Robust estimation of graphical models with measurement error: False discovery control and sensitivity analysis

Shira Mingelgrin[1] **Y. Samuel Wang**[1]

[1]Department of Statistics and Data Science, Cornell University, Ithaca, New York, USA

Abstract

This paper addresses the problem of learning Gaussian graphical models from data contaminated by additive measurement error. Applying traditional procedures to the noisy data may result in many false discoveries. On the other hand, existing procedures that explicitly account for measurement error assume that the measurement error covariance is known or can be estimated precisely. We provide two practical alternatives for the common setting where this side information is not available. The first proposed method follows a partial identification approach and estimates a graph with controlled false discovery rate when given an upper bound on the measurement error variances. In many settings, the precise variance of the errors is unknown, but the scientist may still confidently specify an upper bound. The second framework is a sensitivity analysis that assesses the minimum amount of measurement error required to explain away an estimated edge. This yields an interpretable measure for practitioners to assess which estimated edges are robust and which may be simply due to measurement error. We derive theoretical guarantees and show good empirical performance in simulations. We further illustrate the practical utility of our procedures on a single-cell mRNA sequencing dataset.

1 INTRODUCTION

Graphical modeling is a popular tool in the natural and social sciences for representing conditional independence relationships in multivariate data (Lauritzen, 1996; Uhler, 2018). Specifically, a multivariate system can be represented by an (undirected) conditional independence graph (CIG) where each node in the graph corresponds to a variable, and the absence of an edge between two nodes indicates that the corresponding variables are conditionally independent given all remaining variables. For example, graphical models are used in neuroscience to model brain connectivity patterns (Ortiz et al., 2015), in systems biology to represent gene regulatory networks (Marbach et al., 2012), and in psychology to model interactions of symptoms that co-occur in psychological disorders, such as PTSD (Armour et al., 2017).

In many applications, however, the variables of interest are not directly observed, and the sample is contaminated by measurement error. For instance, in genomics, sequencing protocols to extract gene expression counts do not capture all mRNA molecules in a cell. This leads to sampling noise since only a subset of mRNA transcripts are captured in each cell (Klein et al., 2015). In these settings, the conditional independence relations that hold for the "true" data generally do not hold for the noisy observations. Specifically, the population CIG for the noisy data will typically contain many additional spurious edges that do not exist in the CIG for the "true" data. As a result, naively applying standard methods for graph estimation will result in an unreliable estimated graph with many false discoveries.

There is a vast literature on explicitly correcting for measurement error in general settings (Carroll et al., 2006) and growing interest in accounting for measurement error when estimating a graphical model (Zhang et al., 2018; Katiyar et al., 2019; Nikolakakis et al., 2019; Saeed et al., 2020a; Byrd et al., 2021; Shi et al., 2021; Casanellas et al., 2024; Dai et al., 2022). However, virtually all proposed methods require some form of side information: typically the covariance of the measurement error must be known a priori or can at least be estimated consistently from replicates. Unfortunately, in the typical situation where such side information is not readily available, practitioners are left with no viable alternative aside from simply ignoring the error and potentially suffering many false discoveries. As an alternative, we propose two complementary frameworks that require much weaker prior information when estimating a

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:4528–4542.

Gaussian graphical model (GGM) from data contaminated by additive measurement error.

The first approach follows a partial identification framework (Kline and Tamer, 2023). When the variances of the measurement errors are unknown, the conditional independence graph is not identifiable. However, in most settings, the scientist may confidently specify a reasonable upper bound on the proportion of the total observed variance that is due to measurement error. This upper bound allows the partial correlations to be partially identified and enables our proposed edge-wise hypothesis tests across a range of admissible noise levels. We then combine these tests with a multiple testing correction procedure (Benjamini and Yekutieli, 2001) to produce a reliable estimated graph that controls the false discovery rate even in the presence of measurement error.

The second approach is a sensitivity analysis framework. In the causal inference literature, sensitivity analysis is a common tool to evaluate how strongly a statistically significant result depends on untestable assumptions such as "no unobserved confounding" (Cornfield et al., 1959; Rosenbaum and Rubin, 1983; Imbens, 2003; Franks et al., 2020). In a similar spirit, we propose a sensitivity analysis that allows the practitioner to quantify the minimum amount of measurement error that could explain away an estimated edge. This provides a transparent measure of robustness and allows researchers to reliably distinguish which edges in the estimated graph are robust to the presence of measurement error and which edges may only appear due to a small amount of measurement error.

