Biometrics 000, 1–20 DOI: 000
000 0000

Statistical Inference on High Dimensional Covariate-Dependent Gaussian Graphical Regressions

Xuran Meng
Department of Biostatistics, University of Michigan, Ann Arbor, USA
email: xuranm@umich.edu

and

Jingfei Zhang
Department of Information Systems and Operations Management, Emory University, Atlanta, USA
email: emma.zhang@emory.edu

and

Yi Li
Department of Biostatistics, University of Michigan, Ann Arbor, USA
email: yili@umich.edu

Summary: In many genomic studies, gene co-expression graphs are influenced by subject-level covariates like SNPs. Traditional Gaussian graphical models ignore these covariates and estimate only population-level networks, potentially masking important heterogeneity. Covariate-dependent Gaussian graphical regressions address this limitation by regressing the precision matrix on covariates, thereby modeling how graph structures vary with high-dimensional subject-specific covariates. To fit the model, we adopt a multi-task learning approach that achieves lower error rates than node-wise regressions. Yet, the important problem of statistical inference in this setting remains largely unexplored. We propose a class of debiased estimators based on multi-task learners, which can be computed quickly and separately. In a key step, we introduce a novel projection technique for estimating the inverse covariance matrix, reducing optimization costs to scale with the sample size n. Our debiased estimators achieve fast convergence and asymptotic normality, enabling valid inference. Simulations demonstrate the utility of the method, and an application to a brain cancer gene-expression dataset reveals meaningful biological relationships.

Key words: Debiased inference; Gaussian graphical models; Graphical model with covariates; Multi-task learning; Projection.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

Gaussian graphical models are powerful tools for describing dependencies among response variables in a homogeneous population (Peng et al., 2009). In precision medicine where subject-specific gene co-expression graphs are of interest, precision matrices may depend on external covariates (Saegusa and Shojaie, 2016). For instance, genetic variants and environmental factors affect both genes and their co-expression relationships in gene co-expression graphs (Wang et al., 2012); external factors like gender, age and genetic variants influence functional connectivity between brain regions (Zhang et al., 2023). Our motivating dataset (GSE108476) comprises 178 glioblastoma multiforme patients with both microarray and SNP chip profiling. We aim to analyze 73 KEGG glioma-pathway genes with 120 covariates (118 nearby SNPs plus age and gender) to investigate how these covariates shape gene co-expression patterns. Towards this goal, Zhang and Li (2023) proposed high-dimensional Gaussian graphical regression, modeling the precision matrix as a covariate function with sparse group lasso and node-wise regressions. Zhang and Li (2025) further introduced a joint multi-task learning approach with an efficient augmented Lagrangian algorithm. While these works focus on estimation, inference in high-dimensional settings remains underexplored.

In high-dimensional problems, debiasing is a key tool for inference, as lasso-type estimators suffer from bias in second-order expansions, yielding inaccurate results (Bellec and Zhang, 2022; Cai et al., 2022). Debiasing within the lasso framework has enabled confidence intervals and p-values in high-dimensional linear regressions (Javanmard and Montanari, 2014; Zhang and Zhang, 2014), while Neyman-orthogonal moments with cross-fitting address nuisance parameters (Chernozhukov et al., 2018). Extensions include debiasing without sparsity (Zhu and Bradic, 2018), sample-splitting (Fei et al., 2019; Fei and Li, 2021), and unbiased tensor regression algorithms (Xia et al., 2022). Li et al. (2023) proposed a direct debiasing approach

without extra score calculation, but their i.i.d. auxiliary dataset framework does not apply to covariate-dependent settings where no common covariance matrix exists.

Estimation and inference in high-dimensional Gaussian graphical regressions face challenges from model complexity, dimensionality, and computational burden. A multi-task estimator solving all $O(p^2q)$ coefficients (p Gaussian variables, q covariates) achieves fast convergence, and augmented Lagrangian algorithms enable efficient estimation (Zhang and Li, 2025). However, debiasing all $O(p^2q)$ coefficients remains demanding. Theoretical challenges arise because the design matrix, formed from interactions between p dependent Gaussian variables and q covariates, has a complex joint distribution that is neither sub-Gaussian nor sub-exponential (Cai et al., 2022). Furthermore, unlike standard multi-task problems with independent regressions, the p tasks here are dependent through external covariates, rendering joint inference infeasible.

We tackle these challenges in what is, to our knowledge, the first study to enable inference for network regressions with high-dimensional responses and predictors. Our key insight is that although coefficients are estimated jointly, they can be debiased separately; we thus propose a marginalized debiasing procedure that decomposes inference node by node while maintaining validity. Unlike traditional methods (Javanmard and Montanari, 2014; Cai et al., 2022) with polynomial costs in p and q, we develop a projection technique that maps the optimization from $\mathbb{R}^{(p-1)(q+1)}$ to $\mathbb{R}^n$, solves it efficiently, and maps back, reducing computational cost. This enables scalable inference in high-dimensional Gaussian graphical regression with combined penalties. We establish asymptotic theory and relax sparsity requirements from $O(n^{1/6})$ to $o(n^{1/2})$ compared to Zhang and Li (2023, 2025).

Related to our work, Hudson and Shojaie (2022) explored a debiasing method for covariate-adjusted testing in differential graph analysis, but their approach is limited to a small number of variables; Cai et al. (2022) considered debiased sparse group lasso, but their method is

not directly applicable in our settings with high-dimensional interaction terms and incurs prohibitive computational costs that scale polynomially with p and q.

In the following Section 2, we describe the Gaussian graphical regression model and review its multi-task estimators. Section 3 proposes a new inference method for Gaussian graphical regression models. Section 4 conducts simulations and Section 5 applies our inference approach to analyze a brain tumor data set. Section 6 concludes the paper.

2. The Preamble

Lowercase letters denote scalars, boldface vectors and matrices. We write $\mathbb{N}$ and $\mathbb{R}$ for natural and real numbers, $[n_1 : n_2] = n_1, \ldots, n_2$, and $[n] = 1, \ldots, n$. The Hadamard product is "$\odot$." For sequences, $X_1(n) = O(X_2(n))$ (or $X_1(n) \precsim X_2(n)$) if $|X_1(n)/X_2(n)|$ is bounded in probability, and $X_1(n) = o(X_2(n))$ (or $X_2(n) = w(X_1(n))$) if $X_1(n)/X_2(n) \to 0$ in probability. We denote the ℓ_q norm by $\|\cdot\|_q$.

2.1 *Gaussian graphical regression*

We consider covariate-dependent Gaussian graphical models (Zhang and Li, 2023, 2025) , where $\boldsymbol{U} = (U_1, \ldots, U_q)^\top$ represents a q-dimensional vector of covariates and the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{U} = \mathbf{u}$ is modeled as:

$$\boldsymbol{X} \mid \boldsymbol{U} = \mathbf{u} \sim \mathcal{N}_p(\boldsymbol{\mu}(\mathbf{u}), \boldsymbol{\Omega}^{-1}(\mathbf{u})). \tag{1}$$

The mean vector $\boldsymbol{\mu}(\mathbf{u})$ and precision matrix $\boldsymbol{\Omega}(\mathbf{u})$ are given by

$$\boldsymbol{\mu}(\mathbf{u}) = \boldsymbol{\Gamma}\mathbf{u}, \quad \boldsymbol{\Omega}(\mathbf{u}) = \mathbf{B}_0 + \sum_{h=1}^{q} \mathbf{B}_h u_h, \tag{2}$$

where $\boldsymbol{\Gamma} \in \mathbb{R}^{p\times q}$ is the coefficient matrix for the mean vector, and $\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_q \in \mathbb{R}^{p\times p}$ are symmetric matrices. This framework allows covariate effects to enter through each $\mathbf{B}_h$ in the precision matrix, with diagonal elements fixed as $\Omega(\mathbf{u})_{jj} = \sigma^{jj}$; so residual variances remain covariate-invariant for tractability. The formulation of (1) and (2) is useful as, after

centering the vector $\boldsymbol{Z} = \boldsymbol{X} - \boldsymbol{\Gamma}\mathbf{u} = (Z_1, \ldots, Z_p)^\top$, they can be rewritten as:

$$Z_j = \sum_{k\neq j}^{p} \beta_{jk0} Z_k + \sum_{k\neq j}^{p}\sum_{h=1}^{q} \underbrace{\beta_{jkh} \cdot u_h \cdot Z_k}_{\text{Interaction term}} + \varepsilon_j, \tag{3}$$

where $\beta_{jkh} = -(\mathbf{B}_h)_{jk}/\sigma^{jj}$, ϵ_j is independent of $\boldsymbol{Z}_{-j}$ and $\mathrm{Var}(\epsilon_j) = 1/\sigma^{jj}$, for all j, k and h.

