{\mathrm{C}''}\left(\theta_{0;r} + X_{\setminus r}^{(i)\top}\theta_{0;\setminus r}\right)\right\}$. On the other hand, by (B.2) and (B.3), we have

$$\nabla_\theta \ell_{nv}\left(\widehat{\theta}_{nv}\right) - \nabla_\theta \ell\left(\widetilde{\theta}\right) = -\lambda_n\left(\widehat{z}_{nv} - \widetilde{z}\right). \tag{B.13}$$

Thus, combining (B.9), (B.12), and (B.13) with (B.6) gives

$$\begin{aligned} & -\lambda_n\left(\widehat{z}_{nv} - \widetilde{z}\right) \\ \approx\ & \left\{\nabla_\theta \ell_{nv}\left(\theta_0\right) - \nabla_\theta \ell\left(\theta_0\right)\right\} + \left(\mathcal{Q}_r + \Sigma_{\epsilon;\setminus r}\mathcal{D}_r\right)\widehat{\theta}_{nv} - \mathcal{Q}_r\widetilde{\theta} - \Sigma_{\epsilon;\setminus r}\mathcal{D}_r\theta_0 \\ =\ & \left\{\nabla_\theta \ell_{nv}\left(\theta_0\right) - \nabla_\theta \ell\left(\theta_0\right)\right\} + \mathcal{Q}_r\left(\widehat{\theta}_{nv} - \widetilde{\theta}\right) + \Sigma_{\epsilon;\setminus r}\mathcal{D}_r\left(\widehat{\theta}_{nv} - \theta_0\right). \end{aligned} \tag{B.14}$$

By the triangle inequality, $\|\widehat{z}_{nv} - \widetilde{z}\|_\infty \le \|\widehat{z}_{nv}\|_\infty + \|\widetilde{z}\|_\infty < 2$. Besides, by (B.14), we have

$$\begin{aligned} \left\|\nabla_\theta \ell_{nv}\left(\theta_0\right) - \nabla_\theta \ell\left(\theta_0\right)\right\|_\infty \ \le\ & 2\lambda_n + \left\|\mathcal{Q}_r\right\|_\infty \left\|\widehat{\theta}_{nv} - \widetilde{\theta}\right\|_\infty \\ & + \left\|\Sigma_{\epsilon;\setminus r}\mathcal{D}_r\right\|_\infty \left\|\widehat{\theta}_{nv} - \theta_0\right\|_\infty, \end{aligned} \tag{B.15}$$

and thus rearranging (B.15) gives

$$\begin{aligned} \left\|\widehat{\theta}_{nv} - \widetilde{\theta}\right\|_\infty \ \ge\ & \left\|\mathcal{Q}_r\right\|_\infty^{-1}\left(\left\|\nabla_\theta \ell_{nv}\left(\theta_0\right) - \nabla_\theta \ell\left(\theta_0\right)\right\|_\infty - 2\lambda_n\right) \\ & - \left\|\mathcal{Q}_r\right\|_\infty^{-1}\left\|\Sigma_{\epsilon;\setminus r}\mathcal{D}_r\right\|_\infty \left\|\widehat{\theta}_{nv} - \theta_0\right\|_\infty. \end{aligned} \tag{B.16}$$

Noting that based on true random variables, Lemmas 9 and 10 in [28] show that with suitable range of λ_n, there exist some constants $\widetilde{\alpha} \in (0,1)$ and $\widetilde{\rho} > 0$, such that

$$\left\|\nabla_\theta \ell\left(\theta_0\right)\right\|_\infty \le \frac{\lambda_n \widetilde{\alpha}}{4\left(2 - \widetilde{\alpha}\right)} \tag{B.17}$$

and

$$\left\|\widetilde{\theta} - \theta_0\right\|_\infty \le \widetilde{\rho}\lambda_n \tag{B.18}$$

with large probabilities.

Finally, applying the triangle inequality on $\left\|\widehat{\theta}_{nv}-\theta_0\right\|_\infty$, we have

$$\left\|\widehat{\theta}_{nv}-\theta_0\right\|_\infty \geq \left\|\widehat{\theta}_{nv}-\widetilde{\theta}\right\|_\infty - \left\|\widetilde{\theta}-\theta_0\right\|_\infty,$$

and thus, implementing (B.16), (B.17) and (B.18) gives

$$\begin{aligned}\left\|\widehat{\theta}_{nv}-\theta_0\right\|_\infty \geq\ & \|\mathcal{Q}_r\|_\infty^{-1}\left(\|\nabla_\theta \ell_{nv}(\theta_0)\|_\infty - \frac{\lambda_n\widetilde{\alpha}}{4(2-\widetilde{\alpha})} - 2\lambda_n\right)\\ & - \|\mathcal{Q}_r\|_\infty^{-1}\left\|\Sigma_{\epsilon;\setminus r}\mathcal{D}_r\right\|_\infty\left\|\widehat{\theta}_{nv}-\theta_0\right\|_\infty - \widetilde{\rho}\lambda_n.\end{aligned}$$

Consequently, we have

$$\begin{aligned}&\left\|\widehat{\theta}_{nv}-\theta_0\right\|_\infty\\ \geq\ & \left\{1+\|\mathcal{Q}_r\|_\infty^{-1}\left\|\Sigma_{\epsilon;\setminus r}\mathcal{D}_r\right\|_\infty\right\}^{-1}\|\mathcal{Q}_r\|_\infty^{-1}\\ & \times\left(\|\nabla_\theta \ell_{nv}(\theta_0)\|_\infty - \frac{\lambda_n\widetilde{\alpha}}{4(2-\widetilde{\alpha})} - 2\lambda_n\right)\\ & - \left\{1+\|\mathcal{Q}_r\|_\infty^{-1}\left\|\Sigma_{\epsilon;\setminus r}\mathcal{D}_r\right\|_\infty\right\}^{-1}\widetilde{\rho}\lambda_n\\ =\ & \left\{\|\mathcal{Q}_r\|_\infty+\left\|\Sigma_{\epsilon;\setminus r}\mathcal{D}_r\right\|_\infty\right\}^{-1}\left(\|\nabla_\theta \ell_{nv}(\theta_0)\|_\infty - \frac{\lambda_n\widetilde{\alpha}}{4(2-\widetilde{\alpha})} - 2\lambda_n\right)\\ & - \left\{1+\|\mathcal{Q}_r\|_\infty^{-1}\left\|\Sigma_{\epsilon;\setminus r}\mathcal{D}_r\right\|_\infty\right\}^{-1}\widetilde{\rho}\lambda_n. \qquad \text{(B.19)}\end{aligned}$$

Since (B.19) holds under the range of λ_n specified in [28, pp.3839], the value λ_n taken in such range ensures that the right-hand side of (B.19) is nonnegative. Thus, desired inequality is obtained, and the proof completes. □

Appendix C: Proof of Theorem 5.1

Since the proof of $\widehat{\mathcal{N}}^{\mathrm{D}}(r')=\mathcal{N}^{\mathrm{D}}(r')$ and $\widehat{\mathcal{N}}^{\mathrm{DC}}(r')=\mathcal{N}^{\mathrm{DC}}(r')$ for $r'\in V^{\mathrm{D}}$ is similar to the derivation for $\widehat{\mathcal{N}}^{\mathrm{C}}(r)=\mathcal{N}^{\mathrm{C}}(r)$ and $\widehat{\mathcal{N}}^{\mathrm{CD}}(r)=\mathcal{N}^{\mathrm{CD}}(r)$ for $r\in V^{\mathrm{C}}$, we only show the latter results in the following derivations that consist of three steps.

