

Statistica Sinica Preprint No: SS-2026-0078	
Title	A Unified Debiasing Framework for Inference with Jointly Regularized Multi-Graph Estimators
Manuscript ID	SS-2026-0078
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202026.0078
Complete List of Authors	Qiao Zheng, Gefei Li, Yibo Yan and Riquan Zhang
Corresponding Authors	Riquan Zhang
E-mails	zhangriquan@163.com
Notice: Accepted author version.	

A Unified Debiasing Framework for Inference with Jointly Regularized Multi-Graph Estimators

Qiao Zheng^{1,†}, Gefei Li^{2,†}, Yibo Yan¹ and Riquan Zhang¹

¹ *Shanghai University of International Business and Economics*

² *East China Normal University*

Abstract: Although joint regularization has become a widely used tool for structured multi-graph estimation, statistical inference for multi-graph models is challenging because of the nontrivial coupling effects induced by cross-group penalties. To address this issue, we develop a KKT-based unified debiasing framework for inference with jointly regularized multi-graph estimators and derive an analytic debiasing representation that removes the leading regularization-induced bias while accounting for cross-group coupling. The resulting debiasing representation also admits an identity-based interpretation through the precision–covariance relation, highlighting its applicability beyond likelihood-based constructions to a broad class of multi-graph estimators satisfying the required consistency conditions under standard high-dimensional sparsity regimes. We establish theoretical guarantees, including estimation consistency and entry-wise asymptotic normality, and develop inferential procedures that provide entry-wise confidence intervals and simultaneous tests with asymptotically valid control of the family-wise error rate (FWER) and the false discovery rate (FDR). We further apply the proposed framework to differential structure detection to identify structural changes across groups. Extensive simulation studies and a real-data analysis demonstrate the validity and effectiveness of the proposed debiasing framework for high-dimensional multi-graph inference.

Key words and phrases: Gaussian graphical models, High-dimensional statistics, Joint regularization, Multi-graph models, Simultaneous inference.

Corresponding author: Riquan Zhang. E-mail: zhangriquan@163.com

[†] These authors contributed equally to this work.

1. Introduction

Graphical models provide a principled framework for representing conditional dependence among multivariate random variables, with applications in areas such as biological networks and brain connectivity (Belilovsky et al., 2016; Li et al., 2019). In Gaussian graphical models (GGMs) (Lauritzen, 1996), conditional independence corresponds to zeros in the precision matrix, so recovering the graph reduces to identifying the nonzero pattern of the precision matrix. In modern high-dimensional settings where p is comparable to or larger than n , sparse regularized estimators have become standard tools for learning graphical model structure (Meinshausen and Bühlmaan, 2006; Friedman et al., 2008; Cai et al., 2011).

In many contemporary applications, however, data are collected from multiple related populations rather than a single homogeneous population. A natural strategy is to estimate a separate graphical model for each population, but this can be statistically inefficient and unstable when the sample size for each population is limited. A primary limitation is that it fails to exploit structural information shared across populations (Danaher et al., 2014; Lee and Liu, 2015). For instance, gene regulatory networks from different breast cancer subtypes (e.g., ER+ and ER−) may differ in specific regulatory interactions while sharing a large set of common edges (TCGA, 2008). Such considerations have motivated jointly regularized methods that borrow strength across multiple graphical models.

A substantial literature has studied the joint estimation of multiple graphical models via structured cross-group regularization (Guo et al., 2011; Danaher et al., 2014; Hao et al., 2018; Wang et al., 2020; Majumdar and Michailidis, 2022). Existing approaches range from global sharing schemes (e.g., encouraging a common sparse support across groups) to more

flexible pairwise or weighted coupling schemes that allow for varying degrees of similarity among groups. We refer the reader to [Tsai et al. \(2022\)](#) for a comprehensive review.

As noted by [Tsai et al. \(2022\)](#), although methods for joint estimation have been developed extensively, corresponding advances in statistical inference remain comparatively limited. Existing methods for multi-graph inference are often built on specific regression formulations and tailored to particular inferential targets ([Xia et al., 2015](#); [Belilovsky et al., 2016](#); [Liu, 2017](#)). These methods provide useful tools in their respective settings, but unified inference for a broader class of jointly regularized multi-graph estimators has received relatively little systematic study. Beyond the multi-graph-specific literature, [Neykov et al. \(2018\)](#) develop a unified estimating-equation framework for inference, with CLIME-based graphical model inference as one particular application. Their work provides a closely related perspective to the debiasing principle considered here, while jointly regularized multi-graph estimation and the corresponding inferential procedures are not specifically studied in their work.

To address this gap, we develop a unified debiasing framework for a broad class of jointly regularized multi-graph estimators. Within the Gaussian penalized-likelihood formulation, we use the coupled KKT conditions ([Kuhn and Tucker, 1951](#)) associated with the jointly penalized optimization problem to identify the leading regularization-induced bias and derive an explicit bias correction for each group. The resulting debiasing formula has the same algebraic form as the classical single-graph debiasing formula in [Janková and van de Geer \(2015\)](#). We further reinterpret this formula through the fundamental precision–covariance identity, yielding an algebraic error decomposition. Our main contributions are summarized

as follows:

1. We develop a unified debiasing framework applicable to a broad class of jointly regularized multi-graph estimators. Although the debiasing formula is derived within the Gaussian penalized-likelihood formulation, the algebraic error decomposition allows us to establish its inferential validity for a broader class of initial estimators, including those not necessarily obtained from penalized likelihood.
2. Under appropriate regularity conditions on the initial estimators, we establish entrywise consistency and asymptotic normality of the resulting debiased estimators. Building on these results, we construct confidence intervals and hypothesis tests for individual precision-matrix entries, together with multiple-testing procedures that provide asymptotically valid control of the family-wise error rate (FWER) and the false discovery rate (FDR).
3. We demonstrate the extensibility of the proposed unified debiasing framework across multiple graphs. Specifically, we apply the framework to differential structure detection, establish asymptotic normality for the linear contrast functional of the debiased estimators, and derive corresponding tests for differential edges.
4. Through simulation studies, we evaluate the finite-sample performance of the proposed framework in terms of Gaussian approximation, confidence-interval coverage, and empirical FWER and FDR calibration. An analysis of the WebKB text networks illustrates how the proposed framework can be applied to conduct category-specific edge inference with FDR adjustment, starting from a jointly regularized multi-graph

estimator.

The notation used in this paper is summarized below. For a p -dimensional vector $\mathbf{a} = (a_1, \dots, a_p)^\top \in \mathbb{R}^p$, define the ℓ_q norm as $\|\mathbf{a}\|_q = (\sum_{i=1}^p |a_i|^q)^{1/q}$ for $0 < q < \infty$. Let $\|\mathbf{a}\|_0$ denote the number of nonzero elements in $\mathbf{a}$, and $\|\mathbf{a}\|_\infty := \max_i |a_i|$. For a $p \times q$ matrix $\mathbf{A} = (a_{ij})_{p \times q} \in \mathbb{R}^{p \times q}$, we define the matrix ℓ_1 , ℓ_F , and $\ell_{\max}$ norms as $\|\mathbf{A}\|_1 = \max_j \sum_i |a_{ij}|$, $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} a_{ij}^2}$, and $\|\mathbf{A}\|_{\max} = \max_{i,j} |a_{ij}|$, respectively. For a non-empty set $\mathcal{F}$, denote its cardinality as $|\mathcal{F}|$. For two sequences of real numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n = O(b_n)$ or $a_n \lesssim b_n$ if there exists a constant C such that $|a_n| \leq C|b_n|$ holds for sufficiently large n . We define $a_n \asymp b_n$ if $a_n = O(b_n)$ and $b_n = O(a_n)$, and $a_n = o(b_n)$ if $\lim_{n \rightarrow \infty} a_n/b_n = 0$.

2. Methodology

Suppose we have a heterogeneous dataset with p variables and M groups. $\mathbf{x}^{(m)} = (\mathbf{x}_1^{(m)}, \dots, \mathbf{x}_{n_m}^{(m)})^\top \in \mathbb{R}^{n_m \times p}$ denotes the observations from the m^{th} group, which contains n_m observations, where each $\mathbf{x}_k^{(m)} \in \mathbb{R}^p$ is a p -dimensional row vector. Without loss of generality, we assume that observations in the same group are centered along each variable, i.e., $\sum_{k=1}^{n_m} \mathbf{x}_k^{(m)} = \mathbf{0}$. Further, we assume that $\mathbf{x}_1^{(m)}, \dots, \mathbf{x}_{n_m}^{(m)}$ are i.i.d. from a p -variate Gaussian distribution with mean zero and covariance matrix $\Sigma^{(m)} \in \mathbb{R}^{p \times p}$. For notational simplicity, we additionally assume that the sample sizes are balanced across groups, i.e., $n_1 = \dots = n_M = n$. Denote by $\Theta^{(m)} = (\Sigma^{(m)})^{-1}$ the corresponding precision matrix and define the sample covariance matrix as $\hat{\Sigma}^{(m)} = \frac{1}{n_m} \sum_{k=1}^{n_m} \mathbf{x}_k^{(m)} (\mathbf{x}_k^{(m)})^\top$.

2.1 Unified Penalized Likelihood Framework for Multiple Graphs

2.1 Unified Penalized Likelihood Framework for Multiple Graphs

We begin by introducing a unified penalized likelihood formulation for estimating multiple precision matrices. Consider a p -variate Gaussian distribution with mean zero and covariance matrix Σ . For a precision matrix $\Theta = \Sigma^{-1}$, the log-likelihood (up to an additive constant) is given by $\ell(\Theta) = \log \det(\Theta) - \text{tr}(\hat{\Sigma}\Theta)$, where $\hat{\Sigma}$ denotes the empirical covariance matrix. For M related groups in the multi-graph setting, the corresponding joint negative log-likelihood is $\mathcal{L}(\Theta) = \sum_{m=1}^M \left\{ -\log \det(\Theta^{(m)}) + \text{tr}(\hat{\Sigma}^{(m)}\Theta^{(m)}) \right\}$, where $\Theta = (\Theta^{(1)}, \dots, \Theta^{(M)})$ collects the precision matrices for all groups.

To regularize the high-dimensional parameter space and incorporate structural relationships across groups, we consider the unified penalized likelihood estimator

$$\hat{\Theta} = \arg \min_{\Theta} \{ \mathcal{L}(\Theta) + P(\Theta, \lambda) \}, \quad (2.1)$$

where $P(\Theta, \lambda)$ denotes a general penalty function that imposes structural constraints on the collection of precision matrices and λ collects the associated tuning parameters. In practice, the penalty consists of additive components—for example, a within-group component that encourages sparsity in each individual precision matrix, together with one or more cross-group components that encode structural information shared across groups. More elaborate constructions, such as hierarchical or overlapping penalties that simultaneously couple within- and across-group structures, can also be accommodated within this unified formulation.

The unified formulation encompasses many well-known estimators as special cases. For

2.2 Debiasing for the Unified Penalized Likelihood Framework

instance, if the penalty takes the additive form $P(\Theta, \lambda) = \lambda_1 \sum_{m=1}^M \|\Theta^{(m)}\|_{1,\text{off}}$, then Eq. (2.1) reduces to solving M separate graphical lasso problems (Friedman et al., 2008). More prominent multi-graph estimators—such as JGL (Guo et al., 2011), FGL/GGL (Danaher et al., 2014), LASICH (Saegusa and Shojaie, 2016), and several recent variants—arise from appropriate choices of $P(\Theta, \lambda)$. This highlights the flexibility of the unified penalized likelihood formulation in accommodating both within-group sparsity and structured similarity across groups.