Together, our proposed approaches provide valuable tools for practitioners who require dependable graphical model estimates but may not have precise knowledge of the measurement error variability. This complements the existing literature by providing a novel alternative along the assumption vs guarantee trade-off frontier. Philosophically, our framework avoids exploiting strong (and potentially false) assumptions to gain a (potentially false) increase in precision. Instead, we prioritize robustness by deliberately relying on weaker assumptions that may result in fewer but ultimately more reliable discoveries. We derive theoretical properties of our proposed approaches and demonstrate their empirical effectiveness through simulation studies. We further illustrate the practical utility by applying our procedures to a single-cell mRNA sequencing dataset, identifying robust conditional dependencies while accounting for measurement errors.

2 PRELIMINARIES

Let $X \in \mathbb{R}^p$ be a random vector with distribution $\mathcal{N}(0, \Sigma)$, and let $\Omega = \Sigma^{-1}$. Let $G = (V, E)$ denote the undirected conditional independence graph associated with X, where $V = \{1, \ldots, p\}$ is the set of vertices and E is the set of undirected edges. For $u, v \in V$, the absence of an edge between u and v (i.e., $u - v \notin E$) encodes $X_u \perp\!\!\!\perp X_v \mid X_{V \setminus \{u,v\}}$, where $V \setminus \{u, v\}$ denotes the set of all vertices other than u and v. In the multivariate Gaussian, $X_u \perp\!\!\!\perp X_v \mid X_{V \setminus \{u,v\}}$ is equivalent to $\Omega_{u,v} = 0$ (Lauritzen, 1996). In the typical setting where X is observed directly, popular methods for graph estimation include the graphical lasso (Banerjee et al., 2008; Yuan and Lin, 2007; Friedman et al., 2008), CLIME (Cai et al., 2011), and neighborhood selection (Meinshausen and Bühlmann, 2006).

The false discovery rate (FDR) of a graph estimation procedure is $\mathbb{E}\left((|\hat{E} \setminus E|)/\max\{|\hat{E}|, 1\}\right)$ where $\hat{E}$ is the estimated set of edges. Several procedures have been proposed that control the FDR below a user-specified level α (Drton and Perlman, 2007; Liu, 2013; Janková and van de Geer, 2015; Lee et al., 2019; Li and Maathuis, 2021; Yu et al., 2021; Koka et al., 2024). Our approach is most similar to that of Drton and Perlman (2007), which conducts a hypothesis test for each potential edge and controls the FDR by adjusting for multiple testing (Benjamini and Yekutieli, 2001). However, all of the previous approaches for controlling FDR assume direct access to data associated with the graph and are not applicable when the observed samples are contaminated by measurement error.

We consider the setting where X is corrupted by additive error so that instead of observing X directly, we only observe $W = X + U$. The measurement error, $U \in \mathbb{R}^p$, is independent of X and has mean 0 and diagonal covariance D. In linear regression with measurement error, practitioners are primarily concerned with the problem of attenuation bias, whereby estimates of nonzero regression coefficients are biased toward zero and increase Type II errors (Fuller, 2009). Although attenuation bias also arises in graphical models, a different issue is often more pressing: the population CIG for the contaminated measurements W typically contains many spurious edges that are not present in the CIG for X. Thus, the measurement error may result in increased Type I errors.

For intuition, consider the example graph in Fig. 1a where $X_1 \perp\!\!\!\perp X_3 \mid X_2$. Since U is independent of X, $W_j \perp\!\!\!\perp \{W_k, X_k\}_{\{k \neq j\}} \mid X_j$, but W_j is still dependent with X_j when conditioning on all other nodes. A CIG for W and X together is shown in Fig. 1b, and we can see that $W_1 \perp\!\!\!\perp W_3 \mid X_2$. However, since W_2 is a noisy version of X_2, conditioning on W_2 does not fully remove the dependence between W_1 and W_3 so $W_1 \not\perp\!\!\!\perp W_3 \mid W_2$. The CIG when only considering the observed data, W, and marginalizing out X, the unobserved process of interest, is given in Fig. 1c. Notably, there is an extra edge between W_1 and W_3 even though there is no edge between X_1 and X_3 in the true graph given in Fig. 1a. Thus, using the observed data, W to estimate a CIG for X will result in additional

edges. A numerical example corresponding to this CIG is provided in Appendix B.