2.2 *Multi-task learning for Gaussian graphical regressions*

Consider n independent observations, $(\mathbf{u}^{(i)}, \mathbf{x}^{(i)}) \in \mathbb{R}^q \times \mathbb{R}^p$, where $i \in [n]$. Let $\mathbf{z}^{(i)} = \mathbf{x}^{(i)} - \boldsymbol{\Gamma}\mathbf{u}^{(i)}$. To expose the key ideas, we assume $\boldsymbol{\Gamma}$ is known in the ensuing development and focus on the estimation and inference of β_{jhk}'s. We discuss the situation of unknown $\boldsymbol{\Gamma}$ in S3.3 of the Supplementary Material. Letting $\mathbf{z}_j = (z_j^{(1)}, ..., z_j^{(n)})^\top$ for $j \in [p]$ and $\mathbf{u}_h = (u_h^{(1)}, ..., u_h^{(n)})^\top$ for $h \in [q]$, the Gaussian graphical regression model on the jth response variable can be written as

$$\mathbf{z}_j = \sum_{k\neq j}^{p} \beta_{jk0}\mathbf{z}_k + \sum_{k\neq j}^{p}\sum_{h=1}^{q} \beta_{jkh}\mathbf{u}_h \odot \mathbf{z}_k + \boldsymbol{\varepsilon}_j.$$

Here, $\boldsymbol{\varepsilon}_j \in \mathbb{R}^n$ is the regression error vectors of the node j with $\boldsymbol{\varepsilon}_j \sim \mathcal{N}(0, \frac{1}{\sigma^{jj}}\mathbf{I})$. We write $\boldsymbol{\beta}_j = ((\boldsymbol{\beta}_j)_{(0)}^\top, ..., (\boldsymbol{\beta}_j)_{(q)}^\top)^\top \in \mathbb{R}^{(p-1)(q+1)}$, where $(\boldsymbol{\beta}_j)_{(h)} = (\beta_{j1h}, \ldots, \beta_{jph})^\top \in \mathbb{R}^{p-1}$ groups all the coefficients in the h-th group. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_p^\top)^\top \in \mathbb{R}^{p(p-1)(q+1)}$, and collect all group vectors $(\boldsymbol{\beta}_j)_{(h)}$ defined above into the vector $\mathbf{b}_h$, where $\mathbf{b}_h = ((\boldsymbol{\beta}_1)_{(h)}^\top, \ldots, (\boldsymbol{\beta}_p)_{(h)}^\top)^\top \in \mathbb{R}^{p(p-1)}$. We then introduce the norm $\|\cdot\|_{q_1,q_2}$ based on the group structure of a vector. Let $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{(1)}^\top, \ldots, \boldsymbol{\gamma}_{(q)}^\top)^\top$ be a vector associated with a pre-defined group structure with q groups, where $\boldsymbol{\gamma}_{(j)}$ is the sub-vector corresponding to group j for $j \in [q]$. Then, the ℓ_{q_1,q_2}-norm of $\boldsymbol{\gamma}$ is defined as: $\|\boldsymbol{\gamma}\|_{q_1,q_2} = \left(\sum_j \|\boldsymbol{\gamma}_{(j)}\|_{q_2}^{q_1}\right)^{1/q_1}$, for $0 \leqslant q_1, q_2 \leqslant +\infty$. For example, $\|\boldsymbol{\beta}_j\|_{\infty,2} = \max_h \|(\boldsymbol{\beta}_j)_{(h)}\|_2$. To ease notation, we define a large design matrix $\mathcal{W} \in \mathbb{R}^{np\times p(p-1)(q+1)}$ as:

$$\mathcal{W} = \begin{pmatrix} \mathbf{W}_1 & \cdots & \mathbf{0}_{n\times(p-1)(q+1)} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{n\times(p-1)(q+1)} & \cdots & \mathbf{W}_p \end{pmatrix},$$

where $\mathbf{W}_j = [\mathbf{z}_1, ..., \mathbf{z}_{j-1}, \mathbf{z}_{j+1}, ..., \mathbf{z}_p, \mathbf{z}_1 \odot \mathbf{u}_1, ..., \mathbf{z}_{j-1} \odot \mathbf{u}_1, \mathbf{z}_{j+1} \odot \mathbf{u}_1, ..., \mathbf{z}_p \odot \mathbf{u}_q] \in \mathbb{R}^{n \times (p-1)(q+1)}$, and the response vector $\mathbf{y}$ as $\mathbf{y} = (z_1^{(1)}, ..., z_1^{(n)}, z_2^{(1)}, ..., z_p^{(n)})^\top \in \mathbb{R}^{np}$. The **multi-task learning** simultaneously estimates all $\boldsymbol{\beta}_j$ by minimizing the following loss function:

$$\widehat{\boldsymbol{\beta}} = \text{argmin}_{\boldsymbol{\beta}} \frac{1}{2n} \left\| \mathbf{y} - \mathcal{W}\boldsymbol{\beta} \right\|_2^2 + \lambda_e \|\boldsymbol{\beta}\|_1 + \lambda_g \sum_{h=1}^{q} \|\boldsymbol{b}_h\|_2, \tag{4}$$

where $\lambda_e, \lambda_g > 0$ are two tuning parameters. Here, tasks are defined by nodes in the graphical model, with each node j representing one of p regression problems. (4) jointly estimates these using a group Lasso penalty that couples coefficients $\boldsymbol{b}_h$ across tasks, enforcing shared sparsity within a multi-task learning framework. To ensure symmetry of the estimated $\boldsymbol{\Omega}(\mathbf{u})$ in finite samples, we consider the post-processing step in Zhang and Li (2025).

The regularization in (4) is the *sparse group lasso penalty* (Cai et al., 2022; Zhang and Li, 2023), where each group vector $\boldsymbol{b}_h$ pools coefficients $(\boldsymbol{\beta}_1)(h), \ldots, (\boldsymbol{\beta}_p)(h)$ across p tasks. This multi-task framework leverages shared sparsity to select and estimate covariates, with Zhang and Li (2025) showing error rates can improve by a factor of p. The problem can be solved efficiently via Fenchel convexification (Mifflin, 1977).

3. Segmentally Debiased Multi-task Graphical Regression via Projection

Because $\widehat{\boldsymbol{\beta}}$ is biased by shrinkage, debiasing is essential for valid inference. Although (4) converges quickly, joint debiasing is infeasible since the p regression tasks are dependent with intractable error covariances. Our key insight is that, while coefficients are estimated jointly, they can be debiased separately. We propose segmental debiasing, which reduces complexity and ensures valid inference within each node, akin to generalized estimation equation (GEE) approaches.

Specifically, we segment the debiasing procedure of $\widehat{\boldsymbol{\beta}}$ to each component $\widehat{\boldsymbol{\beta}}_j$ individually, i.e., for each $j \in [p]$, we debias $\widehat{\boldsymbol{\beta}}_j$ separately and denote the estimator by $\widehat{\boldsymbol{\beta}}_j^u$:

$$\widehat{\boldsymbol{\beta}}_j^u = \widehat{\boldsymbol{\beta}}_j + \frac{1}{n} \widehat{\mathbf{M}}_j^\top \mathbf{W}_j^\top (\mathbf{z}_j - \mathbf{W}_j \widehat{\boldsymbol{\beta}}_j). \tag{5}$$

We refer to (5) as the *segmentally adjusted graphical regression* (SAGE) estimator. We later show the SAGE estimator $\widehat{\boldsymbol{\beta}}_j^u$ asymptotically follows a multivariate normal distribution, laying a foundation for inference. In (5), $\widehat{\mathbf{M}}_j = [\widehat{\mathbf{m}}_{j1}, \ldots, \widehat{\mathbf{m}}_{j(p-1)(q+1)}]$ is an estimator of $\boldsymbol{\Sigma}_{\mathbf{W}_j}^{-1}$, where $\boldsymbol{\Sigma}_{\mathbf{W}_j} = \mathbb{E}\ \mathbf{W}_j^\top \mathbf{W}_j/n$. Let $\widehat{\boldsymbol{\Sigma}}_{\mathbf{W}_j} = \mathbf{W}_j^\top \mathbf{W}_j/n$. In existing methods (Javanmard and Montanari, 2014; Cai et al., 2022), $\widehat{\mathbf{m}}_{jl}$'s are typically estimated by solving:

$$\begin{aligned} &\text{argmin}_{\mathbf{m}\in\mathbb{R}^{(p-1)(q+1)}}\ \mathbf{m}^\top \widehat{\boldsymbol{\Sigma}}_{\mathbf{W}_j} \mathbf{m} \\ &\text{subject to} \quad \|H_\alpha(\widehat{\boldsymbol{\Sigma}}_{\mathbf{W}_j}\mathbf{m} - \mathbf{e}_l)\|_{\infty,2} \leqslant \gamma, \end{aligned} \tag{6}$$

where the soft-thresholding operator $H_\alpha(x) = \text{sign}(x)\cdot(|x|-\alpha)_+$ applies pointwise to vectors, $\mathbf{e}_l \in \mathbb{R}^{(p-1)(q+1)}$ is the standard basis vector, and the thresholding scalars α and γ will be defined through theoretical analysis; see Theorem 2.