Step 1: *For $r\in V^C$, let $\widehat{\theta}^C_{rt}(\zeta,b)$ denote the tth component of $\widehat{\theta}^C_{\setminus r}(\zeta,b)$ and let $\widehat{\theta}^{CD}_{rt'}(\zeta,b)$ denote the t'th component of $\widehat{\theta}^{CD}_{r}(\zeta,b)$. Examine*

$$\widehat{\mathcal{N}}^C_b(r;\zeta)=\left\{t\in V^C\setminus\{r\}:\widehat{\theta}^C_{rt}(\zeta,b)\neq 0\right\}$$

and

$$\widehat{\mathcal{N}}^{CD}_b(r;\zeta)=\left\{t'\in V^D:\widehat{\theta}^{CD}_{rt'}(\zeta,b)\neq 0\right\},$$

and show that

$$\widehat{\mathcal{N}}^C_b(r;\zeta)=\mathcal{N}^C(r) \quad \textit{and} \quad \widehat{\mathcal{N}}^{CD}_b(r;\zeta)=\mathcal{N}^{CD}(r) \qquad \text{(C.1)}$$

with probability greater than $1-\left\{c_1\{\max(n,p)\}^{-2}+\exp(-c_2 n)\right\}$, *where* $\mathcal{N}^{C}(r)$ *and* $\mathcal{N}^{CD}(r)$ *for* $r \in V^{C}$ *are the neighbourhood sets defined in Section 2.2.*

In the proof of Lemma A.3, we write

$$\widehat{\theta}^{\mathrm{C}}(r;\zeta,b)=\left(\widehat{\theta}^{\mathrm{C}\top}_{\mathcal{S}_r^{\mathrm{C}}}(r;\zeta,b),\widehat{\theta}^{\mathrm{C}\top}_{\overline{\mathcal{S}}_r^{\mathrm{C}}}(\zeta,b)\right)^{\top}$$

with

$$\widehat{\theta}^{\mathrm{C}}_{\mathcal{S}_r^{\mathrm{C}}}(r;\zeta,b)=\left(\widehat{\theta}^{\mathrm{C}}_{r}(\zeta,b),\widehat{\theta}^{\mathrm{C}\top}_{\mathcal{S}_r^{\mathrm{C}}}(\zeta,b)\right)^{\top}.$$

Let $\widehat{z}^{\mathrm{C}}=\left(\widehat{z}_r,\widehat{z}^{\top}_{\setminus r}\right)^{\top}$ be a p-dimensional vector with the tth component in $\widehat{z}_{\setminus r}$ being $\widehat{z}_t=\operatorname{sign}\left(\widehat{\theta}_{rt}(\zeta,b)\right)$ if $\widehat{\theta}_{rt}(\zeta,b)\neq 0$ and $|\widehat{z}_t|\leq 1$ otherwise, while $\widehat{z}_r$, corresponding to θ_r^{C}, is set to zero since the nodewise term θ_r^{C} is not penalized in (28). To show the sparsity recovery, we consider the primal dual witness (PDW) method (e.g., [16, pp.307]). The strategy of the PDW method is to

(i) $\widehat{\theta}^{\mathrm{C}}_{\overline{\mathcal{S}}_r^{\mathrm{C}}}(\zeta,b)=0_{p-d_r^{\mathrm{C}}-1}$ and
$\widehat{\theta}^{\mathrm{C}}_{\mathcal{S}_r^{\mathrm{C}}}(r;\zeta,b)=\operatorname{argmin}_{\theta_{\mathrm{C};\mathcal{S}_r^{\mathrm{C}}}(r)}\left\{\ell^{\mathrm{C}}_{b,\zeta}\left(\theta^{\mathrm{C}}(r)\right)+\lambda_n^{\mathrm{C}}\left\|\theta^{\mathrm{C}}_{\mathcal{S}_r^{\mathrm{C}}}\right\|_1\right\}$;

(ii) write $\widehat{z}^{\mathrm{C}}=\left(\widehat{z}^{\mathrm{C}\top}_{\mathcal{S}_r^{\mathrm{C}}},\widehat{z}^{\mathrm{C}\top}_{\overline{\mathcal{S}}_r^{\mathrm{C}}}\right)^{\top}$ corresponding to the components of $\widehat{\theta}^{\mathrm{C}}_{\mathcal{S}_r^{\mathrm{C}}}(r;\zeta,b)$ and $\widehat{\theta}^{\mathrm{C}}_{\overline{\mathcal{S}}_r^{\mathrm{C}}}(\zeta,b)$;

(iii) show that

$$\left\|\widehat{z}^{\mathrm{C}}_{\overline{\mathcal{S}}_r^{\mathrm{C}}}\right\|_{\infty}\leq 1. \tag{C.2}$$

Indeed, as discussed in Lemma 11.2 of [16, pp.307], if (C.2) is true, then $\widehat{\theta}^{\mathrm{C}}(r;\zeta,b)=\left(\widehat{\theta}^{\mathrm{C}\top}_{\mathcal{S}_r^{\mathrm{C}}}(r;\zeta,b),0^{\top}_{p-d_r^{\mathrm{C}}-1}\right)^{\top}$ is an optimal solution of (28), and thus, (C.1) holds with probability approaching one (e.g., [16, Theorem 11.3]). So, the remaining task is to show (C.2).

By the KKT conditions, we have

$$\nabla_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\widehat{\theta}^{\mathrm{C}}(r;\zeta,b)\right)+\lambda_n^{\mathrm{C}}\widehat{z}^{\mathrm{C}}=0. \tag{C.3}$$

Adding $-\nabla_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\theta_0^{\mathrm{C}}(r)\right)$ to the both sides of (C.3) gives

$$\begin{aligned}&\nabla_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\widehat{\theta}^{\mathrm{C}}(r;\zeta,b)\right)-\nabla_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\theta_0^{\mathrm{C}}(r)\right)\\ =\;&-\lambda_n^{\mathrm{C}}\widehat{z}^{\mathrm{C}}-\nabla_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\theta_0^{\mathrm{C}}(r)\right).\end{aligned} \tag{C.4}$$

By the Mean Value Theorem (MVT), there exists $\bar{\theta}_{\mathrm{C}}$ which lies on the "line segment" between $\widehat{\theta}^{\mathrm{C}}(r;\zeta,b)$ and $\theta_0^{\mathrm{C}}(r)$, such that

$$\nabla^2_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\bar{\theta}_{\mathrm{C}}\right)\left\{\widehat{\theta}^{\mathrm{C}}(r;\zeta,b)-\theta_0^{\mathrm{C}}(r)\right\}=-\lambda_n^{\mathrm{C}}\widehat{z}^{\mathrm{C}}-\nabla_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\theta_0^{\mathrm{C}}(r)\right).$$

Adding $\nabla^2_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\theta^{\mathrm{C}}_0(r)\right)\left\{\widehat{\theta}^{\mathrm{C}}(r;\zeta,b)-\theta^{\mathrm{C}}_0(r)\right\}$ to both sides of (C.4) yields

$$\begin{aligned}
&\nabla^2_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\theta^{\mathrm{C}}_0(r)\right)\left\{\widehat{\theta}^{\mathrm{C}}(r;\zeta,b)-\theta^{\mathrm{C}}_0(r)\right\}\\
=\;& -\lambda^{\mathrm{C}}_n\widehat{z}^{\mathrm{C}}-\nabla_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\theta^{\mathrm{C}}_0(r)\right)-\Big[\nabla^2_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\bar{\theta}_{\mathrm{C}}\right)\left\{\widehat{\theta}^{\mathrm{C}}(r;\zeta,b)-\theta^{\mathrm{C}}_0(r)\right\}\\
&-\nabla^2_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\theta^{\mathrm{C}}_0(r)\right)\left\{\widehat{\theta}^{\mathrm{C}}(r;\zeta,b)-\theta^{\mathrm{C}}_0(r)\right\}\Big]\\
\triangleq\;& -\lambda^{\mathrm{C}}_n\widehat{z}^{\mathrm{C}}-Y^{\mathrm{C}}-R^{\mathrm{C}}_n, \qquad \text{(C.5)}
\end{aligned}$$

where R^{C}_n is defined (A.28), and $Y^{\mathrm{C}}=\nabla_{\theta^{\mathrm{C}}(r)}\ell^{\mathrm{C}}_{b,\zeta}\left(\theta^{\mathrm{C}}_0(r)\right)$.