Remark 1. *While the formulation in Eq. (2.1) accommodates a broad class of penalty functions, our theoretical analysis requires the penalty to contain at least one component that promotes sparsity in the precision matrices. This sparsity-inducing mechanism ensures compatibility with the sparse parameter space introduced below and is essential for controlling high-dimensional estimation errors and establishing the validity of the debiasing procedure. Such penalties include entrywise ℓ_1 penalty (Tibshirani, 1996), folded concave penalties such as SCAD (Fan and Li, 2001) and MCP (Zhang, 2010). In multi-graph estimation, these sparsity-inducing components can be combined with structured penalties, such as group (Yuan and Lin, 2006) or fused penalties (Tibshirani et al., 2005), to additionally encode shared structure across graphs.*

2.2 Debiasing for the Unified Penalized Likelihood Framework

Due to the regularization term in Eq. (2.1), the penalized estimator $\hat{\Theta} = (\hat{\Theta}^{(1)}, \dots, \hat{\Theta}^{(M)})$ cannot be directly used for statistical inference. Therefore, we seek to construct a debiased estimator by removing the leading bias induced by regularization.

To derive the debiasing correction under the unified penalized likelihood framework, we

2.2 Debiasing for the Unified Penalized Likelihood Framework

begin with the KKT conditions for the optimization problem in Eq. (2.1). In derivative form, the KKT conditions are given by

$$\widehat{\Sigma}^{(m)} - \left(\widehat{\Theta}^{(m)}\right)^{-1} = - \left. \frac{\partial P(\Theta, \lambda)}{\partial \Theta^{(m)}} \right|_{\Theta = \widehat{\Theta}}, \quad m = 1, \dots, M, \quad (2.2)$$

where the partial derivative of $P(\Theta, \lambda)$ with respect to $\Theta^{(m)}$ is interpreted as any valid subgradient whenever the penalty term is not differentiable.

Pre- and post-multiplying Eq. (2.2) by $\widehat{\Theta}^{(m)}$ yields $\widehat{\Theta}^{(m)}\widehat{\Sigma}^{(m)}\widehat{\Theta}^{(m)} - \widehat{\Theta}^{(m)} = -\widehat{\Theta}^{(m)}G^{(m)}\widehat{\Theta}^{(m)}$, where $G^{(m)} = \left. \frac{\partial P(\Theta, \lambda)}{\partial \Theta^{(m)}} \right|_{\Theta = \widehat{\Theta}}$ and $m = 1, \dots, M$. Define $W^{(m)} := \widehat{\Sigma}^{(m)} - \Sigma^{(m)}$. Rearranging the equation above yields that

$$\widehat{\Theta}^{(m)} + \widehat{\Theta}^{(m)}G^{(m)}\widehat{\Theta}^{(m)} - \Theta^{(m)} = -\Theta^{(m)}W^{(m)}\Theta^{(m)} + R^{(m)}, \quad m = 1, \dots, M, \quad (2.3)$$

where $R^{(m)} = -\left(\widehat{\Theta}^{(m)} - \Theta^{(m)}\right)W^{(m)}\Theta^{(m)} - \left(\widehat{\Theta}^{(m)}\widehat{\Sigma}^{(m)} - \mathbf{I}\right)\left(\widehat{\Theta}^{(m)} - \Theta^{(m)}\right)$ and $\mathbf{I}$ denotes the $p \times p$ identity matrix. The term $\widehat{\Theta}^{(m)}G^{(m)}\widehat{\Theta}^{(m)}$ represents the leading bias induced by regularization. While this bias typically converges to zero as the sample size increases, it is generally non-negligible at the $\sqrt{n}$ -scale and therefore distorts the entry-wise asymptotic distribution of $\widehat{\Theta}^{(m)}$. As a result, this bias needs to be explicitly removed in order to obtain valid inferential procedures. Under suitable regularity conditions, the remainder term $R^{(m)}$ is asymptotically negligible and the leading stochastic term $\Theta^{(m)}W^{(m)}\Theta^{(m)}$ is element-wise asymptotically normal for each $m = 1, \dots, M$; see Theorem 3.1.

Based on the preceding analysis, a natural debiasing step is to remove the bias component $\widehat{\Theta}^{(m)}G^{(m)}\widehat{\Theta}^{(m)}$ from the penalized estimator. Specifically, rearranging Eq. (2.3) yields the

2.3 Debiasing Beyond the Unified Penalized Likelihood Framework

approximation $\widehat{\Theta}^{(m)} \approx \Theta^{(m)} - \Theta^{(m)}W^{(m)}\Theta^{(m)} - \widehat{\Theta}^{(m)}G^{(m)}\widehat{\Theta}^{(m)}$. Therefore, we obtain the following debiased estimator under the unified penalized likelihood framework:

$$\widetilde{\Theta}^{(m)} = 2\widehat{\Theta}^{(m)} - \widehat{\Theta}^{(m)}\widehat{\Sigma}^{(m)}\widehat{\Theta}^{(m)}, \quad m = 1, \dots, M. \quad (2.4)$$

Stacking these estimators gives $\widetilde{\Theta} = (\widetilde{\Theta}^{(1)}, \dots, \widetilde{\Theta}^{(M)})$.

It is worth noting that the debiasing formula in Eq. (2.4) coincides algebraically with the classical debiased precision-matrix estimator for the single-graph setting (see, e.g., [Janková and van de Geer \(2015\)](#)). In the multi-graph setting considered here, the structural information shared across graphs has already been incorporated into the penalized estimators $\widehat{\Theta}^{(m)}$ through the coupled optimization problem. Thus, although the debiasing formula in Eq. (2.4) has the same algebraic form as in the single-graph setting, the distinction lies in the initial estimators to which the correction is applied: here, the initial precision-matrix estimators already incorporate structural information shared across graphs.

2.3 Debiasing Beyond the Unified Penalized Likelihood Framework

Although the closed-form debiased estimator in Eq. (2.4) is derived from the Gaussian penalized likelihood, it admits a more general interpretation based only on the fundamental precision-covariance identity $\Theta^{(m)}\Sigma^{(m)} = \mathbf{I}$. In particular, the construction does not require the initial estimators $\widehat{\Theta}^{(m)}$ to be obtained from likelihood-based procedures. Consequently, this debiasing formula can be applied to a broad class of multi-graph estimators, such as constrained ℓ_1 -type estimators ([Lee and Liu, 2015](#); [Cai et al., 2016](#); [Wang et al., 2017](#)).

The parameter of interest for group m is the precision matrix $\Theta_0^{(m)} = (\Sigma_0^{(m)})^{-1}$, where

2.3 Debiasing Beyond the Unified Penalized Likelihood Framework

$\Sigma_0^{(m)}$ denotes the population covariance (or a model-implied covariance-type quantity) of the observed variables. The identity $\Theta_0^{(m)} \Sigma_0^{(m)} = \mathbf{I}$ implies $\Theta_0^{(m)} \Sigma_0^{(m)} \Theta_0^{(m)} - \Theta_0^{(m)} = \mathbf{0}$. Motivated by the corresponding sample analogue, we define the residual mapping

$$\Psi^{(m)}(\Theta) := \Theta \hat{S}^{(m)} \Theta - \Theta, \quad (2.5)$$

where $\hat{S}^{(m)}$ is a covariance-type estimator of $\Sigma_0^{(m)}$ computed from the sample data. At the truth, $\Psi^{(m)}(\Theta_0^{(m)}) = \Theta_0^{(m)}(\hat{S}^{(m)} - \Sigma_0^{(m)})\Theta_0^{(m)}$, which is small under standard concentration of $\hat{S}^{(m)}$. Accordingly, we consider the following debiased estimator

$$\tilde{\Theta}^{(m)} := \hat{\Theta}^{(m)} - \Psi^{(m)}(\hat{\Theta}^{(m)}) = 2\hat{\Theta}^{(m)} - \hat{\Theta}^{(m)} \hat{S}^{(m)} \hat{\Theta}^{(m)}, \quad m = 1, \dots, M. \quad (2.6)$$

This interpretation clarifies that the debiasing step depends only on the precision–covariance identity and the availability of $\hat{S}^{(m)}$, rather than on the particular estimation procedure used to obtain $\hat{\Theta}^{(m)}$. When $\hat{S}^{(m)} = \hat{\Sigma}^{(m)}$, Eq. (2.6) reduces exactly to Eq. (2.4).

Remark 2. *The use of the precision–covariance relation for high-dimensional graphical-model inference has also appeared in [Neykov et al. \(2018\)](#). In their CLIME-based application, inference on a target precision-matrix entry is constructed through a projected estimating equation. At the level of a single precision matrix, this construction is closely related to the precision–covariance residual-correction principle underlying Eq. (2.6), although the correction direction and normalization are constructed differently.*

For jointly regularized multi-graph estimators, a direct application of their construction would still yield graph-specific corrections, since the projection direction and normalization

2.3 Debiasing Beyond the Unified Penalized Likelihood Framework

for each graph depend on its own sample covariance matrix. Thus, cross-graph information would enter through the initial estimators rather than the correction step itself, and explicitly incorporating the joint structure into the latter would require further methodological and theoretical development.

The next proposition 2.1 gives an exact error decomposition for Eq. (2.6), isolating the leading fluctuation term and collecting the remainder terms involving the initial estimation error.

Proposition 2.1 (Error Decomposition). *Fix m and suppress the superscript (m) . Let $\Theta_0 \succ 0$, $\Sigma_0 = \Theta_0^{-1}$, $\Delta = \hat{\Theta} - \Theta_0$, and $V = \hat{S} - \Sigma_0$. Then the estimator in Eq. (2.6) satisfies the identity*

$$\tilde{\Theta} - \Theta_0 = -\Theta_0 V \Theta_0 - \left(\Delta \Sigma_0 \Delta + \Delta V \Theta_0 + \Theta_0 V \Delta + \Delta V \Delta \right). \quad (2.7)$$

The decomposition makes explicit that the leading term $-\Theta_0 V \Theta_0$ in Eq. (2.7) depends only on the fluctuation of $\hat{S} - \Sigma_0$ and does not depend on the specific estimation procedure used to obtain $\hat{\Theta}$. This provides a common representation for analyzing the debiasing step across a broad class of initial estimators. Under suitable regularity conditions, the remaining terms are asymptotically negligible and the leading term is entry-wise asymptotically normal.

Remark 3 (Choices of $\hat{S}^{(m)}$). *The matrix $\hat{S}^{(m)}$ represents a covariance-type estimator of $\Sigma_0^{(m)} = (\Theta_0^{(m)})^{-1}$ computed from the sample data in group m . In the Gaussian (or more generally sub-Gaussian) setting, a natural choice is $\hat{S}^{(m)} = \hat{\Sigma}^{(m)}$, which recovers exactly Eq. (2.4). As a possible extension beyond Gaussianity, one may replace $\hat{\Sigma}^{(m)}$ by a appropriate covariance-type estimator; we leave a detailed treatment to future work.*

3. Theoretical Results

In this section, we study the theoretical properties of the debiased estimator developed in Section 2 under $\widehat{S}^{(m)} = \widehat{\Sigma}^{(m)}$. We establish estimation consistency and entry-wise asymptotic normality, and then develop simultaneous inference procedures with valid control of the FWER and the FDR in high-dimensional settings. Rather than specifying a particular construction of the initial estimator, we formulate the theoretical results under a high-level consistency condition on the initial estimator $\widehat{\Theta} = (\widehat{\Theta}^{(1)}, \dots, \widehat{\Theta}^{(M)})$. This formulation can accommodate a broad class of multi-graph estimators, including JGL (Guo et al., 2011), FGL/GGL (Danaher et al., 2014), and related extensions, provided that the corresponding initial estimators satisfy the stated consistency condition.

We begin by introducing the parameter space, followed by the consistency requirement on the initial estimator, which provides the basis for the subsequent theoretical analysis. We consider a sparse matrix class as follows,

$$\mathcal{U}(C, p, s) := \left\{ \Theta \in \mathbb{S}_{++}^p : \|\Theta\|_1 \leq C, \max_{1 \leq t \leq p} \sum_{i=1}^p I(\Theta_{t,i} \neq 0) \leq s \right\}, \quad (3.8)$$

where $I(\cdot)$ is an indicator function.