More generally, the edge $u - v$ will only be absent in the CIG for W if the underlying X_u and X_v are marginally independent. Thus if X_u and X_v are marginally dependent but conditionally independent given all other nodes (i.e., in the CIG for X, u and v are connected by a path of edges but are not directly connected by an edge), then the CIG for W will contain the spurious edge $u - v$ that is not present in the CIG for X. We can also see in Fig. 2 that $(\Sigma + D)^{-1}$, which encodes the conditional independence relations for W, does not typically share the same support as Σ^{-1}, which encodes the conditional independence relations for X.

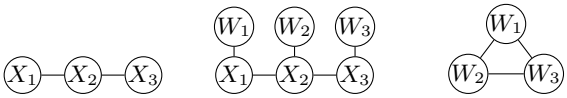


Figure 1: (a) The CIG for the "true" variables X_1, X_2, X_3, where $X_1 \perp\!\!\!\perp X_3 \mid X_2$. (b) The CIG for "true" and noisy variables, where each W_j is connected to its corresponding X_j. (c) The CIG for the noisy observed variables W_1, W_2, W_3.

A recently growing body of work has focused on accounting for measurement error in structure learning procedures, but the proposed methods rely on strong assumptions. Both Bayesian (Byrd et al., 2021; Shi et al., 2021) and frequentist (Saeed et al., 2020b) approaches have been developed to estimate a CIG or sparse precision matrix under measurement error. However, these methods assume that the covariance of the measurement error is either known or can be consistently estimated from replicate measurements, or, similarly, that the partial correlations of the true variables are a known function of the covariance of the noisy variables. Alternatively, under the strict assumption that the true graph is a tree (i.e., a graph with no cycles), the CIG can be recovered from noisy samples when the variances of the measurement errors are unknown but "small enough" (Katiyar et al., 2019; Nikolakakis et al., 2019; Tandon et al., 2021; Casanellas et al., 2024). Finally, several methods have been proposed for recovering directed causal graphs from non-Gaussian linear structural equation models corrupted by measurement error (Zhang et al., 2018; Dai et al., 2022; Yang et al., 2022; Zhang et al., 2025). These methods do not require knowledge of the measurement error covariance, but the non-Gaussian assumption is crucial for identification and estimation. In contrast to the existing literature, we do not require that the variances of the errors are known or can be estimated precisely; instead, we only require an upper bound. Under these weak assumptions, the graph is not identifiable. Nonetheless, our proposed procedure delivers a reliable estimate by controlling the FDR and exhibits strong empirical performance under the weakened identifiability conditions.

3 PROPOSED METHODOLOGY

In the following section, we first introduce the model assumptions and develop adjusted partial correlation tests for recovering the "true" conditional independence relationships from noisy observations. We then detail our false discovery control and sensitivity analysis frameworks.

3.1 ASSUMED MODEL

Given n i.i.d noisy observations of $W \in \mathbb{R}^p$, our goal is to recover the undirected graph $G = (V, E)$ associated with the latent X.

Assumption 1. *The process of interest is $X \sim \mathcal{N}(0, \Sigma)$, and we observe $W = X + U$. The measurement error U, is independent of X has mean 0 with diagonal covariance matrix D so that cov$(W) = \Sigma + D$, which we denote as Θ.*

Assumption 1 notably implies that the measurement errors are independent of the true observations, X, and that $\mathbb{E}(U_u U_v) = 0$ for all $u, v \in V$. This is a reasonable assumption in settings such as single-cell mRNA sequencing, where technical noise is modeled separately for each gene (Brennecke et al., 2013). For notational convenience, let Θ_D and Σ_D denote the diagonal matrices formed by taking the diagonal entries of Θ and Σ and letting all other elements be 0.

Assumption 2. *The measurement error variances satisfy one of the following conditions.*

(a) For some $1 > \delta \geq 0$, $D_{jj} = \delta\, \Theta_{jj}$ for all $j \in V$, or

(b) For some $\delta \geq 0$, $D = \delta I$.

Assumption 2 implies that either (a) the proportion of observed variance that is due to the measurement error is equal across all variables, or (b) the measurement errors have equal variance across all variables. For instance, if each variable is measured by the same type of sensor, it is reasonable that the variability of the measurement errors has similar magnitude. Assumption 2(b) is also a common assumption in many causal discovery procedures (Peters et al., 2014; Loh and Bühlmann, 2014; Ghoshal and Honorio, 2017; Chen et al., 2019). When each coordinate of X has the same variance, then (a) and (b) are equivalent; however, (a) may be more realistic when the variables have drastically different scales. For brevity, the remainder of the paper will work under Assumption 2(a); however, the methods and theory generalize immediately to Assumption 2(b) in a straightforward way.