Let $Y^{\mathrm{C}}=\left(Y^{\top}_{\mathcal{S}^{\mathrm{C}}_r},Y^{\top}_{\overline{\mathcal{S}}^{\mathrm{C}}_r}\right)^{\top}$ and $R^{\mathrm{C}}_n=\left(R^{\mathrm{C}\top}_{n;\mathcal{S}^{\mathrm{C}}_r},R^{\mathrm{C}\top}_{n;\overline{\mathcal{S}}^{\mathrm{C}}_r}\right)^{\top}$ Now, by (C.5) and (i), we have

$$\begin{aligned}
&\begin{pmatrix} Q_{\zeta,\mathcal{S}^{\mathrm{C}}_r\mathcal{S}^{\mathrm{C}}_r} & Q_{\zeta,\mathcal{S}^{\mathrm{C}}_r\overline{\mathcal{S}}^{\mathrm{C}}_r}\\ Q_{\zeta,\overline{\mathcal{S}}^{\mathrm{C}}_r\mathcal{S}^{\mathrm{C}}_r} & Q_{\zeta,\overline{\mathcal{S}}^{\mathrm{C}}_r\overline{\mathcal{S}}^{\mathrm{C}}_r}\end{pmatrix}\begin{pmatrix}\widehat{\theta}^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}(r;\zeta,b)-\theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_r}(r)\\ 0\end{pmatrix}\\
=\;& -\lambda^{\mathrm{C}}_n\begin{pmatrix}\widehat{z}^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}\\ \widehat{z}^{\mathrm{C}}_{\overline{\mathcal{S}}^{\mathrm{C}}_r}\end{pmatrix}-\begin{pmatrix}Y_{\mathcal{S}^{\mathrm{C}}_r}\\ Y_{\overline{\mathcal{S}}^{\mathrm{C}}_r}\end{pmatrix}-\begin{pmatrix}R^{\mathrm{C}}_{n;\mathcal{S}^{\mathrm{C}}_r}\\ R^{\mathrm{C}}_{n;\overline{\mathcal{S}}^{\mathrm{C}}_r}\end{pmatrix},
\end{aligned}$$

and it implies that

$$Q_{\zeta,\mathcal{S}^{\mathrm{C}}_r\mathcal{S}^{\mathrm{C}}_r}\left\{\widehat{\theta}^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}(r;\zeta,b)-\theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_r}(r)\right\}=\lambda^{\mathrm{C}}_n\widehat{z}^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}-Y_{\mathcal{S}^{\mathrm{C}}_r}-R^{\mathrm{C}}_{n;\mathcal{S}^{\mathrm{C}}_r}, \qquad \text{(C.6)}$$

and

$$Q_{\zeta,\overline{\mathcal{S}}^{\mathrm{C}}_r\mathcal{S}^{\mathrm{C}}_r}\left\{\widehat{\theta}^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}(r;\zeta,b)-\theta_{0;\mathcal{S}^{\mathrm{C}}_r}(r)\right\}=\lambda^{\mathrm{C}}_n\widehat{z}^{\mathrm{C}}_{\overline{\mathcal{S}}^{\mathrm{C}}_r}-Y_{\overline{\mathcal{S}}^{\mathrm{C}}_r}-R^{\mathrm{C}}_{n;\overline{\mathcal{S}}^{\mathrm{C}}_r}. \qquad \text{(C.7)}$$

Combining (C.6) and (C.7) yields

$$\begin{aligned}
&Q_{\zeta,\overline{\mathcal{S}}^{\mathrm{C}}_r\mathcal{S}^{\mathrm{C}}_r}Q^{-1}_{\zeta,\mathcal{S}^{\mathrm{C}}_r\mathcal{S}^{\mathrm{C}}_r}\left(-\lambda^{\mathrm{C}}_n\widehat{z}^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}-Y_{\mathcal{S}^{\mathrm{C}}_r}-R^{\mathrm{C}}_{n;\mathcal{S}^{\mathrm{C}}_r}\right)\\
=\;& -\lambda^{\mathrm{C}}_n\widehat{z}^{\mathrm{C}}_{\overline{\mathcal{S}}^{\mathrm{C}}_r}-Y_{\overline{\mathcal{S}}^{\mathrm{C}}_r}-R^{\mathrm{C}}_{n;\overline{\mathcal{S}}^{\mathrm{C}}_r} \qquad \text{(C.8)}
\end{aligned}$$

and thus our target $\widehat{z}^{\mathrm{C}}_{\overline{\mathcal{S}}^{\mathrm{C}}_r}$ can be expressed as

$$\begin{aligned}
\widehat{z}^{\mathrm{C}}_{\overline{\mathcal{S}}^{\mathrm{C}}_r} =\;& \frac{1}{\lambda^{\mathrm{C}}_n}\Big\{Q_{\zeta,\overline{\mathcal{S}}^{\mathrm{C}}_r\mathcal{S}^{\mathrm{C}}_r}Q^{-1}_{\zeta,\mathcal{S}^{\mathrm{C}}_r\mathcal{S}^{\mathrm{C}}_r}\left(\lambda^{\mathrm{C}}_n\widehat{z}^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}+Y_{\mathcal{S}^{\mathrm{C}}_r}+R^{\mathrm{C}}_{n;\mathcal{S}^{\mathrm{C}}_r}\right)\\
&-Y_{\overline{\mathcal{S}}^{\mathrm{C}}_r}-R^{\mathrm{C}}_{n;\overline{\mathcal{S}}^{\mathrm{C}}_r}\Big\}. \qquad \text{(C.9)}
\end{aligned}$$

We now show (C.2). Given

$$\lambda^{\mathrm{C}}_n<\frac{\rho^2}{\eta_1\rho^*_2d^{\mathrm{C}}} \qquad \text{(C.10)}$$

with $\rho_2^* = 288\rho_2$, then (C.9) gives that

$$\begin{aligned}
\|\widehat{z}^{\mathrm{C}}_{\overline{\mathcal{S}}^{\mathrm{C}}_r}\|_\infty &\leq \frac{1}{\lambda^{\mathrm{C}}_n}\Big\{ \left\|Q_{\zeta,\overline{\mathcal{S}}^{\mathrm{C}}_r\mathcal{S}^{\mathrm{C}}_r} Q^{-1}_{\zeta,\mathcal{S}^{\mathrm{C}}_r\mathcal{S}^{\mathrm{C}}_r}\right\|_\infty \left(\lambda^{\mathrm{C}}_n \left\|\widehat{z}^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}\right\|_\infty + \|Y^{\mathrm{C}}\|_\infty + \|R^{\mathrm{C}}_n\|_\infty\right) \\
&\quad + \|Y^{\mathrm{C}}\|_\infty + \|R^{\mathrm{C}}_n\|_\infty \Big\} \\
&\leq \frac{1}{\lambda^{\mathrm{C}}_n}\Big\{ (1-\alpha)\left(\lambda^{\mathrm{C}}_n + \frac{\alpha\lambda^{\mathrm{C}}_n}{8-4\alpha} + \frac{72\eta_1\rho_2 d^{\mathrm{C}}\lambda^{\mathrm{C}2}_n}{\rho^2}\right) + \frac{\alpha\lambda^{\mathrm{C}}_n}{8-4\alpha} \\
&\quad + \frac{72\eta_1\rho_2 d^{\mathrm{C}}\lambda^{\mathrm{C}2}_n}{\rho^2}\Big\} \\
&= 1 - \frac{\alpha}{2} \\
&\leq 1, \qquad \text{(C.11)}
\end{aligned}$$