In (6), the computing cost for estimating each $\widehat{\mathbf{m}}_{jl}$ in the space of $\mathbb{R}^{(p-1)(q+1)}$ increases polynomially with p and q. Alternatively, since $\widehat{\boldsymbol{\Sigma}}_{\mathbf{W}_j}$ has rank n, this motivates us to map the optimization problem from $\mathbb{R}^{(p-1)(q+1)}$ to $\mathbb{R}^n$, making computation feasible. Specifically, consider the $n \times n$ matrix $\boldsymbol{\Xi}_j = \mathbf{W}_j\mathbf{W}_j^\top/n$, with its eigen-decomposition given by $\boldsymbol{\Xi}_j = \mathbf{U}_j\mathbf{D}_j\mathbf{U}_j^\top$, where $\mathbf{U}_j^\top\mathbf{U}_j = \mathbf{U}_j\mathbf{U}_j^\top = \mathbf{I}_n$ (Golub and Reinsch, 1971). The diagonal matrix $\mathbf{D}_j \in \mathbb{R}^{n\times n}$ collects the eigenvalues of $\boldsymbol{\Xi}_j$, which are also identical to the non-zero eigenvalues of $\widehat{\boldsymbol{\Sigma}}_{\mathbf{W}_j}$. Now, define $\mathbf{V}_j = \mathbf{W}_j^\top\mathbf{U}_j\mathbf{D}_j^{-1/2}/\sqrt{n}$. It is then easy to derive that $\mathbf{W}_j/\sqrt{n} = \mathbf{U}_j\mathbf{D}_j^{1/2}\mathbf{V}_j^\top$ and $\mathbf{V}_j^\top\mathbf{V}_j = \mathbf{I}_n$. Correspondingly, we have $\widehat{\boldsymbol{\Sigma}}_{\mathbf{W}_j} = \mathbf{W}_j^\top\mathbf{W}_j/n = \mathbf{V}_j\mathbf{D}_j\mathbf{V}_j^\top$. This allows us to consider the following optimization problem in $\mathbb{R}^n$:

$$\begin{aligned} &\text{argmin}_{\boldsymbol{\theta}\in\mathbb{R}^n}\ \boldsymbol{\theta}^\top\mathbf{D}_j\boldsymbol{\theta}, \\ &\text{subject to} \quad \|H_\alpha(\mathbf{V}_j\mathbf{D}_j\boldsymbol{\theta} - \mathbf{e}_l)\|_{\infty,2} \leqslant \gamma, \end{aligned} \tag{7}$$

where α and γ are defined later in our main results. Denote by $\widehat{\boldsymbol{\theta}}_{jl}$, which solves (7). As the columns of the orthonormal matrix $\mathbf{V}_j$ are eigenvectors of $\widehat{\boldsymbol{\Sigma}}_{\mathbf{W}_j}$ spanning an n-dimensional subspace, the $\boldsymbol{\theta}$ in (7) is a projection of $\mathbf{m}$ in (6) onto the columns of $\mathbf{V}_j$. The following proposition shows the properties of $\widehat{\boldsymbol{\theta}}_{jl}$ and $\widehat{\mathbf{m}}_{jl}$, justifying the utility of this projection, and its proof is in Section S1 of the Supplementary Material.

PROPOSITION 1: If $\widehat{\boldsymbol{\theta}}_{jl}$ is a solution of (7), then $\widehat{\mathbf{m}}_{jl} = \mathbf{V}_j\widehat{\boldsymbol{\theta}}_{jl}$ is a solution of (6). Inversely, if $\widehat{\mathbf{m}}_{jl}$ is the solution of (6), then $\widehat{\boldsymbol{\theta}}_{jl} = \mathbf{V}_j^\top\widehat{\mathbf{m}}_{jl}$ is a solution of (7).

The advantage of (7) lies in its estimation in $\mathbb{R}^n$, a space whose dimension does not depend on p or q and is typically much lower than $(p-1)(q+1)$, avoiding estimating the inverse matrix directly in the original $\mathbb{R}^{(p-1)(q+1)}$ and saving much computation. Once $\widehat{\boldsymbol{\theta}}_{jl}$ is computed, we can map it back to $\mathbb{R}^{(p-1)(q+1)}$ to obtain the estimate of each column of the inverse matrix as $\widehat{\mathbf{m}}_{jl} = \mathbf{V}_j\widehat{\boldsymbol{\theta}}_{jl}$, with $j \in [p-1)(q+1)]$. Moreover, the diagonal structure of $\mathbf{D}_j$ simplifies the computation in (7) by avoiding full matrix operations. Thus, it is clear that our debiased method remains computationally efficient even as the dimension of $\boldsymbol{\Sigma}_{\mathbf{W}_j}$ grows. When estimating $\widehat{\mathbf{m}}_{jl}$, the computational cost with (7) is $O(n)$, whereas the direct applications of optimization in (6) incur a cost of $O(p^2q^2)$; see Remark 1 in Section S1 of the Supplementary Material. Recently, Banerjee et al. (2025) derived a closed-form debiasing matrix for the lasso, but their results do not extend to our covariate-dependent Gaussian setting with sparse group lasso; see Remark 2 in the Supplementary Material.

Importantly, the asymptotic results of $\widehat{\boldsymbol{\beta}}_j^u$ of (5), with $\widehat{\mathbf{M}}_j$ estimated based on $\widehat{\boldsymbol{\theta}}_{jl}$'s, can be established under the following assumptions.

ASSUMPTION 1: Suppose $\mathbf{u}^{(i)}$ are i.i.d mean zero random vectors with a covariance matrix satisfying $\phi_0 \geqslant \lambda_{\max}(\mathrm{Cov}(\mathbf{u}^{(i)})) \geqslant \lambda_{\min}(\mathrm{Cov}(\mathbf{u}^{(i)})) \geqslant 1/\phi_0$ for some constant $\phi_0 > 0$. Moreover, there exists a constant $M > 0$ such that $|u_h^{(i)}| \leqslant M$ for all i and h.

ASSUMPTION 2: Suppose that $\phi_1 \leqslant \lambda_{\min}(\mathrm{Cov}(\mathbf{z}^{(i)})) \leqslant \lambda_{\max}(\mathrm{Cov}(\mathbf{z}^{(i)})) \leqslant \phi_2$ for some constants $\phi_1, \phi_2 > 0$.

ASSUMPTION 3: The dimensions p, q, and the element-wise sparsity s_e satisfy $s_e(\log(p) + \log(q)) = o(\sqrt{n}/\log(n))$. Additionally, the maximum column ℓ_0 norm of $\boldsymbol{\Omega}(\mathbf{u})$ is bounded above by a positive constant $d_0 > 0$.

Assumption 1 characterizes the joint distribution of each row in $\mathbf{W}_j$. This condition is not restrictive, as genetic variants are often encoded as $0, 1$ or $0, 1, 2$ (Chen et al., 2016). Similar assumptions can be found in Zhang and Li (2023, 2025). Assumption 2 imposes bounded eigenvalues on $\mathrm{Cov}(\mathbf{z}^{(i)})$, which, alongside Assumption 1, provides a well-defined characterization of the joint distribution for each row in $\mathbf{W}_j$. Assumption 2 is mild and commonly used in the literature (Chen et al., 2016; Cai et al., 2022; Zhang and Li, 2023). Assumption 3 is also mild, as we assume $\sqrt{n} = \omega(s_e(\log p + \log q))$. This condition is less restrictive compared to Zhang and Li (2025), where s_e, $\log q$ and $\log q$ are required to grow no faster than $n^{1/6}$. Our assumption allows $s_e(\log p + \log q)$ to grow more slowly than $\sqrt{n}$, providing more flexibility. The condition on the ℓ_0 norm of $\mathbf{\Omega}(\mathbf{u})$ follows from $\beta_{jkh} = -(\mathbf{B}_h)_{jk}/\sigma^{jj}$, linking column sparsity of $\mathbf{\Omega}(\mathbf{u})$ to that of $\boldsymbol{\beta}$. This is not overly restrictive, as it only requires each response (e.g., a gene) to interact with finitely many others, a biologically plausible assumption also used in Zhang and Li (2025).

Let candidate models be those evaluated during tuning, and define s_λ as the maximum number of nonzero coefficients across them. We then state the following theorem.