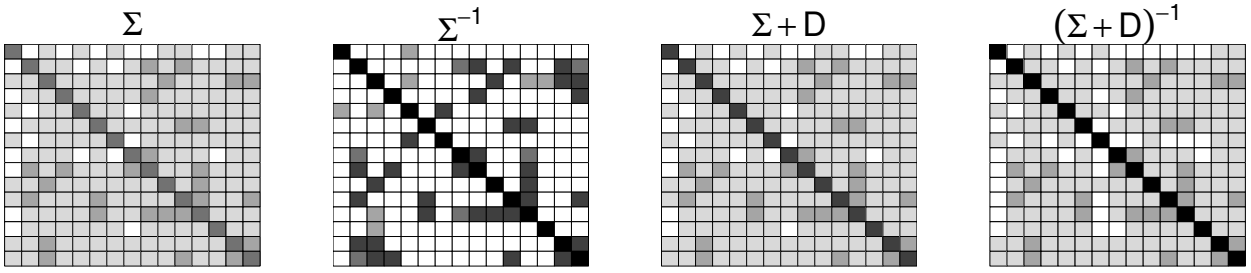


Figure 2: Panels (a) and (b) show a visualization of the covariance and precision matrices of the latent variable X, and panels (c) and (d) show those of the observed variable W. White cells indicate zero entries, while darker shading denotes larger magnitudes. Nonzero entries in (b) and (d) correspond to edges in the CIG. The precision matrix of X is sparse, but the precision matrix of W is dense, leading to many false positives when the graph for X is estimated directly from W.

Assumption 2 asserts a specific structure for the measurement error covariance. While this structure may initially seem restrictive, we emphasize that it does not require δ to be known in advance. Furthermore, we emphasize that all previous work with general graphs essentially assume that D is either known or that the analyst has access to replicates so that it may be consistently estimated. In the many settings where neither are true, our alternative assumption is a drastic relaxation of the only other practical alternative: assuming no measurement error. Although the theory extends straightforwardly beyond the assumed covariance structure, allowing the error variance proportions to differ between variables introduces computational challenges. Nonetheless, we explore the impact of violating Assumption 2 in Section 4.1.1 and see that our procedures still perform well empirically.

For multivariate Gaussians, conditional independence $X_u \perp\!\!\!\perp X_v \mid X_{V\setminus\{u,v\}}$ is equivalent to $\rho_{u,v\cdot V\setminus\{u,v\}} = 0$ where $\rho_{u,v\cdot V\setminus\{u,v\}}$ is the partial correlation of X_u and X_v given $X_{V\setminus\{u,v\}}$ (Lauritzen, 1996). For the remainder of the paper, we will drop the additional notation for brevity and simply use $\rho_{u,v\cdot V}$. The partial correlation may be computed directly from entries of the precision, $\Omega = \Sigma^{-1}$ where $\rho_{u,v\cdot V} = \frac{-\Omega_{u,v}}{\sqrt{\Omega_{uu}\Omega_{vv}}}$. Thus, when X is directly observed, Drton and Perlman (2007) propose including the potential edge $u - v$ in $\hat{E}$ if the following null hypothesis is rejected after adjusting for multiple testing:

$$H_0^{(u,v)} : \rho_{u,v\cdot V} = 0. \tag{1}$$

When we only observe W, we cannot directly estimate or test the empirical partial correlation $\hat{\rho}_{u,v\cdot V}$. However, suppose for the moment that $d = \delta$ is known. We could then adjust, $\hat{\Theta}$, the observed sample covariance matrix of W, to consistently estimate the covariance of X using $\hat{S}_d = \hat{\Theta} - d\hat{\Theta}_D$. We could subsequently estimate the adjusted sample partial correlation for X:

$$\hat{\rho}_{u,v\cdot V}(d) = \frac{-(\hat{S}_d^{-1})_{u,v}}{\sqrt{(\hat{S}_d^{-1})_{u,u}(\hat{S}_d^{-1})_{v,v}}}. \tag{2}$$

We will use S_d and $\rho_{u,v\cdot V}(d)$ to denote the corresponding population quantities. Although $\hat{S}_d$ and $\hat{\rho}_{u,v\cdot V}(d)$ will consistently estimate Σ and $\rho_{u,v\cdot V}$ when $d = \delta$, they have a different distribution than the sample covariance and sample partial correlations calculated directly from the unobserved X. Theorem 1 gives the asymptotic distribution of $\hat{\rho}_{u,v\cdot V}(d)$ and uses intermediate Lemma 1, which describes the asymptotic distribution of $\hat{S}_d^{-1}$. Further details can be found in the proof for Theorem 1 in Appendix A.