where the first step is due to that $\left\|Y_{\mathcal{S}^{\mathrm{C}}_r}\right\|_\infty \leq \|Y^{\mathrm{C}}\|_\infty$, $\left\|Y_{\overline{\mathcal{S}}^{\mathrm{C}}_r}\right\|_\infty \leq \|Y^{\mathrm{C}}\|_\infty$, $\left\|R^{\mathrm{C}}_{n;\mathcal{S}^{\mathrm{C}}_r}\right\|_\infty \leq \|R^{\mathrm{C}}_n\|_\infty$ and $\left\|R^{\mathrm{C}}_{n;\overline{\mathcal{S}}^{\mathrm{C}}_r}\right\|_\infty \leq \|R^{\mathrm{C}}_n\|_\infty$, the second step comes from Assumption (A1), definition of d^{C}, Lemmas A.2 and A.4, and $\left\|\widehat{z}^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}\right\|_\infty \leq 1$ by the construction of $\widehat{z}^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}$, and the third step is due to (C.10).

Hence, by PDW approach, we have (C.1) for every $b = 1, \cdots, B$ and $\zeta \in \mathcal{Z}$ with probability greater than $1 - \left\{c_1\{\max(n,p)\}^{-2} + \exp(-c_2 n)\right\}$.

Step 2: *Let*

$$\widehat{\theta}^{C}_{rt}(\zeta) = \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}^{C}_{rt}(\zeta, b) \quad \textit{and} \quad \widehat{\mathcal{N}}^{C}(r;\zeta) = \left\{t \in V^{C}\setminus\{r\} : \widehat{\theta}^{C}_{rt}(\zeta) \neq 0\right\}.$$

In addition, let

$$\widehat{\theta}^{CD}_{rt'}(\zeta) = \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}^{CD}_{rt'}(\zeta, b) \quad \textit{and} \quad \widehat{\mathcal{N}}^{CD}(r;\zeta) = \left\{t' \in V^{D}\setminus\{r\} : \widehat{\theta}^{CD}_{rt'}(\zeta) \neq 0\right\}.$$

For $r \in V^{C}$, show that

$$\widehat{\mathcal{N}}^{C}(r;\zeta) = \mathcal{N}^{C}(r) \quad \textit{and} \quad \widehat{\mathcal{N}}^{CD}(r;\zeta) = \mathcal{N}^{CD}(r) \qquad \text{(C.12)}$$

with probability greater than $1 - \left\{c_1\{\max(n,p)\}^{-2} + \exp(-c_2 n)\right\}$.

The derivations of two results in (C.12) are similar, so we only present the former result.

Since

$$\begin{aligned}
\widehat{\mathcal{N}}^{\mathrm{C}}(r;\zeta) &= \left\{t \in V^{\mathrm{C}}\setminus\{r\} : \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}^{\mathrm{C}}_{rt}(\zeta, b) \neq 0\right\} \\
&\subset \left\{t \in V^{\mathrm{C}}\setminus\{r\} : \text{there exists } b \in \{1, \cdots, B\}\right.
\end{aligned}$$

$$\begin{aligned} & \text{such that } \widehat{\theta}^{\mathrm{C}}_{rt}(\zeta, b) \neq 0 \Big\} \\ = & \bigcup_{b=1}^{B} \widehat{\mathcal{N}}^{\mathrm{C}}_b(r;\zeta) \\ = & \mathcal{N}^{\mathrm{C}}(r) \end{aligned}$$

with the probability greater than $1 - \Big\{c_1\{\max(n,p)\}^{-2} + \exp(-c_2 n)\Big\}$, where the last step is due to the result in Step 1 and the finiteness of B.

Next, we show that $\mathcal{N}^{\mathrm{C}}(r) \subseteq \widehat{\mathcal{N}}^{\mathrm{C}}(r;\zeta)$ with the probability greater than $1 - \Big\{c_1\{\max(n,p)\}^{-2} + \exp(-c_2 n)\Big\}$. For any $t \in V^{\mathrm{C}} \setminus \{r\}$, by the result in Step 1, for any b, $t \in \widehat{\mathcal{N}}^{\mathrm{C}}_b(r;\zeta)$, i.e., $\widehat{\theta}^{\mathrm{C}}_{rt}(\zeta, b) \neq 0$, with the probability greater than $1 - \Big\{c_1\{\max(n,p)\}^{-2} + \exp(-c_2 n)\Big\}$. By (A.16), we can show the boundness of $\Big\|\widehat{\theta}^{\mathrm{C}}_{\setminus r} - \theta^{\mathrm{C}}_{0;\setminus r}\Big\|_\infty$ and the sign recovery of $\widehat{\theta}^{\mathrm{C}}_{rt}(\zeta, b)$ with the probability greater than $1 - \Big\{c_1\{\max(n,p)\}^{-2} + \exp(-c_2 n)\Big\}$. Therefore, the sign of $\widehat{\theta}^{\mathrm{C}}_{rt}(\zeta, b)$ is the same as that of θ^{C}_{rt} for any b with the probability greater than $1 - \Big\{c_1\{\max(n,p)\}^{-2} + \exp(-c_2 n)\Big\}$, showing that $\widehat{\theta}^{\mathrm{C}}_{rt}(\zeta) \neq 0$, or $\mathcal{N}^{\mathrm{C}}(r) \subseteq \widehat{\mathcal{N}}^{\mathrm{C}}(r;\zeta)$, with the probability greater than $1 - \Big\{c_1\{\max(n,p)\}^{-2} + \exp(-c_2 n)\Big\}$.

Step 3: *Establish the desired result.*

It suffices to show that $\widehat{\mathcal{N}}^{\mathrm{C}}(r) = \bigcup_{\zeta\in\mathcal{Z}} \widehat{\mathcal{N}}^{\mathrm{C}}(r;\zeta)$ and $\widehat{\mathcal{N}}^{\mathrm{CD}}(r) = \bigcup_{\zeta\in\mathcal{Z}} \widehat{\mathcal{N}}^{\mathrm{CD}}(r;\zeta)$ with the probability greater than $1 - \Big\{c_1\{\max(n,p)\}^{-2} + \exp(-c_2 n)\Big\}$ for $r \in V^{\mathrm{C}}$, and thus, we obtain the desired result by similar derivations in Step 2. In this step, we only show $\widehat{\mathcal{N}}^{\mathrm{C}}(r) = \bigcup_{\zeta\in\mathcal{Z}} \widehat{\mathcal{N}}^{\mathrm{C}}(r;\zeta)$ for $r \in V^{\mathrm{C}}$, and similar arguments yield $\widehat{\mathcal{N}}^{\mathrm{CD}}(r) = \bigcup_{\zeta\in\mathcal{Z}} \widehat{\mathcal{N}}^{\mathrm{CD}}(r;\zeta)$.