THEOREM 2: *Suppose* $\boldsymbol{\beta} \in \mathbb{R}^{p(p-1)(q+1)}$ *is* (s_e, s_g)*-sparse, Assumptions 1-3 hold,* $s_e < s_\lambda$ *and* $s_\lambda \log(pq) = O(\sqrt{n}/\log n)$. *For some large constant* $C > 0$, *set*

$$\lambda_e = C\sqrt{\frac{2s_e \log(ep) + s_g \log(eq/s_g)}{ns_e}}, \quad \lambda_g = \sqrt{\frac{s_e}{s_g}}\lambda_e,$$

and let $\alpha = C\sqrt{\log(pq)/n}$, $\gamma = \sqrt{s_e/s_g}\cdot\alpha$, *then with probability at least* $1 - C_1\exp(-C_2\{s_e\log(ep) + s_g\log(eq/s_g)\}/s_e)$ *for some constants* $C_1, C_2 > 0$, *with the existence of* $\widehat{\mathbf{M}}_j$, *the SAGE estimator* $\widehat{\boldsymbol{\beta}}_j^u$ *defined by* (5) *can be decomposed into* $\sqrt{n}(\widehat{\boldsymbol{\beta}}_j^u - \boldsymbol{\beta}_j) = \boldsymbol{\Delta}_j + w_j$, *where*

$$\|\boldsymbol{\Delta}_j\|_\infty \precsim \frac{s_e(\log(p) + \log(q))}{\sqrt{n}}, \quad w_j|\mathbf{W}_j \sim \mathcal{N}(0, \frac{1}{\sigma^{jj}}\widehat{\mathbf{M}}_j^\top \widehat{\mathbf{\Sigma}}_{\mathbf{W}_j}\widehat{\mathbf{M}}_j),$$

and $\widehat{\mathbf{M}}_j$ *is obtained via* (7) *and Proposition 1. Also, for any* $l \in [(p-1)(q+1)]$ *it holds that*

$$\frac{\sqrt{n}\big((\widehat{\boldsymbol{\beta}}_j^u)_l - (\boldsymbol{\beta}_j)_l\big)}{\sqrt{\widehat{\mathbf{m}}_{jl}^\top \widehat{\mathbf{\Sigma}}_{\mathbf{W}_j}\widehat{\mathbf{m}}_{jl}}} \xrightarrow{d} \mathcal{N}(0, \frac{1}{\sigma^{jj}}).$$

The theorem and proof yield several advances. First, we establish the existence of a solution to (7) under suitable α and γ, despite the complex, non-sub-Gaussian design matrix $\mathbf{W}_j$, by deriving tailored probability bounds to control $\boldsymbol{\Delta}_j$. Second, we provide a nonvanishing lower bound for $\widehat{\mathbf{m}}_{jl}^\top \boldsymbol{\Sigma}_{\mathbf{W}_j} \widehat{\mathbf{m}}_{jl}$, ensuring estimator stability through advanced techniques beyond standard concentration bounds (e.g., Lemma 8). Third, since (4) lacks a closed form, the proof carefully handles this under weaker sparsity assumptions than Zhang and Li (2025), as detailed in Section S2 and Lemmas 4–5.

We may apply this theorem to construct confidence intervals. If σ^{jj} were known, an asymptotic $(1-\alpha)$-confidence interval of $(\boldsymbol{\beta}_j)_l$ would be

$$\left[(\widehat{\boldsymbol{\beta}}_j^u)_l - \frac{\Phi^{-1}(1-\alpha/2)}{\sigma^{jj}} \sqrt{\frac{\widehat{\mathbf{m}}_{jl}^\top \widehat{\boldsymbol{\Sigma}}_{\mathbf{W}_j} \widehat{\mathbf{m}}_{jl}}{n}},\ (\widehat{\boldsymbol{\beta}}_j^u)_l + \frac{\Phi^{-1}(1-\alpha/2)}{\sigma^{jj}} \sqrt{\frac{\widehat{\mathbf{m}}_{jl}^\top \widehat{\boldsymbol{\Sigma}}_{\mathbf{W}_j} \widehat{\mathbf{m}}_{jl}}{n}} \right]. \quad (8)$$

As proposed by Zhang and Li (2023), a natural and consistent estimator of $\frac{1}{\sigma^{jj}}$ is $\frac{1}{\widehat{\sigma}^{jj}} = \frac{1}{n-\widehat{s}_j} \|\mathbf{z}_j - \mathbf{W}_j \widehat{\boldsymbol{\beta}}_j^{OLS}\|_2^2$, where $\widehat{\boldsymbol{\beta}}_j^{OLS}$ is the OLS estimator constrained on the set $\widehat{\mathcal{S}}_j$, satisfying $\left(\widehat{\boldsymbol{\beta}}_j^{OLS}\right)_{\widehat{\mathcal{S}}_j} = \left((\mathbf{W}_j)_{\widehat{\mathcal{S}}_j}^\top (\mathbf{W}_j)_{\widehat{\mathcal{S}}_j}\right)^{-1} (\mathbf{W}_j)_{\widehat{\mathcal{S}}_j}^\top \mathbf{z}_j$ and $\left(\widehat{\boldsymbol{\beta}}_j^{OLS}\right)_{\widehat{\mathcal{S}}_j^c} = \mathbf{0}$. Here, $\widehat{\mathcal{S}}_j$ is the set indexing the non-zero estimates under the sparse group lasso. Hence, we can construct the confidence interval for $(\boldsymbol{\beta}_j)_l$ by replacing σ^{jj} with $\widehat{\sigma}^{jj}$ in (8). Moreover, for tests of linear contrasts, such as $(\boldsymbol{\beta}_j)_1 = (\boldsymbol{\beta}_j)_2$, the following corollary addresses the inference for $\mathbf{A}\boldsymbol{\beta}_j$ for fixed matrix $\mathbf{A}$. It forms the basis for testing, for example, whether a QTL modulate co-expressions in a gene pathway consisted of a set of genes.

COROLLARY 1: *Under the same condition of Theorem 2, for any fixed matrix* $\mathbf{A} \in \mathbb{R}^{K\times(p-1)(q+1)}$ *for some* $K \in \mathbb{N}$, *it holds that, with probability at least* $1-C_1\exp(-C_2\{s_e\log(ep)+s_g\log(eq/s_g)\}/s_e)$ *for some constants* $C_1, C_2 > 0$, *the vector* $\mathbf{A}\widehat{\boldsymbol{\beta}}_j^u - \mathbf{A}\boldsymbol{\beta}_j$ *can be decomposed into* $\sqrt{n}(\mathbf{A}\widehat{\boldsymbol{\beta}}_j^u - \mathbf{A}\boldsymbol{\beta}_j) = \boldsymbol{\Delta}_j + w_j$, *where*

$$\|\boldsymbol{\Delta}_j\|_\infty \leqslant \frac{C\|\mathbf{A}\|_\infty s_e(\log(p)+\log(q))}{\sqrt{n}}, \quad w_j \mid \mathbf{W}_j \sim \mathcal{N}_s(0, \frac{1}{\sigma^{jj}} \mathbf{A}\widehat{\mathbf{M}}_j^\top \widehat{\boldsymbol{\Sigma}}_{\mathbf{W}_j} \widehat{\mathbf{M}}_j \mathbf{A}^\top),$$

with $\widehat{\mathbf{M}}_j$ *obtained via (7) and Proposition 1.*

Corollary 1 provides a concrete framework for conducting inference on linear combinations of form $\mathbf{A}\boldsymbol{\beta}_j$. A direct application is testing whether $\Omega_{jk}(\mathbf{u}) = 0$ holds for all $\mathbf{u}$. Recalling the definition of $\boldsymbol{\beta}_j$ in Section 2.2, this hypothesis is equivalent to testing whether $\boldsymbol{\beta}_{jkh} = 0$ for all $h \in [1+q]$. This can be formulated by a specific matrix $\mathbf{A} \in \mathbb{R}^{(1+q)\times(p-1)(q+1)}$ that selects the relevant entries in $\boldsymbol{\beta}_j$. The constructed $\mathbf{A}$ satisfies $\|\mathbf{A}\|_\infty = 1$. Through this construction, Corollary 1 enables valid inference on functional edges across varying covariate levels.

4. Numerical Experiments

To evaluate the finite sample performance of the proposed SAGE estimator, we conduct numerical experiments to: [Aim (i)] assess the bias and variance of the estimates, along with the coverage probability of the confidence intervals; [Aim (ii)] examine how performance varies with sample size; [Aim (iii)] test the feasibility of conducting linear contrast tests; and [Aim (iv)] analyze computation time.