Assumption 3.1 (Sparse Precision Matrices). *For some constants $C > 0$ and $s \geq 1$, the true precision matrices $\Theta_* = (\Theta_*^{(1)}, \dots, \Theta_*^{(M)})$ belong to the sparse matrix class in Eq. (3.8), i.e., $\Theta_*^{(m)} \in \mathcal{U}(C, p, s)$ for $m = 1, \dots, M$.*

Assumption 3.1 characterizes the structural sparsity of the true precision matrices. Such sparsity conditions are commonly adopted in the literature on high-dimensional sparse matrix

estimation (Cai et al., 2011; Zheng et al., 2025). We next state the remaining assumptions and main theoretical results.

Assumption 3.2 (Bounded Condition). *There exists a constant $c_1 > 0$ such that*

$$c_1^{-1} \leq \Lambda_{\min}(\Theta_*^{(m)}) \leq \Lambda_{\max}(\Theta_*^{(m)}) \leq c_1, \quad m = 1, \dots, M.$$

Assumption 3.3 (Dimensionality Condition). *Assume that the dimensionality p satisfies $\log p = o(n)$.*

Assumption 3.4 (Error of the Initial Estimator). *Let $\hat{\Theta} = (\hat{\Theta}^{(1)}, \dots, \hat{\Theta}^{(M)})$ be an initial estimator of the true precision matrices $\Theta_* = (\Theta_*^{(1)}, \dots, \Theta_*^{(M)})$. We assume that the initial estimator satisfies the following ℓ_1 -error bound:*

$$\max_{1 \leq m \leq M} \left\| \hat{\Theta}^{(m)} - \Theta_*^{(m)} \right\|_1 = O_p(\delta_{n,1}),$$

where the required rate of $\delta_{n,1}$ depends on the theoretical result under consideration:

- (a) $\delta_{n,1} \rightarrow 0$,
- (b) $\delta_{n,1} = o\left(\frac{1}{\sqrt{\log p}}\right)$ and $\delta_{n,1} = o(n^{-1/4})$.

Assumptions 3.2 - 3.3 are standard in high-dimensional precision matrix estimation; see, e.g., Bühlmann and van de Geer (2011); Zheng et al. (2026). Assumption 3.2 ensures that the target precision matrices are uniformly well-conditioned, while Assumption 3.3 controls the growth of the dimensionality relative to the sample size. Assumption 3.4 quantifies the accuracy of the initial estimator through its ℓ_1 error rate and distinguishes the requirements

for consistency and inference. Assumption 3.4(a) suffices for consistency results, whereas Assumption 3.4(b) is imposed to control higher-order remainder terms in the debiasing step and to establish entry-wise asymptotic normality and the validity of subsequent inferential procedures.

Remark 4. If $\widehat{\Sigma}^{(m)}$ is replaced by a covariance-type estimator $\widehat{S}^{(m)}$, analogous results can be established provided that $\|\widehat{S}^{(m)} - \Sigma_*^{(m)}\|_{\max} = O_p\left(\sqrt{\log p/n}\right)$ and the resulting leading term admits an entry-wise CLT. A detailed treatment of this extension is left for future work.

Before turning to the inferential properties, we establish an entry-wise consistency bound for the debiased estimator. This result plays a supporting role in the subsequent analysis and guarantees that the debiased estimator remains well-behaved in high dimensions.

Proposition 3.1 (Entry-wise consistency of the debiased estimator). *Suppose that Assumptions 3.1 - 3.3 and 3.4(a) hold. Then the debiased estimator $\widetilde{\Theta} = (\widetilde{\Theta}^{(1)}, \dots, \widetilde{\Theta}^{(M)})$ defined in Eq. (2.4) satisfies*

$$\max_{1 \leq m \leq M} \|\widetilde{\Theta}^{(m)} - \Theta_*^{(m)}\|_{\max} = O_p\left(\sqrt{\frac{\log p}{n}} + \delta_{n,1} \sqrt{\frac{\log p}{n}} + \delta_{n,1}^2\right).$$

In particular, if $\delta_{n,1} \rightarrow 0$ and $\log p = o(n)$, then $\max_{1 \leq m \leq M} \|\widetilde{\Theta}^{(m)} - \Theta_*^{(m)}\|_{\max} \rightarrow 0$ in probability.

The bound in Proposition 3.1 consists of three components: an empirical fluctuation term of order $\sqrt{\log p/n}$ due to covariance estimation, and two remainder terms that stem from the debiasing procedure. This proposition provides uniform entry-wise control of the debiased estimator under the stated conditions on the initial estimator and serves as a preparatory step for establishing asymptotic normality and valid inference.

Theorem 3.1 (Asymptotic normality of the debiased estimator). *Suppose that Assumptions 3.1 - 3.3 and 3.4(b) hold. Then the debiased estimator $\tilde{\Theta}$ defined in Eq. (2.4) admits the following properties.*

(i) **Asymptotic linear representation.** *For each $m = 1, \dots, M$,*

$$\tilde{\Theta}^{(m)} - \Theta_*^{(m)} = -\Theta_*^{(m)} W^{(m)} \Theta_*^{(m)} + R^{(m)},$$

where the remainder term satisfies $\sqrt{n} \|R^{(m)}\|_{\max} = o_p(1)$.

(ii) **Entry-wise asymptotic normality.** *For any fixed (j, l) and each $m = 1, \dots, M$,*

$$\sqrt{n}(\tilde{\Theta}_{jl}^{(m)} - \Theta_{*,jl}^{(m)}) \xrightarrow{d} \mathcal{N}(0, \tau_{m,jl}^2),$$

where $\tau_{m,jl}^2 = \text{Var}\left\{(\Theta_*^{(m)} \mathbf{x}^{(m)})_j (\Theta_*^{(m)} \mathbf{x}^{(m)})_l\right\}$. In particular, under the Gaussian model,

$$\tau_{m,jl}^2 = [\Theta_*^{(m)}]_{jj} [\Theta_*^{(m)}]_{ll} + ([\Theta_*^{(m)}]_{jl})^2.$$

For the subsequent inferential procedures, we specialize to the Gaussian model, under which the asymptotic variance in Theorem 3.1 admits the closed-form expression given above.

Proposition 3.2 (Consistency of variance estimation). *Under the Gaussian model, suppose that the conditions of Theorem 3.1 hold. Then, for any fixed $m \in \{1, \dots, M\}$ and (j, l) , the plug-in variance estimator $\hat{\tau}_{m,jl}^2 = [\tilde{\Theta}^{(m)}]_{jj} [\tilde{\Theta}^{(m)}]_{ll} + ([\tilde{\Theta}^{(m)}]_{jl})^2$ is consistent in the sense that*

$$\hat{\tau}_{m,jl}^2 \xrightarrow{p} \tau_{m,jl}^2.$$

Theorem 3.1 establishes entry-wise asymptotic normality of the debiased estimator, and Proposition 3.2 shows that the associated asymptotic variance can be consistently estimated

via a simple plug-in rule. Together, these results provide the basis for valid entry-wise inference in high-dimensional multi-graph models.

Theorem 3.2 (Entry-wise confidence intervals). *Under the Gaussian model, suppose that the conditions of Theorem 3.1 hold, and let $\alpha \in (0, 1)$ be a nominal level. For any fixed $m \in \{1, \dots, M\}$ and index pair (j, l) , define*

$$\text{CI}_{m,jl}(1 - \alpha) = \left[\tilde{\Theta}_{jl}^{(m)} \pm z_{1-\alpha/2} \frac{\hat{\tau}_{m,jl}}{\sqrt{n}} \right],$$

where $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ -quantile of $\mathcal{N}(0, 1)$ and $\hat{\tau}_{m,jl} = (\hat{\tau}_{m,jl}^2)^{1/2}$ with $\hat{\tau}_{m,jl}^2$ defined in Proposition 3.2. Then

$$\lim_{n \rightarrow \infty} \Pr \left\{ \Theta_{*,jl}^{(m)} \in \text{CI}_{m,jl}(1 - \alpha) \right\} = 1 - \alpha.$$

Corollary 3.1 (Entry-wise hypothesis testing). *Under the Gaussian model, suppose that the conditions of Theorem 3.1 hold. For any fixed $m \in \{1, \dots, M\}$ and (j, l) , consider testing $H_0 : \Theta_{*,jl}^{(m)} = 0$ against the two-sided alternative. Define*

$$T_{m,jl} = \sqrt{n} \frac{\tilde{\Theta}_{jl}^{(m)}}{\hat{\tau}_{m,jl}}.$$

Under H_0 , $T_{m,jl} \xrightarrow{d} \mathcal{N}(0, 1)$, and the level- α test rejects H_0 whenever $|T_{m,jl}| > z_{1-\alpha/2}$.

Theorem 3.2 provides entry-wise confidence intervals based on the debiased estimator with asymptotically correct coverage probability. As a direct consequence, Corollary 3.1 yields a corresponding hypothesis testing procedure.

Theorem 3.3 (Simultaneous inference and FWER control). *Under the Gaussian model, suppose that the conditions of Theorem 3.1 hold. In addition, suppose that $\frac{\{\log(pn)\}^7}{n} \rightarrow 0$ and $\delta_{n,1}^2 \sqrt{n \log(pn)} \rightarrow 0$. Let $\mathcal{S}$ be an index set of interest,*

$$\mathcal{S} \subset \{(m, j, l) : m = 1, \dots, M, 1 \leq j < l \leq p\}, \quad d_s := |\mathcal{S}|.$$

For each $(m, j, l) \in \mathcal{S}$, define the studentized statistic

$$T_{m,jl} = \sqrt{n} \frac{\tilde{\Theta}_{jl}^{(m)} - \Theta_{*,jl}^{(m)}}{\hat{\tau}_{m,jl}}, \quad T_{\max} = \max_{(m,j,l) \in \mathcal{S}} |T_{m,jl}|.$$

Let $c_{1-\alpha}$ be the $(1-\alpha)$ -quantile of $T_{\max}$ estimated via a Gaussian multiplier bootstrap. Define the simultaneous confidence intervals $\text{CI}_{m,jl}^{\text{sim}}(1-\alpha) = [\tilde{\Theta}_{jl}^{(m)} \pm c_{1-\alpha} \frac{\hat{\tau}_{m,jl}}{\sqrt{n}}]$, $(m, j, l) \in \mathcal{S}$.

Then the resulting simultaneous confidence region has asymptotic coverage probability $1-\alpha$:

$$\lim_{n \rightarrow \infty} \Pr \left\{ \Theta_{*,jl}^{(m)} \in \text{CI}_{m,jl}^{\text{sim}}(1-\alpha) \text{ for all } (m, j, l) \in \mathcal{S} \right\} = 1-\alpha.$$

As is standard in simultaneous inference, the critical value $c_{1-\alpha}$ increases with $d_s = |\mathcal{S}|$, reflecting the intrinsic cost of controlling the FWER over multiple entries. In particular, under weak dependence, $c_{1-\alpha}$ typically grows on the order of $\sqrt{\log d_s}$, which is unavoidable even under independence. The Gaussian multiplier bootstrap automatically calibrates this inflation by capturing the dependence structure of the debiased statistics (Chernozhukov et al., 2013).

When d_s is large, such control may become conservative and lead to a substantial loss of power. In large-scale inference problems, it is therefore common to adopt less stringent error

metrics that can yield higher power, such as the false discovery proportion (FDP) and the false discovery rate (FDR). In contrast to FWER, which controls the probability of making at least one false rejection, FDR controls the expected fraction of false rejections among all rejections.

For a given threshold $t > 0$, define the rejection set

$$\mathcal{R}(t) = \{(m, j, l) \in \mathcal{S} : |T_{m,jl}| > t\}, \quad R(t) := |\mathcal{R}(t)|.$$

Let $\mathcal{H}_0 \subset \mathcal{S}$ denote the set of true null hypotheses, and let $R_0(t) = |\mathcal{R}(t) \cap \mathcal{H}_0|$ be the number of false rejections at threshold t . The false discovery proportion and false discovery rate at threshold t are defined as

$$\text{FDP}(t) = \frac{R_0(t)}{R(t) \vee 1}, \quad \text{FDR}(t) = \mathbb{E}\{\text{FDP}(t)\}.$$

In Theorem 3.4, we develop an FDR control procedure based on the debiased statistics.