For ease of presentation, we assume that the extrapolation function is a quadratic polynomial function with two parameters γ_1 and γ_2. Let $(\widehat{\gamma}_1, \widehat{\gamma}_2)$ denote the estimators of (γ_1, γ_2), then

$$\widehat{\theta}^{\mathrm{C}}_{rt}(\zeta) = \widehat{\gamma}_1\zeta + \widehat{\gamma}_2\zeta^2 \quad \text{for} \quad t \in \mathcal{N}^{\mathrm{C}}(r), \quad r \in V^{\mathrm{C}}, \quad \text{and} \quad \zeta \in \mathcal{Z}. \tag{C.13}$$

First, we consider that if $t \in \widehat{\mathcal{N}}^{\mathrm{C}}(r)$, then $\widehat{\theta}^{\mathrm{C}}_{rt} = \widehat{\theta}^{\mathrm{C}}_{rt}(-1) = -\widehat{\gamma}_1 + \widehat{\gamma}_2 \neq 0$, showing that $\widehat{\gamma}_1 \neq \widehat{\gamma}_2$. Thus, by (C.13), there exists $\zeta \in \mathcal{Z}$ so that $\widehat{\theta}^{\mathrm{C}}_{rt}(\zeta) \neq 0$, i.e., $t \in \bigcup_{\zeta\in\mathcal{Z}} \widehat{\mathcal{N}}^{\mathrm{C}}(r;\zeta)$. Thus, $\widehat{\mathcal{N}}^{\mathrm{C}}(r) \subseteq \bigcup_{\zeta\in\mathcal{Z}} \widehat{\mathcal{N}}^{\mathrm{C}}(r;\zeta)$.

Next, suppose that $t \in \bigcup_{\zeta\in\mathcal{Z}} \widehat{\mathcal{N}}^{\mathrm{C}}(r;\zeta)$. Then there exists $\zeta_0 \in \mathcal{Z}$ such that $t \in \widehat{\mathcal{N}}^{\mathrm{C}}(r;\zeta_0)$. By the result in Step 2, we have that $t \in \mathcal{N}^{\mathrm{C}}(r)$, i.e., $\theta^{\mathrm{C}}_{rt} \neq 0$. Then by the result in Step 2 again, we have $\widehat{\theta}^{\mathrm{C}}_{rt}(\zeta) \neq 0$ with the probability approaching one for any $\zeta \in \mathcal{Z}$. Without loss of generality, we consider the case that $\widehat{\theta}^{\mathrm{C}}_{rt}(\zeta) > 0$.

Noting that $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ in (C.13) are the estimates obtained from fitting $\Big\{\Big(\widehat{\theta}^{\mathrm{C}}_{rt}(\zeta), \zeta\Big) : \zeta \in \mathcal{Z}\Big\}$ with the least squares method, we obtain that

$$\widehat{\gamma}_1 = \frac{1}{A_1 A_3 - A_2^2} \left\{ A_3 \sum_{\zeta \in \mathcal{Z}} \zeta \widehat{\theta}^{\mathrm{C}}_{rt}(\zeta) - A_2 \sum_{\zeta \in \mathcal{Z}} \zeta^2 \widehat{\theta}^{\mathrm{C}}_{rt}(\zeta) \right\}$$

and

$$\widehat{\gamma}_2 = \frac{1}{A_1 A_3 - A_2^2} \left\{ -A_2 \sum_{\zeta \in \mathcal{Z}} \zeta \widehat{\theta}^{\mathrm{C}}_{rt}(\zeta) + A_1 \sum_{\zeta \in \mathcal{Z}} \zeta^2 \widehat{\theta}^{\mathrm{C}}_{rt}(\zeta) \right\},$$

where $A_r = \sum_{\zeta \in \mathcal{Z}} \zeta^{r+1}$ for $r = 1, 2, 3$.

Finally, taking $\zeta = -1$ in (C.13) yields

$$\widehat{\theta}^{\mathrm{C}}_{rt} = \frac{1}{A_2^2 - A_1 A_3} \sum_{i=1}^{M} (A_3 - A_2 \zeta_i + A_2 - A_1 \zeta_i)\, \zeta_i \widehat{\theta}^{\mathrm{C}}_{rt}(\zeta_i), \qquad \text{(C.14)}$$

where we use $\mathcal{Z} = \{\zeta_0, \zeta_1, \cdots, \zeta_M\}$ with $\zeta_0 = 0$. Noting that for any $\zeta_i \in \mathcal{Z}$,

$$A_3 - A_2 \zeta_i + A_2 - A_1 \zeta_i = \sum_{k>i} (\zeta_k - \zeta_i) \left(\zeta_k^3 + \zeta_k^2\right) + \sum_{k<i} (\zeta_k - \zeta_i) \left(\zeta_k^3 + \zeta_k^2\right),$$

we write (C.14) as

$$\widehat{\theta}^{\mathrm{C}}_{rt} = \frac{1}{A_2^2 - A_1 A_3} \sum_{i=1}^{M} \left[\zeta_i \widehat{\theta}^{\mathrm{C}}_{rt}(\zeta_i) \left\{ \sum_{k>i} (\zeta_k - \zeta_i) \left(\zeta_k^3 + \zeta_k^2\right) + \sum_{k<i} (\zeta_k - \zeta_i) \left(\zeta_k^3 + \zeta_k^2\right) \right\} \right].$$

Moreover, further computations give

$$\begin{aligned} \widehat{\theta}^{\mathrm{C}}_{rt} &= \frac{1}{A_2^2 - A_1 A_3} \sum_{i=1}^{M} \sum_{k=i+1}^{M} \left[(\zeta_k - \zeta_i) \left\{ \zeta_i \widehat{\theta}^{\mathrm{C}}_{rt}(\zeta_i) \left(\zeta_k^3 + \zeta_k^2\right) \right.\right. \\ &\quad \left.\left. - \zeta_k \widehat{\theta}^{\mathrm{C}}_{rt}(\zeta_k) \left(\zeta_i^3 + \zeta_i^2\right) \right\} \right] \\ &\geq \frac{1}{A_2^2 - A_1 A_3} \sum_{i=1}^{M} \sum_{k=i+1}^{M} (\zeta_k - \zeta_i)\, \widehat{\theta}^{\mathrm{C}}_{rt}(\zeta_i) \zeta_k \zeta_i \left(\zeta_k^2 - \zeta_i^2\right) \\ &> 0 \end{aligned}$$

due to that we assume $\widehat{\theta}^{\mathrm{C}}_{rt}(\zeta) > 0$. Therefore, we conclude that $\widehat{\theta}^{\mathrm{C}}_{rt} \neq 0$ with the probability approaching one. Thus, $t \in \widehat{\mathcal{N}}^{\mathrm{C}}(r)$ and $\widehat{\mathcal{N}}^{\mathrm{C}}(r) \supseteq \bigcup_{\zeta \in \mathcal{Z}} \widehat{\mathcal{N}}^{\mathrm{C}}(r; \zeta)$. □

Appendix D: Proof of Theorem 5.2

D.1. Proof of Part (b)