For Aim (i), we simulate n samples $\{(\mathbf{x}^{(i)}, \mathbf{u}^{(i)}) : i \in [n]\}$ from (2) with $n = 400$, where $\mathbf{x}^{(i)} \in \mathbb{R}^p$ represents p response variables, and $\mathbf{u}^{(i)} \in \mathbb{R}^q$ is the external covariate vector, such as SNPs. The elements $\mathbf{u}^{(i)}$ are generated independently from a Bernoulli(0.5) distribution. Given $\mathbf{u}^{(i)}$, we set $\Omega(\mathbf{u})_{jj} = \sigma^{jj} = 1$, and for the off-diagonal elements, we choose $h = 1, 2$ to allow non-zero values in $\mathbf{B}_h$, meaning the number of effective covariates is 2. For each $h \in \{1, 2\}$ (corresponding to the matrix $\mathbf{B}_h$), the values at positions $(j, j+1)$ and $(j+1, j)$ are set to 0.3, with $j = 1$. We also set $\mathbf{B}_0 = \mathbf{I}$. Once $\boldsymbol{\Omega}(\mathbf{u}^{(i)})$ are generated, we sample $\mathbf{x}^{(i)}$ from $\mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}^{-1}(\mathbf{u}^{(i)}))$ for each $i \in [n]$. For each simulation configuration, we generate 200 independent datasets. Based on the parameter setting, the non-zero coefficients in $\boldsymbol{\beta}$ are located at indexes $\text{ind}_1 = p$, $\text{ind}_2 = 2p - 1$, $\text{ind}_3 = (p-1)(q+1) + p$ and $\text{ind}_4 = (p-1)(q+1) + 2p - 1$, respectively.

As a benchmark for assessing biases, we first compute the pre-debiased multi-task learning estimates (4) (Zhang and Li, 2025). We then compare two methodologies for bias correction

and statistical inference. The first is the SAGE estimator $\widehat{\boldsymbol{\beta}}_j^u$ in the multi-task learning framework, with inference results provided in Theorem 2. In this methodology as well as for the pre-debiased multi-task estimates, we consider two fixed tuning parameter choices, $\lambda_e = 0.3$ and 0.6, both of order $\sqrt{\log(p)/n}$ (as required by the theorem), along with cross-validated λ_e and λ_g. For the fixed λ_e and λ_g, we set λ_g as $\lambda_e/\sqrt{2}$. The second is an oracle method, serving as a "golden standard" approach by assuming prior knowledge of the non-zero sets $\mathcal{S}_j$ and applying ordinary least squares (OLS) to each non-zero set $\mathcal{S}_j$ to obtain $\widehat{\boldsymbol{\beta}}_j^{oracle}$ for inference. To ensure comparability with the oracle estimates, we focus on specific sets in the debiasing procedure, such as $\mathcal{S}$ where $\boldsymbol{\beta}_{\mathcal{S}} \neq 0$. This approach also allows us to evaluate whether the debiased estimates within $\mathcal{S}$ exhibit a second-order normal distribution centered around the true values. The debiasing parameters are set to $\alpha = 1/\sqrt{n}$ and $\gamma = 2/\sqrt{n}$ to satisfy the theoretical requirements for asymptotic normality. Following Javanmard and Montanari (2014), these parameters influence finite-sample performance only at a second-order level. In other words, as long as the parameters scale correctly, the asymptotic properties of the debiased estimator remain unaffected.

[Figure 1 about here.]

We first visually assess the performance of the bias correction of the proposed debiased method by presenting histogram figures of the SAGE estimates $\widehat{\boldsymbol{\beta}}^u$ across 200 repetitions, plotting the histogram of $\widehat{\boldsymbol{\beta}}^u_{\text{ind}_1}$ for $(p, q) = (70, 120)$, respectively. We select the $\widehat{\boldsymbol{\beta}}$ obtained from $(\lambda_e, \lambda_g) = (0.3, 0.212)$ and show in Figure 1 how the SAGE estimates concentrate around the true parameter value of -0.3. As shown in Figure 1, the undebiased estimates display a clear bias away from the true value of -0.3. However, after applying our debiasing procedure, the SAGE estimates are tightly concentrated around -0.3. The resulting histogram appears nearly normal and symmetric around the line -0.3, demonstrating the approximate

normality of our proposed estimator. We next discuss the performance of statistical inference as reported in Table 1 over 200 data replicates.

[Table 1 about here.]

The tabulated results confirm that the proposed SAGE estimator consistently outperforms the undebiased estimator across all four indices. The undebiased estimates show larger pre-bias values, reflecting systematic bias, while the SAGE estimates achieve substantially reduced post-bias values, offering more accurate parameter estimates. Our SAGE estimator also performs comparably to the oracle estimator, which achieves minimal bias and near-nominal coverage probability. This highlights the robustness of the SAGE estimator in providing reliable inference without knowledge of the true support. Additionally, the theoretical variance aligns well with empirical variance, though variations in tuning parameters can affect this alignment. For instance, smaller tuning parameters may lead to model-based variance overestimating true variance, causing slight under-coverage in confidence intervals, likely due to finite sample effects. We further compare our method with the post-selection inference of Lee et al. (2016), implemented in the node-wise regression setting (3), using same datasets. Unlike their lasso-based approach, our method employs group lasso. While their procedure ensures valid coverage via polyhedral selection, in our covariate-dependent graphical regression the design matrix has a complex distribution, so the selection event is not polyhedral. Consequently, their method produces overly conservative intervals and low power, as shown in the last row of Table 1 and further in Section S3.1 of the Supplementary Material. By contrast, our SAGE estimator attains accurate coverage and substantially higher power. We also benchmark against the framework of Ning and Liu (2017) via point-wise tests, where our method again outperforms as shown in Section S3.2 of the Supplementary Material.

For Aim (ii), we conduct additional experiments with $n = 200, 800$ and present the results in Section S3.1 of the Supplementary Material. These results, along with Table 1, show

that when the sample size is small, the under-coverage of confidence intervals becomes more pronounced. However, as the sample size increases, such as $n = 800$, the empirical variance aligns more closely with the theoretical values and the bias of the SAGE estimates becomes closer to 0. We also vary p and q in Section S3.4 of the Supplementary Material and observe similar patterns.

For Aim (iii), we focus on the case of $j = 1$, where the non-zero indices for $\boldsymbol{\beta}_1$ are p and $2p - 1$. We estimate the following linear combinations and conduct testing for the corresponding linear contrasts respectively:

$$\text{I} : (\boldsymbol{\beta}_1)_p - (\boldsymbol{\beta}_1)_{2p-1}; \quad \text{II} : (\boldsymbol{\beta}_1)_1 - (\boldsymbol{\beta}_1)_p; \quad \text{III} : (\boldsymbol{\beta}_1)_1 + 2(\boldsymbol{\beta}_1)_2; \quad \text{IV} : \begin{pmatrix} 2(\boldsymbol{\beta}_1)_1 - (\boldsymbol{\beta}_1)_p \\ (\boldsymbol{\beta}_1)_2 + (\boldsymbol{\beta}_1)_p \end{pmatrix}.$$

Case I examines the relationship between two non-zero parameters, Case II compares a non-zero parameter with a zero parameter, Case III assesses the relationship between two zero parameters, and Case IV explores simultaneous inference. We set $(\lambda_e, \lambda_g) = (0.6, 0.424)$. Based on Corollary 1, the non-zero rows of matrix $\mathbf{A}$ corresponding to the four cases is given by $(\mathbf{e}_p - \mathbf{e}_{2p-1})^\top$, $(\mathbf{e}_1 - \mathbf{e}_{2p-1})^\top$, $(\mathbf{e}_1 + 2\mathbf{e}_2)^\top$ and $\binom{(2\mathbf{e}_1 - \mathbf{e}_p)^\top}{(\mathbf{e}_2 + \mathbf{e}_p)^\top}$, and the values of $\mathbf{A}\boldsymbol{\beta}$ for the four cases are $0, 0.3, 0$ and $(0.3, -0.3)^\top$.

[Table 2 about here.]

Table 2 presents the results based on 200 repetitions. As shown in Table 2, the results of Corollary 1 hold under the examined hypotheses. In particular, under Case IV which involves two different dimensions and shares the same $(\boldsymbol{\beta}_1)_p$ in both values, the standardized bivariate SAGE estimates, i.e., $\sqrt{n}(\mathbf{A}\widehat{\mathbf{M}}_j^\top \widehat{\boldsymbol{\Sigma}}_{\mathbf{W}_j} \widehat{\mathbf{M}}_j \mathbf{A}^\top)^{-1/2}(\mathbf{A}\widehat{\boldsymbol{\beta}}_1 - \mathbf{A}\boldsymbol{\beta}_1)$, approximately follows a two-dimensional normal distribution with mean 0 and an identity covariance matrix. The Q-Q plots in Figure 2 further confirm the asymptotic standard normal distribution. The joint distribution shown in Figure 2(d) shows a small Pearson correlation of only 0.057 between

the two dimensions, and the histograms for both variables exhibit a shape consistent with the standard normal distribution.