Lemma 1. *Let* $\mathrm{Iss}(A)$ *denote the* $p(p+1)/2 \times p(p+1)/2$ *Isserlis matrix associated with a symmetric matrix* $A \in \mathbb{R}^{p\times p}$*, with entries*

$$\mathrm{Iss}(A)_{ij,lk} = A_{il}A_{jk} + A_{ik}A_{jl},$$

for $1 \le i \le j \le p$ *and* $1 \le l \le k \le p$*.*

Assume that S_d *is positive definite, then*

$$\sqrt{n-1}\,(\hat{S}_d^{-1} - S_d^{-1}) \xrightarrow{d} \mathcal{N}(0, V_S),$$

where

$$V_S = \mathrm{Iss}(S_d^{-1})\mathrm{Iss}(I)^{-1}M\mathrm{Iss}(\Theta)M\mathrm{Iss}(I)^{-1}\mathrm{Iss}(S_d^{-1}),$$

I *denotes the* $p \times p$ *identity matrix, and* M *is the diagonal* $p(p+1)/2 \times p(p+1)/2$ *matrix with entries*

$$M_{ij,ij} = \begin{cases} 1, & i < j \\ 1-d, & i = j. \end{cases}$$

Theorem 1. *Under the null hypothesis that* $\rho_{u,v\cdot V} = 0$*, the test statistic* $T_{u,v}(d)$ *when* $d = \delta$ *is asymptotically normally distributed:*

$$\sqrt{n-p+1}\,T_{u,v}(d) = \frac{\hat{\rho}_{u,v\cdot V}(d)}{\sqrt{\widehat{\mathrm{Var}}(\hat{\rho}_{u,v\cdot V}(d))}} \xrightarrow{d} \mathcal{N}(0,1) \tag{3}$$

and the corresponding unadjusted p-values are

$$\pi_{u,v}(d) = 2\left[1 - \Phi\left(\sqrt{n-p+1}\,|T_{u,v}(d)|\right)\right] \quad (4)$$

where Φ is the cdf of the standard normal distribution.

Note in Theorem 1, the variance is the estimated asymptotic variance of the adjusted partial correlation. By the continuous mapping theorem, we can replace the unknown true covariance with the sample covariance to obtain a consistent estimator. Additionally, when calculating the p-value, we scale the test statistic by $\sqrt{n-p+1}$ as a finite sample correction for the degrees of freedom, as we condition on $(p-2)$ variables (Anderson, 2003).

3.2 GRAPH ESTIMATION UNDER PARTIAL IDENTIFICATION

In most situations, the scientist does not have precise knowledge of the measurement error variance. Thus, each partial correlation (and subsequently the CIG) is not identified and consistent recovery of the graph is impossible. However, the scientist can typically specify a reasonable upper bound on the proportion of the observed variance that may be due to measurement error. In practice, this bound can be derived from prior knowledge of the precision of the measurement devices, such as in astronomy (Jin et al., 2025). Alternatively, previous experiments with replicates may provide an upper bound, as seen in systems biology (Klein et al., 2015; Grün et al., 2014). We denote this upper bound as $\delta_{\max}$. For example, the scientist may be confident that the variance of the measurement error does not exceed the variance of the true process; this would imply $\delta_{\max} = 1/2$. In other settings, the scientist might be confident that $\delta_{\max}$ may be even smaller. In order to calculate $\hat{\rho}_{u,v\cdot V}(\delta_{\max})$, we must also constrain $\delta_{\max} \leq \sup\left\{t \in [0,1) : \lambda_{\min}(\hat{S}_t) > 0\right\}$ so that $\hat{S}_d$ is positive definite. Under this limited prior knowledge, the partial correlations (and subsequently the CIG) are now partially identified. In this section, we propose a robust estimate of the CIG that controls the FDR.