By Lemma A.3 and the fact that $\|\cdot\|_\infty \le \|\cdot\|_2$, we have

$$\left\|\widehat{\theta}^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}(r;\zeta,b) - \theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_r}(r)\right\|_\infty \le \frac{6\sqrt{d^{\mathrm{C}}_r}\lambda^{\mathrm{C}}_n}{\rho}$$

and

$$\left\|\widehat{\theta}_{\mathcal{S}^{\mathrm{D}}_{r'}}(r';\zeta,b) - \theta^{\mathrm{D}}_{0;\mathcal{S}^{\mathrm{D}}_{r'}}(r')\right\|_\infty \le \frac{6\sqrt{d^{\mathrm{D}}_{r'}}\lambda^{\mathrm{D}}_n}{\rho}$$

for $r \in V^{\mathrm{C}}$ and $r' \in V^{\mathrm{D}}$. By the definition (30), we have

$$\begin{aligned}\left\|\widehat{\theta}^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}(r;\zeta) - \theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_r}(r)\right\|_\infty &\le \frac{1}{B}\sum_{b=1}^{B}\left\|\widehat{\theta}^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}(r;\zeta,b) - \theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_r}(r)\right\|_\infty \\ &< \frac{6\sqrt{d^{\mathrm{C}}_r}\lambda^{\mathrm{C}}_n}{\rho}\end{aligned}$$

and

$$\begin{aligned}\left\|\widehat{\theta}^{\mathrm{D}}_{\mathcal{S}^{\mathrm{D}}_{r'}}(r';\zeta) - \theta^{\mathrm{D}}_{0;\mathcal{S}^{\mathrm{D}}_{r'}}(r')\right\|_\infty &\le \frac{1}{B}\sum_{b=1}^{B}\left\|\widehat{\theta}^{\mathrm{D}}_{\mathcal{S}^{\mathrm{D}}_{r'}}(r';\zeta,b) - \theta^{\mathrm{D}}_{0;\mathcal{S}^{\mathrm{D}}_{r'}}(r')\right\|_\infty \\ &< \frac{6\sqrt{d^{\mathrm{D}}_{r'}}\lambda^{\mathrm{D}}_n}{\rho}\end{aligned}$$

for $r \in V^{\mathrm{C}}$ and $r' \in V^{\mathrm{D}}$. Finally, let $\zeta \to -1$, we obtain

$$\left\|\widehat{\theta}^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}(r) - \theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_r}(r)\right\|_\infty \le \frac{6\sqrt{d^{\mathrm{C}}_r}\lambda^{\mathrm{C}}_n}{\rho}$$

and

$$\left\|\widehat{\theta}^{\mathrm{D}}_{\mathcal{S}^{\mathrm{D}}_{r'}}(r') - \theta^{\mathrm{D}}_{0;\mathcal{S}^{\mathrm{D}}_{r'}}(r')\right\|_\infty \le \frac{6\sqrt{d^{\mathrm{D}}_{r'}}\lambda^{\mathrm{D}}_n}{\rho}$$

for $r \in V^{\mathrm{C}}$ and $r' \in V^{\mathrm{D}}$. Hence, we complete the proof. □

D.2. Proof of Part (a)

As discussed in [20, pp.1301], to show the correctness of sign recovery, i.e., $\mathrm{sign}\left(\widehat{\theta}^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}(r)\right) = \mathrm{sign}\left(\theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_r}(r)\right)$ and $\mathrm{sign}\left(\widehat{\theta}^{\mathrm{D}}_{\mathcal{S}^{\mathrm{D}}_{r'}}(r')\right) = \mathrm{sign}\left(\theta^{\mathrm{D}}_{0;\mathcal{S}^{\mathrm{D}}_{r'}}(r')\right)$ for $r \in V^{\mathrm{C}}$ and $r' \in V^{\mathrm{D}}$, it suffices to check the boundness of $\left\|\widehat{\theta}^{\mathrm{C}}_{\mathcal{S}^{\mathrm{C}}_r}(r) - \theta^{\mathrm{C}}_{0;\mathcal{S}^{\mathrm{C}}_r}(r)\right\|_\infty$ and $\left\|\widehat{\theta}^{\mathrm{D}}_{\mathcal{S}^{\mathrm{D}}_{r'}}(r') - \theta^{\mathrm{D}}_{0;\mathcal{S}^{\mathrm{D}}_{r'}}(r')\right\|_\infty$. Since Theorem 5.2 (b) holds, we directly obtain the desired result. □

TABLE 1
Simulation results for the estimators of Θ under Scenario 1

Model	(n, p^{C})	σ_ϵ^2	Method	Estimator of Θ_0				Time
				$\|\Delta_\Theta\|_1$	$\|\Delta_\Theta\|_F$	Spe	Sen	
Lattice	(400,20)	0.15	naive	2.773	14.986	0.533	1.000	91.64
			corrected-Q	1.691	2.021	0.988	1.000	1033.50
			corrected-L	1.789	3.844	0.970	1.000	441.48
		0.50	naive	2.711	15.833	0.207	1.000	99.12
			corrected-Q	1.741	3.768	1.000	1.000	1892.40
			corrected-L	1.846	4.219	0.959	1.000	563.95
		0.75	naive	3.111	15.653	0.071	1.000	105.17
			corrected-Q	1.956	3.318	0.988	1.000	2497.65
			corrected-L	2.208	4.605	0.959	1.000	847.95
			true	1.316	1.954	1.000	1.000	90.84
	(400, 100)	0.15	naive	2.555	86.625	0.749	1.000	194.15
			corrected-Q	0.822	6.061	1.000	1.000	5327.51
			corrected-L	0.873	9.734	1.000	1.000	5022.92
		0.50	naive	3.434	90.056	0.237	1.000	287.68
			corrected-Q	1.064	7.320	0.995	1.000	6922.95
			corrected-L	1.083	7.537	0.998	1.000	5518.98
		0.75	naive	3.911	96.357	0.094	1.000	373.78
			corrected-Q	1.549	14.247	0.995	1.000	11243.00
			corrected-L	1.655	15.849	0.991	1.000	9829.60
			true	0.664	4.924	1.000	1.000	180.85
	(200, 400)	0.15	naive	4.895	106.474	0.298	1.000	1200.00
			corrected-Q	1.234	34.828	1.000	0.996	49068.53
			corrected-L	1.655	37.441	0.993	0.972	48660.14
		0.50	naive	8.243	521.504	0.082	0.903	1592.17
			corrected-Q	2.072	56.988	0.999	0.993	66674.65
			corrected-L	2.148	60.969	0.986	0.996	66236.65
		0.75	naive	11.209	621.102	0.030	1.000	2139.54
			corrected-Q	3.908	66.474	0.998	0.956	91102.46
			corrected-L	4.342	72.230	0.878	1.000	90503.39
			true	0.586	1.954	1.000	1.000	1534.17
Hub	(400,20)	0.15	naive	4.857	10.375	0.637	1.000	86.54
			corrected-Q	1.933	2.448	1.000	0.944	975.98
			corrected-L	1.566	1.680	0.972	1.000	473.42
		0.50	naive	4.678	9.716	0.357	1.000	93.57
			corrected-Q	1.354	0.799	1.000	1.000	1753.40
			corrected-L	1.822	1.962	1.000	0.944	664.59
		0.75	naive	4.574	10.173	0.110	1.000	101.40
			corrected-Q	1.209	0.674	1.000	1.000	1953.88
			corrected-L	2.298	2.314	1.000	0.908	853.10
			true	1.651	1.207	1.000	1.000	86.45
	(400, 100)	0.15	naive	9.108	43.036	0.735	1.000	185.09
			corrected-Q	2.259	3.183	1.000	1.000	5634.23
			corrected-L	2.422	3.583	0.997	0.978	5278.04
		0.50	naive	9.671	47.410	0.432	1.000	253.01
			corrected-Q	2.366	3.477	1.000	1.000	6425.84
			corrected-L	2.815	3.624	0.997	1.000	5931.45
		0.75	naive	9.676	49.876	0.163	1.000	361.21
			corrected-Q	2.659	3.146	0.998	1.000	9838.12
			corrected-L	3.309	3.761	1.000	0.936	8412.04
			true	1.677	1.364	1.000	1.000	174.51
	(200, 400)	0.15	naive	10.179	206.529	0.433	1.000	1149.53
			corrected-Q	2.735	15.989	0.999	0.995	44539.17
			corrected-L	3.659	16.856	0.988	0.982	41143.39
		0.50	naive	12.633	261.196	0.165	1.000	1566.76
			corrected-Q	3.280	27.334	0.998	0.989	46597.28
			corrected-L	3.678	34.372	0.998	0.957	45821.30
		0.75	naive	15.659	332.615	0.041	1.000	1771.28
			corrected-Q	3.581	31.680	0.997	0.955	66573.01
			corrected-L	3.704	35.842	0.996	1.000	64129.85
			true	0.206	7.686	1.000	1.000	1101.59