[Figure 2 about here.]

For Aim (iv), we conduct an experiment on our debiasing algorithm in (7), comparing it to the optimization problem in (6) similar to previous works (Javanmard and Montanari, 2014; Cai et al., 2022), to evaluate the computation time. For simplicity, we focus on solving (6) and (7), where we set $l = 1$ for the comparison of computation time. Table S.8 in Section S3.5 of the Supplementary Material compares the computation time between our debiasing procedure by (7) and the direct optimization of (6). The results indicate that the projection method is much faster. As p increases, the projection method remains feasible, while the direct optimization of (6) becomes increasingly computationally prohibitive.

5. Analysis of glioblastoma multiforme gene expression graphs

Glioblastoma multiforme is a lethal brain cancer, and existing therapies are largely ineffective (Kwiatkowska et al., 2013). To develop effective treatments, such as novel gene therapies, a better understanding of the disease's molecular mechanisms is critical. We apply our methods to infer the effects of single nucleotide polymorphisms (SNPs) on gene co-expression in a Glioblastoma trial. The dataset, publicly available via the NIH Gene Expression Omnibus database (labeled as GSE108476), comprises $n = 178$ glioblastoma multiforme patients with both microarray and SNP chip profiling.

Following preprocessing described by Gusev et al. (2018), we analyze expression levels of 73 glioma-pathway genes from KEGG (Kanehisa and Goto, 2000). Covariates include 118 nearby SNPs (within 2kb upstream and 0.5kb downstream, coded as "0/1"), plus age and gender, yielding 120 covariates and 317,988 parameters ($73 \times 36 \times 121$). We apply Gaussian graphical regression to assess their influence on network structure, tuning (λ_e, λ_g) by cross-validation

as in simulations, and then use the debiased SAGE estimator for inference. Parameters $\alpha = 1/\sqrt{178}$ and $\gamma = 2/\sqrt{178}$ are set per Theorem 2, affecting only second-order finite-sample performance; data-driven tuning may improve results but with more computational cost.

[Figure 3 about here.]

[Figure 4 about here.]

We first established the population-level gene network using multi-task learning estimates prior to debiasing (Figure 3(a)), then applied debiasing and significance testing. As shown in Figures 3(b) and 3(c), this reduced the number of edges, retaining only those with strong evidence. Biologically, the PI3K/AKT/MTOR pathway is central in glioblastoma (Network, 2008), with EGFR driving tumor progression via this pathway (Melenhorst et al., 2008). At $\alpha = 0.001$, EGFR shows few direct links but connects broadly through SHC4 (Figure 3(c)); at $\alpha = 0.05$, it exhibits numerous connections, underscoring its network impact. These findings are consistent with prior reports on EGFR's role in glioblastoma (Network, 2008) and with Zhang and Li (2025), supporting the importance of targeting interconnected pathways in treatment.

The fitted graphical regression identified nine eQTLs at $\alpha = 0.05$ ("rs10509346", "rs1347069", "rs6701524", "rs723210", "rs9303511", "rs503314", "rs728655", "rs759950", "rs306098"), of which three ("rs10509346", "rs1347069", "rs759950") remained significant at $\alpha = 0.001$. At this level, "rs10509346" revealed a positive MTOR–EGF and negative SHC2–RAF1 correlation (Figure 4(c)), both linked to cancer progression (Hua et al., 2019); "rs1347069" indicated upregulated PI3K/AKT signaling via AKT1–IGF1R (Figure 4(f)) (Manning and Toker, 2017); and "rs759950" showed a negative GADD45G–CAMK2A correlation (Figure 4(i)), suggesting opposing roles in growth regulation (Coultrap and Bayer, 2012). Additional eQTLs were significant at $\alpha = 0.05$, e.g., "rs6701524" with PDGFRB–CAMK1 and

"rs503314" with CCND1–CDKN2A (see Section S4 of the Supplementary Material). The full estimated network is shown in Figure 4. These findings may inform oncologists in tailoring treatments to patients' genetic profiles and tumor microenvironments (Li et al., 2023).

6. Discussions

We have developed a *segmentally adjusted graphical regression* (SAGE) estimator for multi-task Gaussian graphical regression models, enabling valid statistical inference in high-dimensional settings. In addition, we proposed a projection-based approach to compute the SAGE estimates, which simplifies the estimation of the inverse variance-covariance matrix and enhances computational efficiency. We have shown that the SAGE estimator asymptotically follows a normal distribution, paving the way for inference in network regression analysis, as further confirmed by our simulations.

In finite-sample settings, positive definiteness of the estimated $\mathbf{\Omega}(\mathbf{u})$ is not automatically guaranteed. To address this, we suggest a diagonal scaling strategy that enforces positive definiteness post estimation; see Zhang and Li (2023) for additional details on this post hoc rescaling approach. While this strategy is effective in practice, further work is needed to formally quantify the uncertainty introduced by this correction.

We take a frequentist approach, while recent Bayesian methods (Ni et al., 2022; Niu et al., 2024; Zeng et al., 2024) address low-dimensional settings; extending them to high-dimensional inference is a promising direction. Additionally, our inference procedure is currently designed to operate on each node individually. For simultaneous inference across multiple nodes, one could employ procedures such as the Bonferroni correction or false discovery rate (FDR) control. However, these methods do not explicitly account for the dependencies among nodes and may therefore suffer from reduced power. A more promising alternative could be a joint debiasing approach, which could improve efficiency and reduce variance by leveraging

the dependence structure across tasks. Developing such methods may pose theoretical and computational challenges, which we intend to investigate in future work.

Acknowledgments

The work was supported by NIH (Meng and Li) and NSF (DMS-2329296 and DMS-2326893; Zhang). We thank the Editor, AE, and referee for many insightful comments that improved the content and presentation of the paper.

Supplementary Material

Web Appendices S1 and S2 (referenced in Section 3); Appendix S3, Tables S.1–S.8 (Section 4); Appendix S4, Figures S.1–S.2 (Section 5); and the data and `MatLab` code for simulations and real data analyses (Sections 4 and 5; see our GitHub site) are available with this article at the *Biometrics* website on Oxford Academic.

Data Availability

The dataset is available from the National Institutes of Health Gene Expression Omnibus database (GSE108476) via https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108476.

References

Banerjee, S., Saunderson, J., Srivastava, R., and Rajwade, A. (2025). Fast debiasing of the lasso estimator. *arXiv preprint arXiv:2502.19825* .

Bellec, P. C. and Zhang, C.-H. (2022). De-biasing the lasso with degrees-of-freedom adjustment. *Bernoulli* **28,** 713–743.

Cai, T. T., Zhang, A. R., and Zhou, Y. (2022). Sparse group lasso: Optimal sample complexity, convergence rate, and statistical inference. *IEEE Transactions on Information Theory* **68,** 5975–6002.

Chen, M., Ren, Z., Zhao, H., and Zhou, H. (2016). Asymptotically normal and efficient estimation of covariate-adjusted Gaussian graphical model. *Journal of the American Statistical Association* **111,** 394–406.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* **21,** C1–C68.

Coultrap, S. J. and Bayer, K. U. (2012). CaMKII regulation in information processing and storage. *Trends in Neurosciences* **35,** 607–618.

Fei, Z. and Li, Y. (2021). Estimation and inference for high dimensional generalized linear models: A splitting and smoothing approach. *Journal of Machine Learning Research* **22,** 1–32.

Fei, Z., Zhu, J., Banerjee, M., and Li, Y. (2019). Drawing inferences for high-dimensional linear models: A selection-assisted partial regression and smoothing approach. *Biometrics* **75,** 551–561.

Golub, G. H. and Reinsch, C. (1971). Singular value decomposition and least squares solutions. In *Handbook for Automatic Computation: Volume II: Linear Algebra*, pages 134–151. Springer.

Gusev, Y., Bhuvaneshwar, K., Song, L., Zenklusen, J.-C., Fine, H., and Madhavan, S. (2018). The REMBRANDT study, a large collection of genomic data from brain cancer patients. *Scientific Data* **5,** 180158.

Hua, H., Kong, Q., Zhang, H., Wang, J., Luo, T., and Jiang, Y. (2019). Targeting mTOR for cancer therapy. *Journal of Hematology & Oncology* **12,** 1–19.

Hudson, A. and Shojaie, A. (2022). Covariate-adjusted inference for differential analysis of high-dimensional networks. *Sankhya A* **84,** 345–388.

Javanmard, A. and Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research* **15,** 2869–2909.