Theorem 3.4 (FDR control via the Benjamini–Yekutieli procedure). *Under the Gaussian model, suppose that the conditions of Theorem 3.1 hold. For each $(m, j, l) \in \mathcal{S}$, define the normal-approximation p-value $\hat{p}_{m,jl} = 2\{1 - \Phi(|T_{m,jl}|)\}$, where $\Phi(\cdot)$ denotes the standard normal distribution function and $T_{m,jl} = \sqrt{n} \frac{\tilde{\Theta}_{jl}^{(m)}}{\hat{\tau}_{m,jl}}$ is the studentized debiased statistic. Let $\hat{p}_{(1)} \leq \dots \leq \hat{p}_{(d_s)}$ denote the ordered p-values, and let (m_r, j_r, l_r) be the index triple corresponding to $\hat{p}_{(r)}$. Fix a target FDR level $q \in (0, 1)$ and define $c_{d_s} = \sum_{r=1}^{d_s} r^{-1}$. Define*

$$\hat{r}_{\text{BY}} = \max \left\{ 1 \leq r \leq d_s : \hat{p}_{(r)} \leq \frac{r}{d_s} \cdot \frac{q}{c_{d_s}} \right\},$$

with the convention that $\hat{r}_{\text{BY}} = 0$ if the set is empty, and reject all hypotheses corresponding to indices (m_r, j_r, l_r) for $1 \leq r \leq \hat{r}_{\text{BY}}$. Then the resulting multiple testing procedure achieves FDR control asymptotically in the sense that

$$\limsup_{n \rightarrow \infty} \text{FDR} \leq q.$$

To accommodate the complex and unknown dependence structure among the test statistics $\{T_{m,j,l}\}_{(m,j,l) \in \mathcal{S}}$, we adopt the Benjamini–Yekutieli (BY) procedure (Benjamini and Yekutieli, 2001), which introduces the harmonic factor $c_{d_s} = \sum_{r=1}^{d_s} r^{-1}$. As stated in Benjamini and Yekutieli (2001), although this correction is conservative, it provides FDR control without imposing explicit assumptions on the dependence structure of the test statistics.

4. Application to Differential Structure Detection

While the debiased estimators and their associated inferential theory facilitate valid entry-wise inference for individual graphs, many scientific questions of practical interest extend beyond single-graph analysis to focus on structural differences across multiple graphs.

In a wide range of applications—such as gene regulatory networks across disease subtypes, brain connectivity networks across experimental conditions, and financial networks across time periods—the primary aim is to identify edges whose strengths differ significantly across related populations. Such differences often encode meaningful structural changes and are of central interest for scientific interpretation. This motivates the problem of differential structure detection in high-dimensional multi-graph models.

For any fixed pair of indices (j, l) , an important inferential task is to test whether the

corresponding edge differs across graphs. A canonical example is the pairwise differential hypothesis $H_0 : \Theta_{*,jl}^{(m_1)} = \Theta_{*,jl}^{(m_2)}$, $m_1 \neq m_2$, or more generally, linear contrast hypotheses of the form $H_0 : \mathbf{c}^\top (\Theta_{*,jl}^{(1)}, \dots, \Theta_{*,jl}^{(M)}) = 0$, where $\mathbf{c} \in \mathbb{R}^M$ specifies a contrast of interest.

In this section, we leverage the previously established entry-wise asymptotic normality to develop valid inferential procedures for detecting differential network structures. Importantly, no additional estimation or debiasing step is required. Differential inference is achieved by applying linear transformations to the debiased estimators, allowing the previously established asymptotic results to be extended to linear contrasts. This leads to theoretically justified procedures for testing and confidence interval construction for edge-level differences, even in high-dimensional settings where the number of variables far exceeds the sample size. Theorem 4.1 formalizes the asymptotic distribution of general linear contrasts of the debiased estimators, providing the foundation for differential network inference in high dimensions.

Theorem 4.1 (Asymptotic normality for differential edge inference). *Suppose that the conditions of Theorem 3.1 hold. Let $\tilde{\Theta}_{jl} = (\tilde{\Theta}_{jl}^{(1)}, \dots, \tilde{\Theta}_{jl}^{(M)})^\top$ denote the vector of the debiased estimators corresponding to the $(j, l)^{\text{th}}$ entry across M graphs. For any fixed contrast vector $\mathbf{c} \in \mathbb{R}^M$ satisfying $\|\mathbf{c}\|_2 = 1$, define the linear contrast functional $\mathbf{c}^\top \Theta_{*,jl} = \mathbf{c}^\top (\Theta_{*,jl}^{(1)}, \dots, \Theta_{*,jl}^{(M)})$. Then the contrast-based debiased estimator $\mathbf{c}^\top \tilde{\Theta}_{jl}$ satisfies*

$$\sqrt{n} \left\{ \mathbf{c}^\top \tilde{\Theta}_{jl} - \mathbf{c}^\top \Theta_{*,jl} \right\} \xrightarrow{d} \mathcal{N} \left(0, \sigma_{jl}^2(\mathbf{c}) \right),$$

where $\sigma_{jl}^2(\mathbf{c}) = \mathbf{c}^\top \Sigma_{jl} \mathbf{c}$, $\Sigma_{jl} = \text{Cov} \left(v_{1,j}^{(1)} v_{1,l}^{(1)}, \dots, v_{1,j}^{(M)} v_{1,l}^{(M)} \right)$, with $\mathbf{v}_1^{(m)} = \Theta_*^{(m)} \mathbf{x}_1^{(m)}$.

Corollary 4.1 (Pairwise differential edge testing). *Under the Gaussian model, suppose that the conditions of Theorem 4.1, consider two graphs $m_1 \neq m_2$ and the null hypothesis $H_0 : \Theta_{*,jl}^{(m_1)} = \Theta_{*,jl}^{(m_2)}$. Define $\mathbf{c} = \frac{1}{\sqrt{2}}(0, \dots, 1, \dots, -1, \dots, 0)^\top$, where the nonzero entries correspond to m_1 and m_2 . Let*

$$T_{jl}^{(m_1, m_2)} = \sqrt{n} \frac{\tilde{\Theta}_{jl}^{(m_1)} - \tilde{\Theta}_{jl}^{(m_2)}}{\hat{\sigma}_{jl}(m_1, m_2)},$$

where $\hat{\sigma}_{jl}^2(m_1, m_2) = \hat{\tau}_{m_1, jl}^2 + \hat{\tau}_{m_2, jl}^2$ with $\hat{\tau}_{m, jl}^2$ defined in Proposition 3.2.

Then under H_0 , we have $T_{jl}^{(m_1, m_2)} \xrightarrow{d} \mathcal{N}(0, 1)$. Consequently, the test rejects H_0 at level α whenever $|T_{jl}^{(m_1, m_2)}| > z_{1-\alpha/2}$.

5. Simulation

In this section, we conduct a series of numerical experiments to systematically evaluate the performance of the debiased estimator and inference framework for high-dimensional multi-graph models with structural similarities.

Throughout the simulations, we fix the number of graphs $M = 5$, which represents a typical moderate-scale multi-graph setting. Additional experiments with other values of M lead to qualitatively similar conclusions and are omitted for brevity.

5.1 Simulation Setting

Suppose that the preliminary matrix for the m^{th} group is constructed as $\Theta_{\text{raw}}^{(m)} = \Theta_0 + \rho \Delta^{(m)}$, $m = 1, \dots, M$, where Θ_0 denotes a common precision structure shared across all groups, $\Delta^{(m)}$ represents the group-specific difference, and $\rho > 0$ controls the magnitude of heterogeneity

across groups.

To ensure positive definiteness, we further apply a diagonal loading step and define $\Theta^{(m)} = \Theta_{\text{raw}}^{(m)} + \sigma \mathbf{I}_p$, where $\sigma > 0$ is chosen such that $\lambda_{\min}(\Theta^{(m)}) \geq 0.1$. Given $\Theta^{(m)}$, observations from the m^{th} group are independently and identically distributed as $\mathbf{x}_k^{(m)} \sim \mathcal{N}(\mathbf{0}, \Sigma^{(m)})$, where $\Sigma^{(m)} = (\Theta^{(m)})^{-1}$ and $k = 1, \dots, n$.

Under the “common structure + group-specific difference” setting, we consider three model settings corresponding to different structures of the common precision matrix Θ_0 . In all settings, the difference components $\Delta^{(m)}$ are generated as random sparse matrices.

- **Model I (Random sparse common structure):** For each pair $1 \leq j < l \leq p$, the off-diagonal entry $\Theta_{0,jl}$ is nonzero with probability π_0 , and is generated as $\Theta_{0,jl} = s_{jl}w_{jl}$, where $s_{jl} \in \{-1, +1\}$ is a random sign and $w_{jl} \sim \text{Unif}(0.3, 0.6)$. Otherwise, $\Theta_{0,jl} = 0$. Let $\Theta_{0,jj} = 1$ for any $j = 1, \dots, p$, and $\Theta_{0,lj} = \Theta_{0,jl}$.
- **Model II (Block-structured common structure):** The p variables are partitioned into 10 disjoint blocks of equal size according to their indices. For variable pairs within the same block, off-diagonal entries are generated in the same manner as in Model I with probability π_{in} . For variable pairs belonging to different blocks, the corresponding entries are set to 0.
- **Model III (Banded common structure):** The common precision matrix Θ_0 is constructed to have a banded structure with bandwidth 3. Specifically, for $1 \leq j < l \leq p$, the off-diagonal entry $\Theta_{0,jl}$ is nonzero if and only if $1 \leq l - j \leq 3$, in which case it is generated as $\Theta_{0,jl} = s_{jl}w_{jl}$, where s_{jl} and w_{jl} are defined as in Model I. For all other pairs (j, l) with $|l - j| > 3$, we set $\Theta_{0,jl} = 0$. Let $\Theta_{0,jj} = 1$ for any $j = 1, \dots, p$,

and $\Theta_{0,lj} = \Theta_{0,jl}$.

For all model settings, the group-specific components $\Delta^{(m)}$ are generated as random sparse symmetric matrices. For each pair $1 \leq j < l \leq p$, the entry $\Delta_{jl}^{(m)}$ is nonzero with probability π_Δ and is generated as $\Delta_{jl}^{(m)} = \tilde{s}_{jl}^{(m)} \tilde{w}_{jl}^{(m)}$, where $\tilde{s}_{jl}^{(m)} \in \{-1, +1\}$ and $\tilde{w}_{jl}^{(m)} \sim \text{Unif}(0.2, 0.4)$. Otherwise, $\Delta_{jl}^{(m)} = 0$. The diagonal entries of $\Delta^{(m)}$ are set to zero, and $\Delta_{lj}^{(m)} = \Delta_{jl}^{(m)}$. Throughout the simulations, we set $\pi_0 = 0.05$, $\pi_{\text{in}} = 0.2$, $\pi_\Delta = 0.05$, and $\rho = 1$.

5.2 Initial Estimators and Inference Procedures

The proposed debiased inference framework does not rely on a specific choice of the initial multi-graph estimator. Instead, it provides a general debiased framework for inference which applies to a broad class of existing multi-graph estimators satisfying the ℓ_1 -consistency condition in Assumption 3.4. To illustrate its general applicability, we consider several representative initial estimators commonly used in the literature.

Importantly, the focus of our numerical experiments is not on comparing the estimation accuracy of different initial estimators. Instead, we treat these estimators solely as inputs to the inference stage and examine whether the debiased procedure can correct the regularization-induced bias and deliver valid inference across a diverse range of initial estimation strategies.