TABLE 2
Simulation results for the estimators of Θ under Scenario 2

Model	(n, p^{D})	π	Method	Estimator of Θ_0				Time
				$\|\Delta_\Theta\|_1$	$\|\Delta_\Theta\|_F$	Spe	Sen	
Lattice	(400,20)	0.70	naive	2.856	27.741	0.876	0.871	193.70
			corrected-Q	1.586	6.607	0.988	0.974	6569.73
			corrected-L	1.665	6.021	0.984	0.967	6011.44
		0.80	naive	2.571	23.328	0.840	0.868	169.62
			corrected-Q	1.606	5.473	1.000	0.974	6016.77
			corrected-L	1.635	5.786	1.000	0.946	5640.03
		0.90	naive	2.447	21.853	0.858	0.889	168.50
			corrected-Q	1.564	5.129	0.994	0.972	4503.77
			corrected-L	1.576	5.258	0.986	1.000	4098.05
			true	0.879	3.335	1.000	0.996	157.84
	(400, 15)	0.70	naive	2.843	19.047	0.779	0.681	150.85
			corrected-Q	2.338	12.111	1.000	0.945	4358.09
			corrected-L	2.359	12.772	1.000	0.912	3965.89
		0.80	naive	2.670	17.959	0.823	0.727	147.23
			corrected-Q	1.981	10.231	1.000	0954	3847.65
			corrected-L	2.106	11.670	1.000	0.945	3352.87
		0.90	naive	2.469	15.573	0.856	0.743	142.59
			corrected-Q	1.590	7.038	1.000	0.989	3573.40
			corrected-L	1.595	7.514	0.989	1.000	3145.58
			true	0.885	2.105	1.000	0.964	143.25
	(15, 20)	0.70	naive	7.769	64.859	0.083	1.000	400.98
			corrected-Q	2.034	19.371	0.922	0.952	13448.10
			corrected-L	2.170	21.330	0.922	0.950	10980.45
		0.80	naive	5.313	55.756	0.143	1.000	377.21
			corrected-Q	1.130	25.334	0.944	0.903	10339.48
			corrected-L	1.345	28.740	0.940	0.900	9453.09
		0.90	naive	5.017	54.200	0.159	1.000	375.65
			corrected-Q	1.130	21.670	0.968	0.968	8352.76
			corrected-L	1.168	24.114	0.950	0.968	6658.41
			true	1.760	4.999	0.986	0.973	357.60
Hub	(400,20)	0.70	naive	6.781	15.941	0.676	0.844	199.32
			corrected-Q	3.648	4.281	0.961	0.945	6342.88
			corrected-L	3.970	4.641	0.950	0.928	5877.10
		0.80	naive	7.013	15.918	0.692	0.833	193.14
			corrected-Q	2.971	4.205	0.987	1.000	5947.03
			corrected-L	3.458	4.334	0.976	1.000	5360.14
		0.90	naive	5.059	10.779	0.720	0.904	191.62
			corrected-Q	1.615	1.074	1.000	1.000	5530.17
			corrected-L	2.178	1.555	1.000	1.000	4972.26
			true	1.510	0.977	1.000	1.000	122.51
	(400, 15)	0.70	naive	4.511	11.085	0.628	0.846	196.60
			corrected-Q	2.100	4.800	0.960	0.996	6071.42
			corrected-L	2.763	5.085	0.960	1.000	5433.79
		0.80	naive	4.572	11.818	0.668	0.615	191.03
			corrected-Q	1.903	3.955	0.989	1.000	5488.92
			corrected-L	2.411	4.028	0.980	1.000	4930.77
		0.90	naive	3.879	8.063	0.658	1.000	189.50
			corrected-Q	1.611	1.833	1.000	1.000	5006.33
			corrected-L	2.235	2.006	0.990	1.000	4470.89
			true	0.996	0.661	1.000	1.000	125.35
	(15, 20)	0.70	naive	13.580	71.308	0.050	1.000	411.35
			corrected-Q	7.356	36.161	0.918	0.944	12590.44
			corrected-L	7.620	39.233	0.915	0.940	9733.56
		0.80	naive	10.354	43.816	0.054	1.000	387.85
			corrected-Q	6.553	26.891	0.938	0.933	9653.78
			corrected-L	6.855	28.115	0.930	0.926	8860.21
		0.90	naive	9.406	36.304	0.055	1.000	385.04
			corrected-Q	4.973	18.947	0.929	0.954	9044.13
			corrected-L	5.235	20.276	0.920	0.947	8033.49
			true	2.931	4.650	0.996	0.961	362.98

TABLE 3
Simulation results for the estimators of Θ under Scenario 3

Model	(n,p)	(σ_ϵ^2,π)	Method	Estimator of Θ_0				Time
				$\|\Delta_\Theta\|_1$	$\|\Delta_\Theta\|_F$	Spe	Sen	
Lattice	(400,20)	(0.15,0.9)	naive	1.878	10.671	0.811	0.839	194.62
			corrected-Q	1.111	3.787	0.982	0.939	6634.58
			corrected-L	1.305	4.096	0.976	0.916	5873.66
		(0.50,0.8)	naive	1.921	11.277	0.746	0.806	197.85
			corrected-Q	1.630	5.264	0.952	0.942	7465.11
			corrected-L	1.718	5.569	0.953	0.906	6450.14
		(0.75,0.7)	naive	2.305	11.869	0.686	0.839	207.18
			corrected-Q	1.692	6.218	0.953	0.967	8892.17
			corrected-L	1.705	6.286	0.933	0.905	7253.09
			true	0.853	2.210	1.000	0.977	164.68
	(400,100)	(0.15,0.9)	naive	1.893	53.422	0.718	0.744	200.67
			corrected-Q	1.685	18.655	0.998	0.928	7459.16
			corrected-L	1.753	20.217	0.993	0.933	6534.33
		(0.50,0.8)	naive	2.110	64.900	0.646	0.744	207.73
			corrected-Q	1.800	18.708	0.993	0.928	7853.72
			corrected-L	1.902	22.020	0.995	0.917	7079.10
		(0.75,0.7)	naive	2.266	58.259	0.581	0.739	210.09
			corrected-Q	1.812	23.935	0.987	0.933	8511.38
			corrected-L	2.150	26.992	0.950	0.933	7725.46
			true	1.200	9.068	1.000	0.965	154.01
	(200,300)	(0.15,0.9)	naive	2.538	131.564	0.581	0.866	797.64
			corrected-Q	1.914	50.116	0.994	0.958	15948.73
			corrected-L	2.466	58.713	0.960	0.950	11456.58
		(0.50,0.8)	naive	5.223	174.617	0.278	0.872	830.01
			corrected-Q	3.767	110.748	0.967	0.958	18943.55
			corrected-L	3.956	112.699	0.952	0.958	13149.10
		(0.75,0.7)	naive	5.991	169.521	0.260	0.879	881.60
			corrected-Q	4.136	118.195	0.959	0.956	23578.61
			corrected-L	4.574	122.356	0.950	0.947	19744.23
			true	0.971	16.218	0.999	0.958	740.78
Hub	(400,20)	(0.15,0.9)	naive	3.397	5.098	0.599	1.000	190.93
			corrected-Q	1.142	1.200	1.000	1.000	6743.89
			corrected-L	1.194	1.777	1.000	1.000	5376.11
		(0.50,0.8)	naive	5.190	8.414	0.500	0.667	197.11
			corrected-Q	2.301	3.918	0.981	1.000	7568.44
			corrected-L	2.357	4.042	0.972	1.000	6340.29
		(0.75,0.7)	naive	5.351	11.984	0.439	1.000	203.92
			corrected-Q	2.266	4.578	0.979	1.000	8413.69
			corrected-L	3.359	4.896	0.961	0.989	7148.33
			true	0.468	0.755	1.000	0.984	12.28
	(400, 100)	(0.15,0.9)	naive	9.670	27.398	0.769	0.705	211.46
			corrected-Q	5.447	14.104	0.992	0.947	9453.22
			corrected-L	6.172	16.136	0.985	0.932	7944.25
		(0.50,0.8)	naive	10.235	35.899	0.628	0.716	215.00
			corrected-Q	7.274	24.237	0.958	0.947	9844.06
			corrected-L	7.445	26.099	0.953	0.942	8563.74
		(0.75,0.7)	naive	12.733	60.029	0.673	0.715	213.98
			corrected-Q	7.811	29.614	0.961	0.953	10689.40
			corrected-L	7.906	32.677	0.938	0.942	9437.51
			true	3.337	11.003	0.998	0.973	158.45
	(200,300)	(0.15,0.9)	naive	9.401	83.764	0.401	1.000	906.42
			corrected-Q	4.585	32.320	0.987	1.000	16530.48
			corrected-L	4.740	33.265	0.974	1.000	12447.50
		(0.50,0.8)	naive	13.650	132.395	0.206	0.875	938.32
			corrected-Q	5.254	117.863	0.957	0.977	20679.23
			corrected-L	5.635	120.466	0.946	0.960	16544.83
		(0.75,0.7)	naive	11.708	159.760	0.176	1.000	989.26
			corrected-Q	6.257	120.503	0.951	0.996	29450.77
			corrected-L	6.433	121.659	0.945	0.981	20335.40
			true	1.101	1.694	0.999	1.000	904.81