Kanehisa, M. and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28,** 27–30.

Kwiatkowska, A., Nandhu, M. S., Behera, P., Chiocca, E. A., and Viapiano, M. S. (2013). Strategies in gene therapy for glioblastoma. *Cancers* **5,** 1271–1305.

Lee, J. D., Sun, D. L., Sun, Y., and Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics* **44,** 907–927.

Li, S., Cai, T. T., and Li, H. (2023). Transfer learning in large-scale gaussian graphical models with false discovery rate control. *Journal of the American Statistical Association* **118,** 2171–2183.

Li, S., Schmid, K. T., de Vries, D. H., Korshevniuk, M., Losert, C., Oelen, R., van Blokland, I. V., BIOS Consortium, s.-e. C., Groot, H. E., Swertz, M. A., et al. (2023). Identification of genetic variants that impact gene co-expression relationships using large-scale single-cell data. *Genome Biology* **24,** 80.

Manning, B. D. and Toker, A. (2017). AKT/PKB signaling: navigating the network. *Cell* **169,** 381–405.

Melenhorst, W. B., Mulder, G. M., Xi, Q., Hoenderop, J. G., Kimura, K., Eguchi, S., and van Goor, H. (2008). Epidermal growth factor receptor signaling in the kidney: key roles in physiology and disease. *Hypertension* **52,** 987–993.

Mifflin, R. (1977). Semismooth and semiconvex functions in constrained optimization. *SIAM Journal on Control and Optimization* **15,** 959–972.

Network, C. G. A. R. (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455,** 1061–1068.

Ni, Y., Stingo, F. C., and Baladandayuthapani, V. (2022). Bayesian covariate-dependent gaussian graphical models with varying structure. *Journal of Machine Learning Research* **23,** 1–29.

Ning, Y. and Liu, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *THE ANNALS* **45,** 158–195.

Niu, Y., Ni, Y., Pati, D., and Mallick, B. K. (2024). Covariate-assisted bayesian graph learning for heterogeneous data. *Journal of the American Statistical Association* **119,** 1985–1999.

Peng, J., Wang, P., Zhou, N., and Zhu, J. (2009). Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association* **104,** 735–746.

Saegusa, T. and Shojaie, A. (2016). Joint estimation of precision matrices in heterogeneous populations. *Electronic Journal of Statistics* **10,** 1341.

Wang, Y., Joseph, S. J., Liu, X., Kelley, M., and Rekaya, R. (2012). SNPxGE2: a database for human snp–coexpression associations. *Bioinformatics* **28,** 403–410.

Xia, D., Zhang, A. R., and Zhou, Y. (2022). Inference for low-rank tensors—no need to debias. *The Annals of Statistics* **50,** 1220–1245.

Zeng, Z., Li, M., and Vannucci, M. (2024). Bayesian covariate-dependent graph learning with a dual group spike-and-slab prior. *arXiv preprint arXiv:2409.17404* .

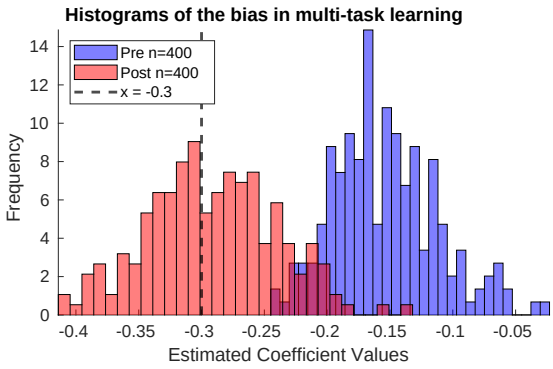
Zhang, C.-H. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **76,** 217–242.

Zhang, J. and Li, Y. (2023). High-dimensional Gaussian graphical regression models with covariates. *Journal of the American Statistical Association* **118,** 2088–2100.

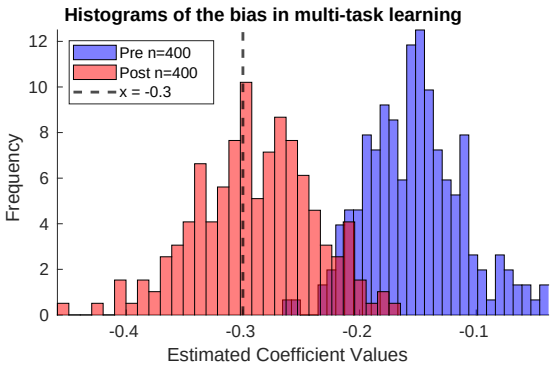
Zhang, J. and Li, Y. (2025). Multi-task learning for Gaussian graphical regressions with high dimensional covariates. *Journal of Computational and Graphical Statistics* **34,** 961–970.

Zhang, J., Sun, W. W., and Li, L. (2023). Generalized connectivity matrix response regression with applications in brain connectivity studies. *Journal of Computational and Graphical Statistics* **32,** 252–262.

Zhu, Y. and Bradic, J. (2018). Linear hypothesis testing in dense high-dimensional linear models. *Journal of the American Statistical Association* **113,** 1583–1600.

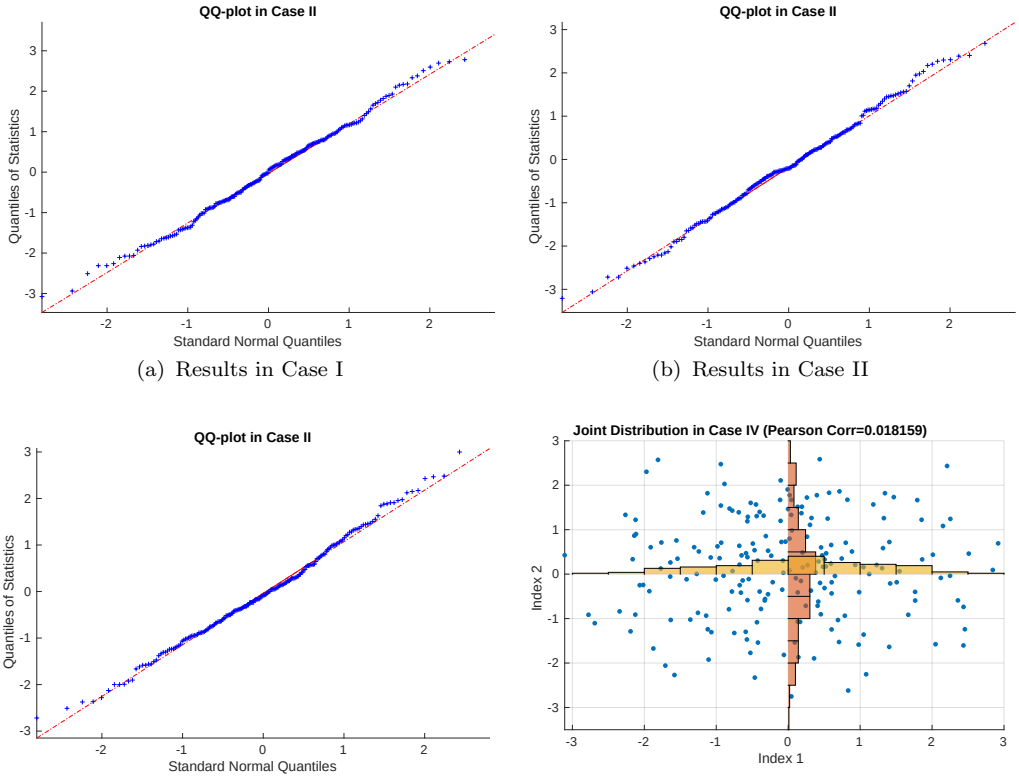


(a) Histogram with $\boldsymbol{\beta}_{\text{ind}_1}$



(b) Histogram with $\boldsymbol{\beta}_{\text{ind}_3}$

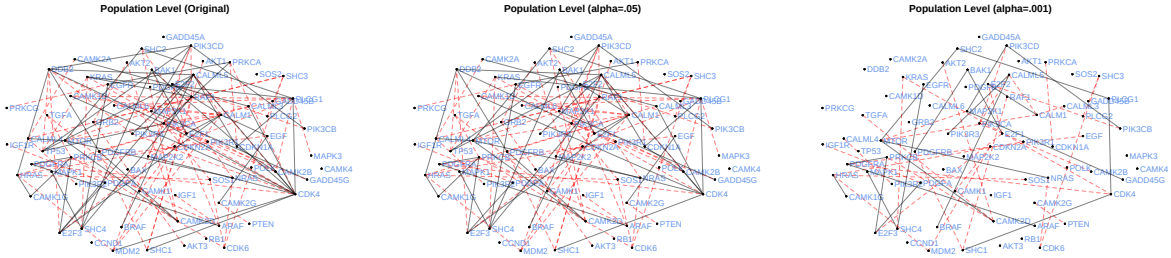
Figure 1: Histograms of pre-debiased estimates (referred to as Pre) and SAGE estimates (referred to as Post) for $\boldsymbol{\beta}_{\text{ind}_1}$ and $\boldsymbol{\beta}_{\text{ind}_3}$ with $p = 70$ and $q = 120$.