Six representative methods are used to evaluate the debiased inference framework, including three separate estimators (Glasso (Friedman et al., 2008), Clime (Cai et al., 2011) and Node-wise approach (Meinshausen and Bühlmaan, 2006)) and three joint estimators

5.3 Evaluation Metrics and Experimental Design

(FGL and GGL (Danaher et al., 2014), and JSEM (Ma and Michailidis, 2016)). In practice, Glasso, CLIME, and Node-wise approach are implemented using the R packages ‘*glasso*’, ‘*fastclime*’, and ‘*huge*’, respectively. The joint estimators FGL/GGL are implemented via the ‘*JGL*’ R package, while JSEM is implemented using the code provided by the authors on their GitHub repository.

Based on each initial estimator, we consider two inference strategies with different inferential scopes. The first strategy is a naive plug-in inference approach, which is only used for entry-wise inference and directly treats the regularized estimator as unbiased and constructs test statistics and confidence intervals based on its empirical variance. Although widely used in practice, such plug-in procedures are known to suffer from substantial bias in high-dimensional settings due to regularization. The second strategy is the proposed debiased inference method. Starting from the same initial estimator, we apply the debiasing step to explicitly remove the leading regularization-induced bias. The resulting debiased estimators provide the basis for valid entry-wise inference as well as simultaneous inference procedures, including FWER and FDR control.

5.3 Evaluation Metrics and Experimental Design

The performance of the proposed debiased inference framework is evaluated from an inferential perspective rather than an estimation accuracy perspective. Specifically, we focus on three complementary aspects: entry-wise distributional approximation, confidence interval coverage, and multiple testing error control.

First, to assess the finite-sample Gaussian approximation, we construct standardized test statistics based on both the naive plug-in estimator and the debiased estimator. The

empirical distributions of these statistics are compared with the standard normal distribution using histograms and Q–Q plots, highlighting the effect of debiasing on distributional accuracy. Second, we evaluate the empirical coverage probabilities of nominal 95% confidence intervals constructed using both inference procedures. For a fixed entry (m, j, l) , the coverage probability is computed as the proportion of simulation replicates in which the true parameter $\Theta_{jl}^{(m)}$ falls within the corresponding confidence interval. Third, we investigate the performance of the proposed method in large-scale hypothesis testing. Specifically, we examine the empirical FWER and FDR under the simultaneous inference and multiple testing procedures developed in Section 3. The empirical power is also reported to illustrate the trade-off between error control and detection ability.

In all experiments, we consider a small number of representative (p, n) configurations covering the regimes $p < n$, $p = n$, and $p > n$. We set $(p, n) \in \{(50, 200), (100, 200), (200, 200), (400, 200)\}$ unless otherwise stated. Although equal sample sizes are used across groups for convenience, the proposed procedure does not require this restriction, as debiasing and subsequent within-group inference use each group’s corresponding sample covariance matrix and sample size. Each simulation setting is repeated independently for $R = 200$ replications to ensure stable empirical results.

5.4 Simulation Results

We first examine the finite-sample distributional behavior of the debiased test statistics. Figures 1 - 4 report the empirical histograms and Q–Q plots of entry-wise debiased statistics under Model I with $p = 100$, separately for signal edges $\left(\Theta_{*,jl}^{(m)} \neq 0\right)$ and null edges $\left(\Theta_{*,jl}^{(m)} = 0\right)$. Results for other dimensional regimes and structural settings are deferred to

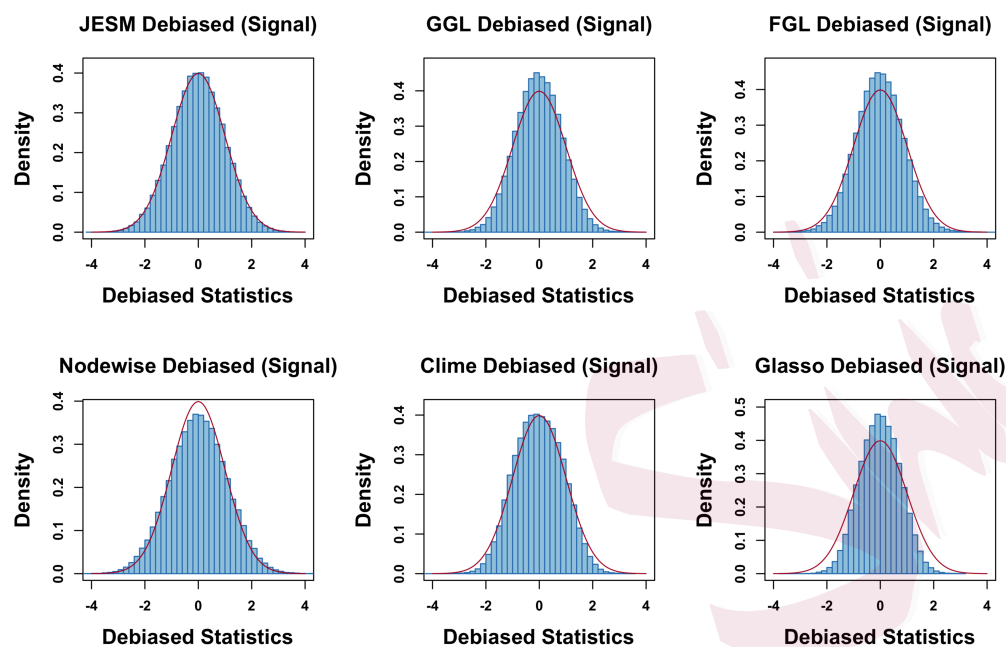


Figure 1: Empirical distributions of debiased statistics for signal edges.

the Supplementary Material.

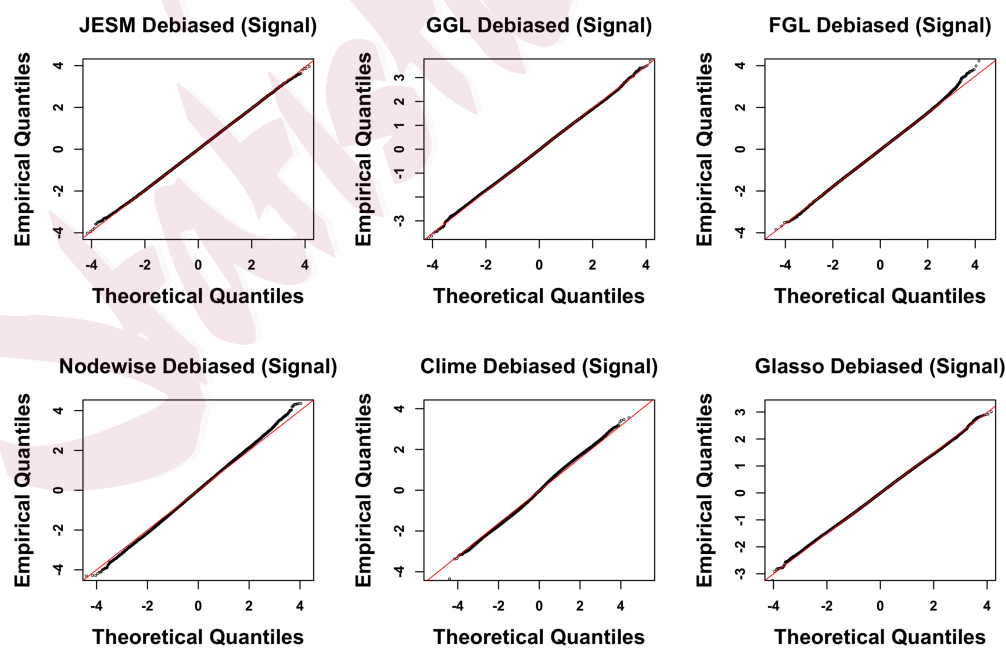


Figure 2: Quantile-quantile plots of debiased statistics for signal edges.

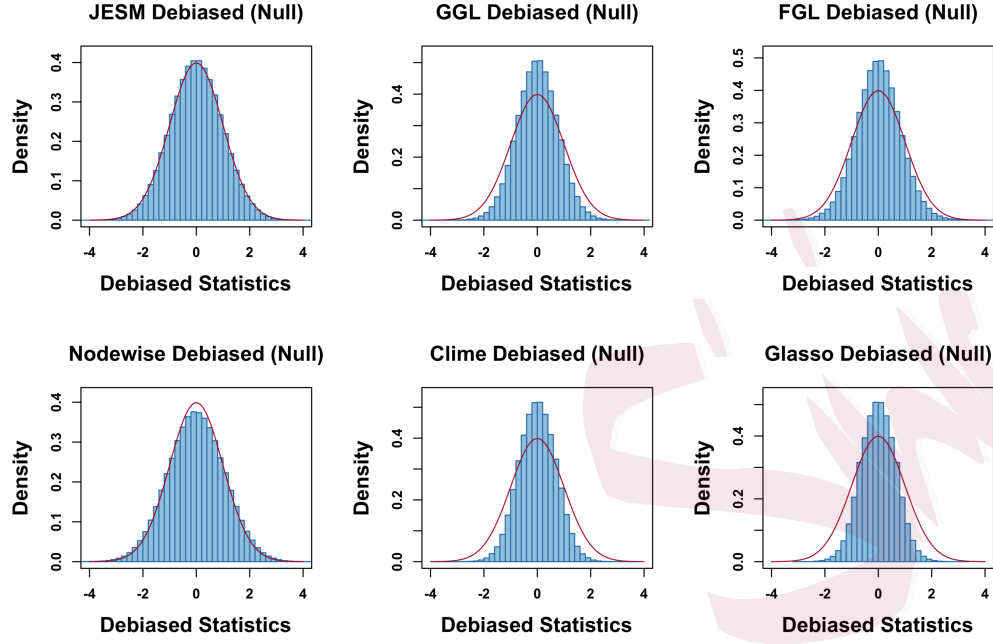


Figure 3: Empirical distributions of debiased statistics for null edges.

Across both signal and null edges, the empirical histograms of debiased statistics closely match the $\mathcal{N}(0, 1)$ density for all initial estimators. The corresponding Q–Q plots exhibit near-linear alignment with the theoretical normal quantiles.

To further quantify inferential accuracy, we examine the empirical coverage probabilities of nominal 95% confidence intervals constructed using the debiased estimators. Figure 5 summarizes the coverage performance across Models I – III and dimensions $p \in \{50, 100, 200, 400\}$, separately for signal and null edges.

Across all settings, the confidence intervals constructed from the debiased estimators achieve coverage probability close to the nominal level for both signal and null edges. With the exception of Clime, the coverage rates for signal and null edges are essentially equal across methods. A plausible explanation is that Clime is built on constraints calibrated to the matrix $\ell_{\max}$ norm, which tends to emphasize worst-case entry-wise control; in finite

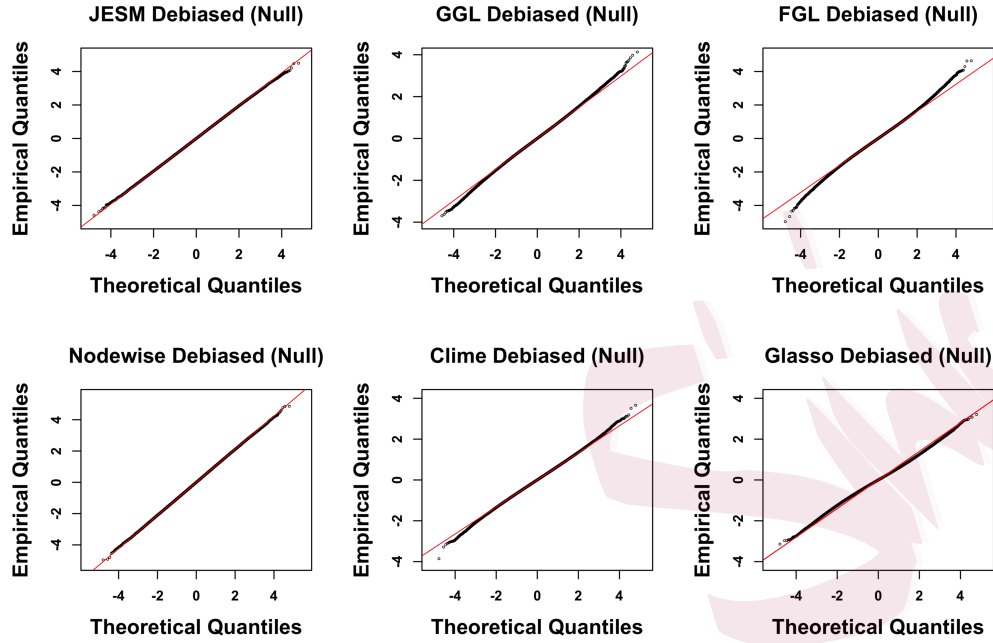


Figure 4: Quantile-quantile plots of debiased statistics for null edges.

samples, this may exacerbate misestimation for relatively weak signal edges, leading to a noticeable drop in signal-edge coverage.