References

[1] Bandara, S., Schlöder, J. P., Eils, R., Bock, H. G., and Meyer, T. (2009). Optimal experimental design for parameter estimation of a cell signaling model. *PLoS Computational Biology*, 5, e1000558. doi:10.1371/journal.pcbi.1000558

[2] Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., and Sudman, S. (1991). *Measurement Error in Surveys*. John Wiley & Sons, Inc., Hoboken, New Jersey. MR3236959

[3] Buonaccorsi, J. P. (2010). *Measurement Error: Models, Methods, and Applications*. Chapman & Hall/CRC, New York. MR2682774

[4] Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Model*. Chapman and Hall, New York. MR1630517

[5] Chen, S., Witten, D. M., and Shojaie, A. (2015). Selection and estimation for mixed graphical models. *Biometrika*, 102, 47–64. MR3335095

[6] Chen, L.-P. and Yi, G. Y. (2020). Model selection and model averaging for analysis of truncated and censored data with measurement error. *Electronic Journal of Statistics*, 14, 4054–4109. MR4170184

[7] Chen, L.-P. and Yi, G. Y. (2021a). Analysis of noisy survival data with graphical proportional hazards measurement error models. *Biometrics*, 77, 956–969. MR4320670

[8] Chen, L.-P. and Yi, G. Y. (2021b). Semiparametric methods for left-truncated and right-censored survival data with covariate measurement error. *Annals of the Institute of Statistical Mathematics*, 73, 481–517. MR4247068

[9] Cheng, J., Li, T., Levina, E., and Zhu, J. (2017). High-dimensional mixed graphical models. *Journal of Computational and Graphical Statistics*, 26, 367–378. MR3640193

[10] Cook, J. R. and Stefaski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, 89, 1314–1328. MR1379467

[11] Dalal, O. and Rajaratnam, B. (2017). Sparse Gaussian graphical model estimation via alternating minimization. *Biometrika*, 104, 379–395 MR3698260

[12] Fan, J., Liu, H., Ning, Y., and Zou, H. (2017). High dimensional semiparametric latent graphical model for mixed data. *Journal of the Royal Statistical Society, Series B*, 79, 405–421. MR3611752

[13] Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso *Biostatistics*, 9, 432–441.

[14] Fuller, W. A. (1987). *Measurement Error Models*. Wiley, New York. MR0898653

[15] Gustafson, P. (2004). *Measurement Error and Misclassification in Statistics and Epidemiology*. Chapman & Hall/CRC, New York. MR2005104

[16] Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC press, New York. MR3616141

[17] Küchenhoff, H., Mwalili, S. M., and Leasaffre, E. (2006). A general method for dealing with misclassificationin regression: The misclassification SIMEX. *Biometrics*, 62, 85–96. MR2226560

[18] Lee, J. and Hastie, T. J. (2015). Learning the structure of mixed graphical models. *Journal of Computational and Graphical Statistics*, 24, 230–253. MR3328255

[19] Meinshausen, N and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *Annals of Statistics*, 34, 1436–1462. MR2278363

[20] Ravikumar, P., Wainwright, M. J., and Lafferty, J. (2010). High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *The Annals of Statistics*, 38, 1287–1319. MR2662343

[21] Ravikumar, P., Wainwright, M. J., Raskutti, G., and Yu, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electronic Journal of Statistics*, 5, 935–980 MR2836766

[22] Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D., and Nolan, G. (2005). Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308, 523–529.

[23] Sun, H. and Li, H. (2012). Robust Gaussian graphical modeling via ℓ_1 penalization. *Biometrics*, 68, 1197–1206. MR3040026

[24] Tan, K. M., Ning, Y., Witten, D. M., and Liu, H. (2016). Replicates in high dimensions, with applications to latent variable graphical models. *Biometrika*, 103, 761–777. MR3620438

[25] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58, 267–288. MR1379242

[26] Wainwright, M. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press, New York. MR3967104

[27] Wang, H., Li, R., and Tsai, C. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*, 94, 553–568. MR2410008

[28] Yang, E., Ravikumar, P., Allen, G. I., and Liu, Z. (2015). Graphical models via univariate exponential family distribution. *Journal of Machine Learning Research*, 16, 3813–3847. MR3450553

[29] Yuan, M. and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94, 19–35 MR2367824

[30] Yi, G. Y. (2017). *Statistical Analysis with Measurement Error and Misclassication: Strategy, Method and Application*. Springer, New York. MR3676914

[31] Yi, G. Y., Delaigle, A., and Gustafson, P. (2021). *Handbook of Measurement Error Models*. Chapman & Hall/CRC, Boca Raton, FL.

[32] Yi, G. Y. and He, W. (2017). Analysis of case-control data with interacting misclassified covariates. *Journal of Statistical Distributions and Application*, 4:16. DOI 10.1186/s40488-017-0069-0 MR3924881

[33] Yi, G. Y., Ma, Y., Spiegelman, D., and Carroll, R. J. (2015). Functional and structural methods With mixed measurement error and misclassification in

covariates. *Journal of the American Statistical Association*, 110, 681–696. MR3367257

[34] Yi, G. Y., Tan, X., and Li, R. (2015). Variable selection and inference procedures for marginal analysis of longitudinal data with missing observations and covariate measurement error. *Canadian Journal of Statistics*, 43, 498–518. MR3433673

[35] Yi, G. Y., Yan, Y., Liao, X., and Spiegelman, D. (2018). Parametric regression analysis with covariate misclassification in main study/validation study designs. *The International Journal of Biostatistics*. DOI: 10.1515/ijb-2017-0002 MR3962651

[36] Yörük, E., Ochs, M. F., Geman, D., and Younes, L. (2011). A comprehensive statistical model for cell signaling and protein activity inference. *IEEE/ACM Trans Comput Biol Bioinform*, 8, 592–606.

[37] Zhou, S., van de Geer, S., and Bühlmann, P. (2009). Adaptive lasso for high-dimensional regression and Gaussian graphical modeling. arXiv:0903.2515 MR2572443