(a) Results in Case I (b) Results in Case II

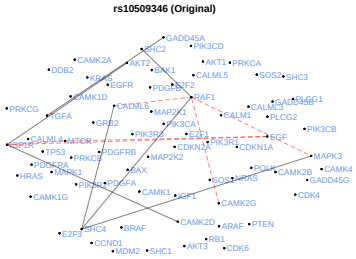
(c) Results in Case III (d) Results in Case IV

Figure 2: The example figures demonstrate the asymptotic standard normal distribution behavior of the standardized SAGE estimates, $\sqrt{n}(\mathbf{A}\widehat{\mathbf{M}}_j^\top \widehat{\boldsymbol{\Sigma}}_{\mathbf{W}_j}\widehat{\mathbf{M}}_j\mathbf{A}^\top)^{-1/2}(\mathbf{A}\widehat{\boldsymbol{\beta}}_1 - \mathbf{A}\boldsymbol{\beta}_1)$, in the cases with $(p, q) = (70, 120)$. Figures 2(a), 2(b) and 2(c) present QQ plots for the Case I, Case II and Case III, respectively, illustrating the asymptotic normality. Figure 2(d) shows the joint distribution of two asymptotically independent standard normal variables under Case IV.

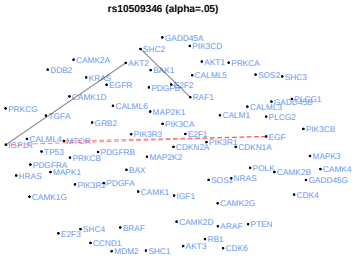


(a) Original Selected Edges (b) Significance Level 0.05 (c) Significance Level 0.001

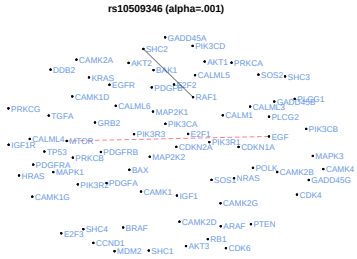
Figure 3: Population-level gene co-expression graph (left), shown with significance levels of 0.05 (middle) and 0.001 (right). Positive partial correlations are shown with red dashed lines, while negative correlations are indicated by black solid lines.



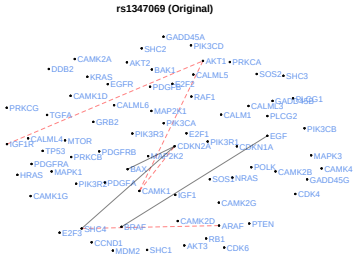
(a) Original Selected Edges



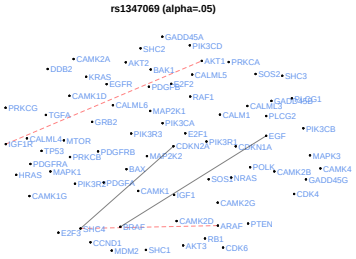
(b) Significance Level 0.05



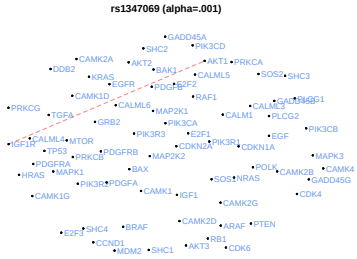
(c) Significance Level 0.001



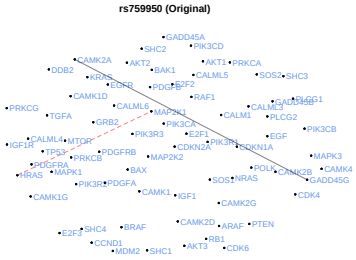
(d) Original Selected Edges



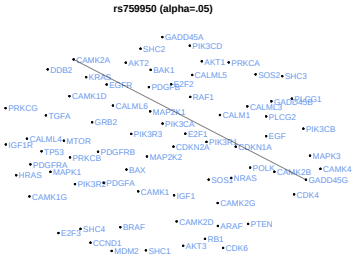
(e) Significance Level 0.05



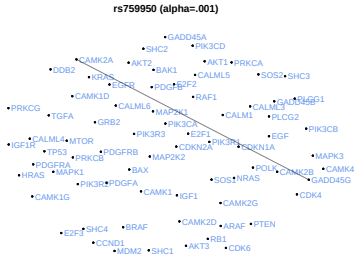
(f) Significance Level 0.001



(g) Original Selected Edges



(h) Significance Level 0.05



(i) Significance Level 0.001

Figure 4: The first, second and third rows display the SNP effects for "rs10509346", "rs1347069" and "rs759950", respectively, at significance levels of 0.05 (middle) and 0.001 (right). Positive partial correlations are shown with red dashed lines, while negative correlations are indicated by black solid lines

Table 1: Simulation results with $n = 400$. Standard deviations are shown in parentheses. **Pre-Bias**: the average bias of $\widehat{\boldsymbol{\beta}}$ prior to debiasing; **Post-Bias**: the average bias of the SAGE estimates post debiasing; **Emp-SD**: the empirical standard deviation of debiased estimates after standardization with the theoretical value being 1; **Cov-Prob**: the estimated coverage probability of the 95% confidence interval; **Rej-0**: the probability that the statistical inference rejects the true parameter being zero; **AIL**: average interval length.

(λ_e, λ_g)		$\boldsymbol{\beta}_{\text{ind}_1}$	$\boldsymbol{\beta}_{\text{ind}_2}$	$\boldsymbol{\beta}_{\text{ind}_3}$	$\boldsymbol{\beta}_{\text{ind}_4}$
(.3,.212)	Pre-Bias	.177(.052)	.175(.054)	.178(.049)	.175(.053)
	Post-Bias	.051(.073)	.050(.075)	.055(.068)	.053(.073)
	EmpSD	1.23	1.30	1.22	1.30
	Cov-Prob	84.5%	87.5%	85.5%	88.5%
	Rej-0	95.5%	96.5%	98.0%	96.0%
(.6,.414)	Pre-Bias	.281(.018)	.280(.018)	.281(.018)	.280(.018)
	Post-Bias	.024(.057)	.025(.060)	.031(.056)	.029(.055)
	Emp-SD	1.02	1.07	1.04	1.04
	Cov-Prob	93.5%	92.5%	93.5%	92.5%
	Rej-0	99.5%	99.5%	99.5%	100%
Cross Validation	Pre-Bias	.157(.052)	.156(.049)	.157(.052)	.156(.050)
	Post-Bias	.006(.050)	.006(.048)	.008(.049)	.006(.048)
	Emp-SD	1.28	1.20	1.25	1.24
	Cov-Prob	87.0%	89.5%	87.5%	89.5%
	Rej-0	100%	100%	100%	100%
	AIL	.074(.002)	.074(.002)	.074(.002)	.074(.002)
$\widehat{\boldsymbol{\beta}}^{oracle}$	Bias	-.002(.049)	-.003(.046)	-.000(.047)	-.001(.048)
	Emp-SD	.971	.912	.944	.952
	Cov-Prob	96.5%	98.0%	96.0%	97.0%
	Rej-0	100%	100%	100%	100%
	AIL	.088			
Post Selection	Bias	.030(.055)	.034(.052)	.031(.054)	.034(.053)
	Cov-Prob	97.5%	98.0%	96.0%	94.5%
	AIL	1.07(1.00)	1.00(0.95)	1.04(1.02)	1.02(1.05)
	Rej-0	63.0%	68.0%	68.5%	62.0%

Table 2: Results for the four cases with $n = 400$ and $n = 800$. **Emp-AVE**: empirical mean of SAGE estimates after standardization with the theoretical value being 0; **Emp-SD**: empirical standard deviation of SAGE estimates after standardization with the theoretical value being 1; **Cov-Prob**: estimated coverage probability of the 95% confidence interval. V1 and V2 correspond to the two vector values being tested in Case IV.

(p, q)		I	II	III	IV	
					V1	V2
$n = 400$	Emp-Ave	-.003	-.148	-.021	-.029	.162
	Emp-SD	1.23	1.22	1.12	1.25	1.16
	Cov-Prob	92.5%	89.0%	94.5%	90.0%	
$n = 800$	Emp-Ave	.006	-.118	-.076	-.126	.073
	Emp-SD	1.10	1.14	1.09	1.15	1.06
	Cov-Prob	93%	92.5%	93.5%	91.0%	