To complement the coverage diagnostics in Figure 5, Table 1 reports the empirical FWER and FDR, together with the corresponding power in parentheses, across Models I–III, all considered initial estimators, and $p \in \{50, 100, 200, 400\}$. Both empirical error rates remain below their respective nominal levels across the reported settings, indicating conservative error control. The power results further illustrate the trade-off between error control and detection ability. For both procedures, power increases as the nominal level increases from 0.05 to 0.20. With the sample size fixed, power decreases as the dimension increases. These patterns are observed across all three models and the six initial estimators.

Taken together, across both signal and null edges, the debiased statistics exhibit a well-behaved Gaussian approximation, the resulting confidence intervals achieve coverage

5.4 Simulation Results

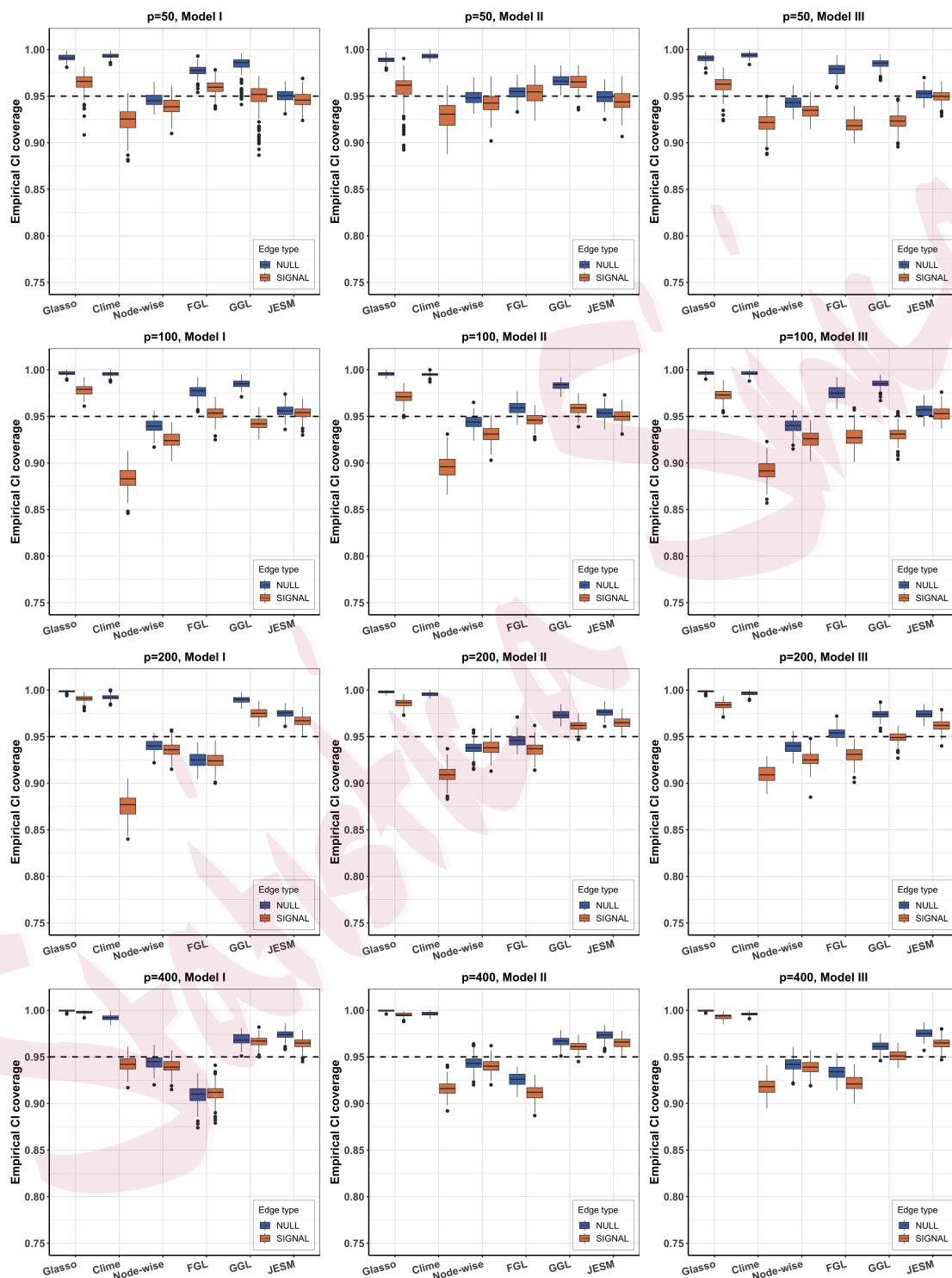


Figure 5: Empirical entry-wise confidence interval coverage of debiased estimators under three data-generating models (Models I – III) and four dimensions $p \in \{50, 100, 200, 400\}$. (For each panel, boxplots summarize the distribution of coverage probabilities across 200 Monte Carlo replications at the nominal confidence level 0.95. Coverage is evaluated separately for NULL and SIGNAL edges. The dashed horizontal line indicates the nominal coverage level.)

5.4 Simulation Results

close to the nominal level, and the FDR/FWER calibration remains stable across settings. These results collectively indicate that the proposed debiasing framework delivers reliable distributional approximation and valid inference over a range of initial estimators and high-dimensional regimes.

Importantly, similar performance is observed across a range of initial estimators, including both single-graph and jointly regularized multi-graph methods. These findings support the applicability of the proposed debiasing framework across different initial estimation methods, provided that the required consistency conditions are satisfied.

Model Method		FWER (power), nominal level α												FDR (power), target level q											
		$p = 50$			$p = 100$			$p = 200$			$p = 400$			$p = 50$			$p = 100$			$p = 200$			$p = 400$		
		0.05	0.10	0.20	0.05	0.10	0.20	0.05	0.10	0.20	0.05	0.10	0.20	0.05	0.10	0.20	0.05	0.10	0.20	0.05	0.10	0.20	0.05	0.10	0.20
I	FGL	0.035 (0.811)	0.070 (0.838)	0.145 (0.869)	0.045 (0.677)	0.085 (0.707)	0.175 (0.736)	0.035 (0.494)	0.065 (0.520)	0.140 (0.553)	0.035 (0.283)	0.065 (0.312)	0.140 (0.341)	0.040 (0.836)	0.074 (0.870)	0.156 (0.901)	0.032 (0.697)	0.066 (0.732)	0.133 (0.762)	0.039 (0.524)	0.079 (0.559)	0.164 (0.588)	0.032 (0.304)	0.066 (0.335)	0.125 (0.370)
	GGL	0.040 (0.786)	0.085 (0.814)	0.165 (0.846)	0.045 (0.644)	0.080 (0.671)	0.170 (0.706)	0.045 (0.468)	0.090 (0.497)	0.170 (0.527)	0.035 (0.256)	0.070 (0.285)	0.130 (0.316)	0.036 (0.811)	0.074 (0.848)	0.140 (0.878)	0.042 (0.671)	0.087 (0.709)	0.168 (0.739)	0.042 (0.495)	0.081 (0.530)	0.168 (0.562)	0.041 (0.283)	0.077 (0.318)	0.161 (0.347)
	Clime	0.035 (0.818)	0.075 (0.846)	0.145 (0.879)	0.040 (0.675)	0.080 (0.703)	0.170 (0.732)	0.040 (0.498)	0.080 (0.523)	0.150 (0.557)	0.040 (0.288)	0.085 (0.318)	0.165 (0.350)	0.032 (0.848)	0.066 (0.882)	0.136 (0.913)	0.037 (0.700)	0.076 (0.736)	0.155 (0.766)	0.039 (0.516)	0.078 (0.552)	0.153 (0.581)	0.041 (0.306)	0.079 (0.343)	0.172 (0.375)
	JSEM	0.040 (0.811)	0.085 (0.840)	0.175 (0.867)	0.035 (0.665)	0.065 (0.694)	0.135 (0.722)	0.045 (0.486)	0.085 (0.515)	0.170 (0.546)	0.035 (0.283)	0.075 (0.309)	0.155 (0.343)	0.033 (0.828)	0.067 (0.863)	0.139 (0.897)	0.042 (0.686)	0.082 (0.717)	0.172 (0.750)	0.032 (0.512)	0.061 (0.549)	0.119 (0.577)	0.043 (0.305)	0.083 (0.340)	0.178 (0.371)
	Glasso	0.035 (0.797)	0.070 (0.829)	0.140 (0.859)	0.035 (0.663)	0.075 (0.689)	0.160 (0.723)	0.040 (0.477)	0.080 (0.506)	0.150 (0.537)	0.035 (0.276)	0.075 (0.303)	0.135 (0.333)	0.032 (0.823)	0.068 (0.857)	0.138 (0.888)	0.034 (0.689)	0.067 (0.724)	0.132 (0.756)	0.040 (0.502)	0.076 (0.536)	0.159 (0.567)	0.035 (0.299)	0.073 (0.336)	0.148 (0.365)
	Nodewise	0.030 (0.804)	0.065 (0.831)	0.130 (0.861)	0.045 (0.655)	0.085 (0.684)	0.170 (0.715)	0.040 (0.479)	0.080 (0.507)	0.155 (0.537)	0.035 (0.272)	0.075 (0.300)	0.135 (0.331)	0.032 (0.829)	0.068 (0.861)	0.138 (0.893)	0.034 (0.682)	0.067 (0.714)	0.132 (0.745)	0.040 (0.498)	0.076 (0.533)	0.159 (0.565)	0.035 (0.298)	0.073 (0.333)	0.148 (0.365)
II	FGL	0.040 (0.805)	0.085 (0.831)	0.155 (0.864)	0.040 (0.662)	0.080 (0.689)	0.155 (0.721)	0.040 (0.482)	0.075 (0.512)	0.135 (0.543)	0.040 (0.266)	0.075 (0.293)	0.150 (0.322)	0.045 (0.829)	0.089 (0.864)	0.178 (0.898)	0.042 (0.687)	0.086 (0.721)	0.163 (0.753)	0.032 (0.504)	0.066 (0.536)	0.126 (0.567)	0.042 (0.284)	0.087 (0.319)	0.174 (0.349)
	GGL	0.040 (0.815)	0.075 (0.841)	0.165 (0.873)	0.045 (0.675)	0.085 (0.702)	0.180 (0.734)	0.035 (0.495)	0.065 (0.523)	0.135 (0.552)	0.040 (0.293)	0.085 (0.320)	0.170 (0.348)	0.041 (0.841)	0.082 (0.877)	0.155 (0.910)	0.035 (0.704)	0.066 (0.740)	0.140 (0.773)	0.043 (0.519)	0.081 (0.553)	0.173 (0.585)	0.035 (0.316)	0.071 (0.350)	0.147 (0.384)
	Clime	0.040 (0.797)	0.080 (0.823)	0.155 (0.855)	0.035 (0.660)	0.065 (0.688)	0.140 (0.716)	0.035 (0.486)	0.070 (0.512)	0.150 (0.542)	0.045 (0.267)	0.090 (0.295)	0.180 (0.323)	0.035 (0.826)	0.065 (0.858)	0.133 (0.892)	0.034 (0.678)	0.069 (0.712)	0.135 (0.744)	0.041 (0.502)	0.082 (0.538)	0.153 (0.568)	0.038 (0.287)	0.074 (0.320)	0.145 (0.352)
	JSEM	0.035 (0.802)	0.065 (0.833)	0.135 (0.864)	0.040 (0.656)	0.080 (0.687)	0.150 (0.716)	0.035 (0.480)	0.065 (0.504)	0.130 (0.537)	0.040 (0.271)	0.080 (0.297)	0.160 (0.331)	0.033 (0.826)	0.063 (0.862)	0.128 (0.893)	0.042 (0.686)	0.085 (0.717)	0.164 (0.750)	0.036 (0.503)	0.072 (0.535)	0.139 (0.570)	0.039 (0.297)	0.073 (0.329)	0.149 (0.362)
	Glasso	0.035 (0.821)	0.070 (0.851)	0.145 (0.881)	0.045 (0.681)	0.085 (0.707)	0.175 (0.741)	0.045 (0.493)	0.080 (0.522)	0.175 (0.553)	0.035 (0.289)	0.065 (0.318)	0.125 (0.348)	0.036 (0.842)	0.068 (0.874)	0.135 (0.907)	0.039 (0.708)	0.079 (0.742)	0.169 (0.776)	0.037 (0.523)	0.073 (0.560)	0.149 (0.591)	0.038 (0.317)	0.074 (0.349)	0.163 (0.383)
	Nodewise	0.035 (0.826)	0.070 (0.856)	0.145 (0.884)	0.045 (0.683)	0.085 (0.715)	0.175 (0.742)	0.045 (0.509)	0.080 (0.537)	0.175 (0.566)	0.035 (0.292)	0.065 (0.321)	0.125 (0.351)	0.036 (0.844)	0.068 (0.881)	0.135 (0.910)	0.039 (0.707)	0.079 (0.742)	0.169 (0.773)	0.037 (0.528)	0.073 (0.562)	0.149 (0.595)	0.038 (0.312)	0.074 (0.348)	0.163 (0.379)
III	FGL	0.035 (0.816)	0.080 (0.845)	0.155 (0.877)	0.045 (0.679)	0.090 (0.707)	0.175 (0.735)	0.035 (0.497)	0.070 (0.526)	0.140 (0.557)	0.035 (0.279)	0.065 (0.307)	0.135 (0.337)	0.034 (0.838)	0.064 (0.873)	0.128 (0.905)	0.039 (0.705)	0.083 (0.742)	0.155 (0.772)	0.039 (0.519)	0.073 (0.554)	0.156 (0.585)	0.042 (0.304)	0.080 (0.337)	0.165 (0.370)
	GGL	0.040 (0.812)	0.080 (0.840)	0.160 (0.873)	0.040 (0.672)	0.075 (0.700)	0.145 (0.730)	0.045 (0.490)	0.085 (0.519)	0.180 (0.552)	0.045 (0.288)	0.085 (0.318)	0.170 (0.346)	0.036 (0.843)	0.075 (0.876)	0.159 (0.911)	0.038 (0.691)	0.078 (0.726)	0.145 (0.757)	0.038 (0.518)	0.082 (0.551)	0.150 (0.583)	0.034 (0.310)	0.062 (0.343)	0.133 (0.375)
	Clime	0.040 (0.819)	0.075 (0.847)	0.155 (0.876)	0.035 (0.679)	0.075 (0.708)	0.135 (0.737)	0.035 (0.492)	0.070 (0.518)	0.145 (0.549)	0.040 (0.289)	0.080 (0.317)	0.155 (0.348)	0.033 (0.847)	0.070 (0.880)	0.126 (0.911)	0.040 (0.704)	0.084 (0.737)	0.161 (0.769)	0.031 (0.513)	0.059 (0.544)	0.125 (0.576)	0.031 (0.308)	0.069 (0.344)	0.139 (0.373)
	JSEM	0.035 (0.830)	0.070 (0.857)	0.135 (0.888)	0.040 (0.693)	0.085 (0.723)	0.175 (0.751)	0.040 (0.514)	0.080 (0.543)	0.155 (0.573)	0.035 (0.293)	0.065 (0.323)	0.140 (0.353)	0.033 (0.858)	0.072 (0.890)	0.134 (0.923)	0.033 (0.714)	0.070 (0.746)	0.133 (0.779)	0.042 (0.541)	0.077 (0.572)	0.162 (0.607)	0.037 (0.324)	0.076 (0.358)	0.148 (0.388)
	Glasso	0.040 (0.839)	0.085 (0.868)	0.180 (0.899)	0.040 (0.698)	0.080 (0.725)	0.150 (0.754)	0.035 (0.511)	0.065 (0.542)	0.135 (0.572)	0.035 (0.310)	0.070 (0.339)	0.150 (0.369)	0.033 (0.859)	0.067 (0.894)	0.130 (0.924)	0.036 (0.719)	0.074 (0.754)	0.140 (0.786)	0.032 (0.539)	0.061 (0.573)	0.129 (0.606)	0.032 (0.329)	0.062 (0.363)	0.124 (0.397)
	Nodewise	0.040 (0.816)	0.070 (0.842)	0.145 (0.873)	0.040 (0.669)	0.080 (0.697)	0.155 (0.727)	0.040 (0.495)	0.080 (0.525)	0.155 (0.556)	0.035 (0.279)	0.070 (0.310)	0.145 (0.340)	0.038 (0.837)	0.074 (0.870)	0.153 (0.900)	0.041 (0.695)	0.082 (0.726)	0.174 (0.762)	0.037 (0.517)	0.075 (0.554)	0.144 (0.585)	0.038 (0.297)	0.079 (0.331)	0.151 (0.364)

Table 1: Empirical FWER and FDR with corresponding power in parentheses, based on debiased statistics across Models I–III and $p \in \{50, 100, 200, 400\}$. The three columns within each dimension correspond to nominal levels $\alpha = 0.05, 0.10, 0.20$ for FWER and target levels $q = 0.05, 0.10, 0.20$ for FDR.

6. Real Data Analysis: WebKB Text Networks

We illustrate the proposed debiased inference framework on the publicly available WebKB corpus, which consists of webpages collected from computer science departments at Cornell University, the University of Wisconsin, the University of Texas, and the University of Washington. The data is available at <https://www.cs.cmu.edu/~webkb/>. Each webpage is manually labeled into one of four semantic categories, *student*, *faculty*, *course*, and *project*, reflecting its primary functional role in the academic environment. These four categories naturally form four related populations: the common institutional context may give rise to shared dependency structure among terms, while distinct communication purposes across page types can lead to category-specific deviations in conditional dependence structure.

6.1 Preprocessing Procedure

The corpus is preprocessed following standard text-mining practice. We remove tokens with fewer than three characters, remove stopwords, and apply the Porter stemming algorithm to the remaining words (Porter, 1980; Schütze et al., 2008). After stemming, we build a document-term frequency table. Let $f_{ij}^{(k)}$ denote the count of term j in webpage i from category k , where $k \in \{\text{stu}, \text{fac}, \text{course}, \text{proj}\}$, $i = 1, \dots, n_k$ and $N = \sum_k n_k$ is the total number of webpages in the pooled corpus. For the pooled corpus, index all webpages by $r = 1, \dots, N$, and let f_{rj} denote the count of term j in the webpage r . For each term j , define the pooled entropy weight

$$e_j = 1 + \frac{\sum_{r=1}^N p_{rj} \log p_{rj}}{\log N}, \quad p_{rj} = \frac{f_{rj}}{\sum_{r=1}^N f_{rj}},$$

6.2 Modeling and Edge Inference

with the convention that $0 \log 0 = 0$. For webpage i in category k , define the weighted feature

$$x_{ij}^{(k)} = e_j \log \left(1 + f_{ij}^{(k)} \right).$$

To control dimensionality, we rank terms by their pooled log-entropy weights $\{e_j\}$ and retain the top $p = 50$ terms for the main analysis. For each category k , we form the design matrix $\mathbf{X}^{(k)} \in \mathbb{R}^{n_k \times p}$ and standardize each column to have mean zero and unit variance within the category. To ensure that the standardization is well-defined, any candidate term with zero within-category variance is excluded and replaced by the next-ranked term, so that all retained features can be properly standardized in every category. The resulting sample sizes are $n_{\text{stu}} = 1625$, $n_{\text{fac}} = 1116$, $n_{\text{course}} = 926$, $n_{\text{proj}} = 501$, yielding a moderate multi-population setting with $p = 50$ and heterogeneous sample sizes across categories. We also conduct an additional analysis with $p = 100$, following the same term-selection and preprocessing procedures while keeping the group-specific sample sizes unchanged.

6.2 Modeling and Edge Inference

For each category $k \in \{\text{stu}, \text{fac}, \text{course}, \text{proj}\}$, we model the standardized term vector by a Gaussian graphical model with precision matrix $\Theta^{(k)} \in \mathbb{R}^{p \times p}$. A nonzero off-diagonal entry $\Theta_{jl}^{(k)}$ indicates conditional dependence between the j^{th} and l^{th} terms given all remaining terms.

Our goal is to quantify how the debiasing step refines an initial sparse, jointly regularized estimator. We proceed in two stages as follows,

1. **Initial jointly regularized estimator.** We compute an initial estimator of $\Theta =$

6.3 Summary and Visualization of Debiased Edge Discoveries

$(\Theta^{(\text{stu})}, \Theta^{(\text{fac})}, \Theta^{(\text{course})}, \Theta^{(\text{proj})})$ using a jointly regularized method (e.g., JGL), which estimates the precision matrices jointly under cross-group similarity regularization. Denote the resulting estimator by $\hat{\Theta}_{\text{init}}$, and define the corresponding initial edge set

$$E_{\text{init}}^{(k)} = \{(j, l) : \hat{\Theta}_{\text{init},jl}^{(k)} \neq 0, 1 \leq j < l \leq p\}.$$

2. **Debiasing and edge inference.** Starting from the initial jointly regularized estimator, we apply the proposed debiasing step to construct edge-wise debiased test statistics and p -values $\{p_{jl}^{(k)}\}_{1 \leq j < l \leq p}$ for all candidate edges in each category k . We then apply the BY procedure at level $q = 0.1$, which controls the FDR within each category, and obtain the corresponding set of significant edges

$$E_{\text{sig}}^{(k)} = \{(j, l) : p_{jl}^{(k)} \text{ is significant under the BY procedure at level } q = 0.1, 1 \leq j < l \leq p\}.$$

6.3 Summary and Visualization of Debiased Edge Discoveries

To summarize the refinement induced by debiasing and within-category BY adjustment, we report the overlaps and discrepancies between $E_{\text{init}}^{(k)}$ and $E_{\text{sig}}^{(k)}$ for $p = 50$ in Table 2. We define

$$E_{\text{kept}}^{(k)} = E_{\text{init}}^{(k)} \cap E_{\text{sig}}^{(k)}, \quad E_{\text{lost}}^{(k)} = E_{\text{init}}^{(k)} \setminus E_{\text{sig}}^{(k)}, \quad E_{\text{gained}}^{(k)} = E_{\text{sig}}^{(k)} \setminus E_{\text{init}}^{(k)}.$$

For $p = 50$, Table 2 shows that, on the WebKB data, the debiasing procedure combined with within-category FDR control produces a substantial refinement of the initial jointly regularized estimator. Across all categories, the number of debiased significant edges $|E_{\text{sig}}^{(k)}|$

6.3 Summary and Visualization of Debiased Edge Discoveries

is markedly smaller than the number of nonzero edges from the initial estimator $|E_{\text{init}}^{(k)}|$ (e.g., Student: $165 \rightarrow 88$; Project: $110 \rightarrow 48$). Moreover, only a fraction of the initial edges

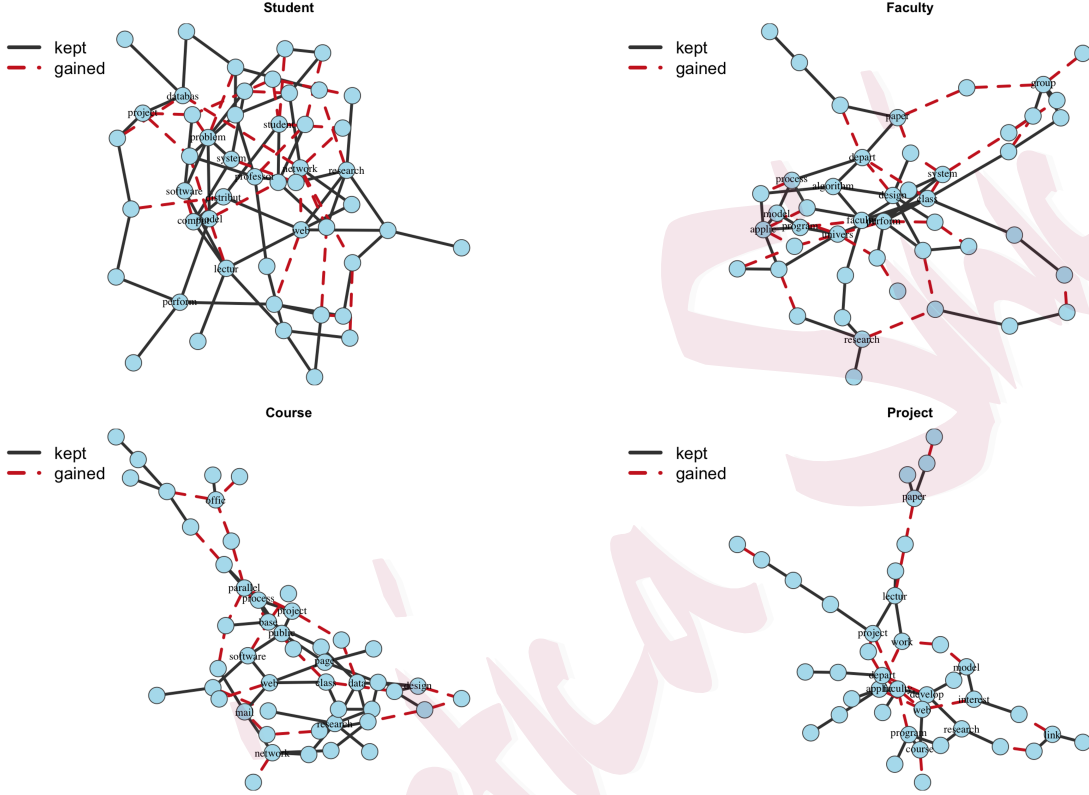


Figure 6: Visualization of WebKB data: Debiased edge discoveries with FDR control at level $q = 0.1$ under $p = 50$.

are retained after inference ($|E_{\text{kept}}^{(k)}|$), while a non-negligible portion of the initial edges are removed ($|E_{\text{lost}}^{(k)}|$). This pattern indicates that many edges selected by the initial regularized fit are not supported by sufficient statistical evidence after regularization-induced bias is corrected and multiplicity is accounted for.

At the same time, the refinement is not merely “pruning” the initial network. A substantial number of edges are newly detected after debiasing ($|E_{\text{gained}}^{(k)}|$), suggesting that debiasing can also identify edges that may be overly shrunk or missed by the coupled regularization.

Category k	n_k	q	$ E_{\text{init}}^{(k)} $	$ E_{\text{sig}}^{(k)} $	$ E_{\text{kept}}^{(k)} $	$ E_{\text{lost}}^{(k)} $	$ E_{\text{gained}}^{(k)} $
Student	1625	0.1	165	88	57	108	31
Faculty	1116	0.1	145	74	46	99	28
Course	926	0.1	132	69	43	89	26
Project	501	0.1	110	48	29	81	19

Table 2: Summary of WebKB data: Debiased edge inference with FDR control at level $q = 0.1$ under $p = 50$.

Overall, Table 2 shows that the procedure yields fewer discoveries at $p = 50$ while identifying additional edges not selected by the initial estimator.

We further visualize the refined edge discoveries at $p = 50$ in Figure 6. In each category-specific panel, we display only the debiased significant edge set $E_{\text{sig}}^{(k)}$. Within $E_{\text{sig}}^{(k)}$, retained edges $E_{\text{kept}}^{(k)}$ are shown as solid black lines, whereas newly detected edges $E_{\text{gained}}^{(k)}$ are shown as red lines.

In the additional analysis with $p = 100$, the procedure likewise retains some initial edges, removes others, and identifies additional edges. The number of significant edges is smaller than the initial edge count for the Faculty, Course, and Project categories, whereas it increases from 400 to 474 for the Student category. These results further illustrate that the procedure can both remove and add edges when refining the initial network estimates. The corresponding table and network visualization are provided in the Supplementary Material.

7. Conclusion

This paper develops a unified debiasing framework for statistical inference in high-dimensional multi-graph models. Exploiting the KKT optimality conditions, we derive an analytic debiasing representation for jointly regularized precision matrix estimators that corrects the

leading regularization-induced bias while accounting for cross-graph coupling. The resulting formula admits a universal interpretation as a residual adjustment associated with the fundamental precision–covariance identity $\Theta^{(m)}\Sigma^{(m)} = \mathbf{I}$, which clarifies that the debiasing step does not rely on likelihood-based constructions and can be paired with a broad class of initial estimators satisfying the required consistency conditions. This yields entry-wise asymptotic normality and enables confidence intervals, simultaneous testing with asymptotically valid control of both FWER and FDR, as well as differential structure detection via linear contrasts of the debiased estimators. Empirically, the debiased statistics exhibit accurate Gaussian approximation and deliver stable inferential performance across a range of representative initial multi-graph estimators.

A notable feature of the proposed approach is its modularity: the debiasing step serves as an inference layer that can be paired with many jointly regularized multi-graph estimators, provided the required consistency conditions hold and the regularizer includes a sparsity-inducing component. This separation allows practitioners to choose an estimator that encodes domain-specific cross-group similarity, and then apply a unified inferential correction. Future work includes sharpening theory for specific coupling patterns, extending inference beyond entry-wise targets to higher-level network functionals, and developing robust variants for non-Gaussian or heavy-tailed settings.

Supplementary Material

The online Supplementary Material contains the related algorithms, detailed proofs, additional simulation results and a real-data example with $p = 100$.

Acknowledgements

This research of Riquan Zhang is supported by the National Natural Science Foundation of China (No. 12371272 and 12531013). The research of Yibo Yan is supported by the the National Natural Science Foundation of China (No. 12401390).

References

- Belilovsky, E., G. Varoquaux, and M. Blaschko (2016). Testing for differences in Gaussian graphical models: Applications to brain connectivity. *Advances in Neural Information Processing Systems(NIPS)* 29(1), 595–603.
- Benjamini, Y. and D. Yekutieli (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* 29(4), 1165–1188.
- Bühlmann, P. and S. van de Geer (2011). *Statistics for High-dimensional Data: Methods, Theory and Applications*. Springer Science & Business Media.
- Cai, T. T., H. Li, W. Liu, and J. Xie (2016). Joint estimation of multiple high-dimensional precision matrices. *Statistica Sinica* 26(2), 445–464.
- Cai, T. T., W. Liu, and X. Luo (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association* 106(494), 594–607.
- Chernozhukov, V., D. Chetverikov, and K. Kato (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Annals of Statistics* 41(6), 2786–2819.
- Danaher, P., P. Wang, and D. M. Witten (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 76(2), 373–397.
- Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96(456), 1348–1360.
- Friedman, J., T. Hastie, and R. Tibshirani (2008). Sparse inverse covariance estimation with the graphical lasso.

REFERENCES

- Biostatistics* 9(3), 432–441.
- Guo, J., E. Levina, G. Michailidis, and J. Zhu (2011). Joint estimation of multiple graphical models. *Biometrika* 98(1), 1–15.
- Hao, B., W. W. Sun, Y. Liu, and G. Cheng (2018). Simultaneous clustering and estimation of heterogeneous graphical models. *Journal of Machine Learning Research* 18(217), 1–58.
- Janková, J. and S. van de Geer (2015). Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics* 9(1), 1205–1229.
- Kuhn, H. and A. Tucker (1951). Nonlinear programming. In *Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 481–492.
- Lauritzen, S. L. (1996). *Graphical Models*. Clarendon Press.
- Lee, W. and Y. Liu (2015). Joint estimation of multiple precision matrices with common structures. *Journal of Machine Learning Research* 16(1), 1035–1062.
- Li, Y., J. Hu, C. Zhang, D.-J. Yu, and Y. Zhang (2019). ResPRE: High-accuracy protein contact prediction by coupling precision matrix with deep residual neural networks. *Bioinformatics* 35(22), 4647–4655.
- Liu, W. (2017). Structural similarity and difference testing on multiple sparse Gaussian graphical models. *Annals of Statistics* 45(6), 2680–2707.
- Ma, J. and G. Michailidis (2016). Joint structural estimation of multiple graphical models. *Journal of Machine Learning Research* 17(166), 1–48.
- Majumdar, S. and G. Michailidis (2022). Joint estimation and inference for data integration problems based on multiple multi-layered Gaussian graphical models. *Journal of Machine Learning Research* 23(1), 1–53.
- Meinshausen, N. and P. Bühlmann (2006). High-dimensional graphs and variable selection with the lasso. *Annals of Statistics* 34(3), 1436–1462.
- Neykov, M., Y. Ning, J. S. Liu, and H. Liu (2018). A unified theory of confidence regions and testing for high-dimensional estimating equations. *Statistical Science* 33(3), 427–443.

REFERENCES

- Porter, M. F. (1980). An algorithm for suffix stripping. *Program* 14(3), 130–137.
- Saegusa, T. and A. Shojaie (2016). Joint estimation of precision matrices in heterogeneous populations. *Electronic Journal of Statistics* 10(1), 1341–1392.
- Schütze, H., C. D. Manning, and P. Raghavan (2008). *Introduction to Information Retrieval*, Volume 39. Cambridge University Press Cambridge.
- TCGA (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455(7216), 1061–1068.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58(1), 267–288.
- Tibshirani, R., M. Saunders, S. Rosset, J. Zhu, and K. Knight (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 67(1), 91–108.
- Tsai, K., O. Koyejo, and M. Kolar (2022). Joint Gaussian graphical model estimation: A survey. *Wiley Interdisciplinary Reviews: Computational Statistics* 14(6), e1582.
- Wang, B., R. Singh, and Y. Qi (2017). A constrained ℓ_1 minimization approach for estimating multiple sparse Gaussian or nonparanormal graphical models. *Machine Learning* 106(9), 1381–1417.
- Wang, Y., S. Segarra, and C. Uhler (2020). High-dimensional joint estimation of multiple directed Gaussian graphical models. *Electronic Journal of Statistics* 14(1), 2439–2483.
- Xia, Y., T. Cai, and T. T. Cai (2015). Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika* 102(2), 247–266.
- Yuan, M. and Y. Lin (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 68(1), 49–67.
- Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics* 38(2), 894–942.
- Zheng, Q., R. Zhang, G. Li, and Y. Zhong (2025). Efficient distributed differential network estimation. *Stat* 14(1),

REFERENCES

e70049.

Zheng, Q., R. Zhang, and Y. Zhong (2026). Communication-efficient large precision matrix estimation by distributed refitted cross-validation. *Journal of Systems Science and Complexity*, 1–18.

Qiao Zheng

School of Statistics and Data Science, Shanghai University of International Business and Economics, Shanghai, 201620, China.

E-mail: zhengqiao@suibe.edu.cn

Gefei Li

KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai, 200062, China.

E-mail: gefeili197@gmail.com

Yibo Yan

School of Statistics and Data Science, Shanghai University of International Business and Economics, Shanghai, 201620, China.

E-mail: yyb820nb@sina.com

Riquan Zhang

School of Statistics and Data Science, Shanghai University of International Business and Economics, Shanghai, 201620, China.

E-mail: zhangriquan@163.com