We outline the procedure in Alg. 1. Specifically, for each potential edge $u - v$, we test the composite null hypothesis:

$$H_0^{(u,v,\delta_{\max})} : \exists d \in [0, \delta_{\max}] \text{ such that } \rho_{u,v\cdot V}(d) = 0. \quad (5)$$

Because the partial correlations are only partially identified, the hypothesis in Eq. (5) that we can test with observed data is necessary but not sufficient for the hypothesis in Eq. (1), which we cannot directly test. Thus, even with infinite data, we may fail to reject Eq. (5) when Eq. (1) is false. Crucially, controlling the Type I error and FDR for Eq. (5) also controls the Type I error and FDR for Eq. (1).

To test the hypothesis in Eq. (5), we use the test statistic

$$T^{\star}_{u,v} = \min_{d \in [0, \delta_{\max})} |T_{u,v}(d)|, \quad (6)$$

which can be compared to the quantiles of a standard normal to produce a p-value. Under the null, the p-value is stochastically lower bounded by a uniform distribution and thus is a "valid" p-value. Because we are testing $m = \frac{p(p-1)}{2}$ hypotheses, adjusting for multiple testing is essential to avoid inflated Type I error rates. Since the tests for each potential edge may be dependent, we apply the Benjamini and Yekutieli (2001) (BY) procedure to control the FDR. The procedure adjusts each p-value as follows: The p-values are ordered in increasing order, and define $k = \max_i \left\{\pi^{(i)} \leq \frac{i}{m}\frac{\alpha}{c(m)}\right\}$, where $c(m) = \sum_{i=1}^{m} \frac{1}{i}$. We reject the null hypotheses for all indices $i \leq k$ ensuring that the proportion of false discoveries is controlled at level α. Theorem 2 formally states that our procedure controls the FDR and follows from Theorem 1 and Benjamini and Yekutieli (2001).

Algorithm 1: Hypothesis Testing

Input: Sample covariance $\hat{\Theta}$, level α, $\delta_{\max} \in [0, 1)$
Output: Estimated edge set $\hat{E}$
for *each pair* $u < v$ **do**
  $T^{\star}_{u,v} = \min_{d \in [0, \delta_{\max}]} |T_{u,v}(d)|$
  $\pi^{\star}_{u,v} = 2\left[1 - \Phi\left(\sqrt{n-p+1}\,T^{\star}_{uv}\right)\right]$

Adjust $\pi^{\star}_{u,v}$ according to the BY procedure

Define: $\hat{E} = \{(u,v) : H_0^{(u,v,\delta_{\max})}$ rejected at level $\alpha\}$
return $\hat{E}$

Theorem 2. *Suppose Assumptions 1 and 2(a) hold for some $\delta \leq \delta_{\max}$. Then, the set of edges $\hat{E}$ estimated by Alg. 1 controls the asymptotic FDR below level α; i.e.,*

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{|\hat{E} \setminus E|}{\max\{|\hat{E}|, 1\}}\right] \leq \alpha. \quad (7)$$

3.3 SENSITIVITY ANALYSIS

We now propose a sensitivity analysis approach that is "dual" to the false discovery control perspective. Our approach adapts ideas from the causal inference literature where a sensitivity analysis for unobserved confounding allows the practitioner to "stress test" a statistically significant result by asking "What is the minimum amount of unmeasured confounding that, if properly accounted for, would overturn the current result?" (Cornfield et al., 1959; Rosenbaum and Rubin, 1983; VanderWeele and Ding, 2017; Cinelli and Hazlett, 2019). In practice, no analysis can perfectly adjust for all confounding. However, if the magnitude of unobserved confounding required to overturn the finding is small, then the scientist should not be

confident that the result would hold in an "idealized analysis" that properly adjusts for all confounding. If the confounding required to overturn the result exceeds what the scientist believes to be plausible, then the finding can be considered robust.

Analogously, our proposed approach allows a practitioner to assess the robustness of an estimated edge in $\hat{E}$ by asking "What is the smallest amount of measurement error that, if properly accounted for, would result in removing this estimated edge?" If the size of the measurement error (parameterized by the proportion of observed variance due to measurement error) required to remove the edge is much larger than what the scientist deems plausible, then they may be confident that the edge would persist in the ideal analysis which properly accounts for the true measurement error. On the other hand, if the minimum amount of measurement error might be plausible for their setting, the practitioner should be wary that the estimated edge may be removed in the ideal analysis.

Algorithm 2: Sensitivity Analysis

Input: Sample covariance $\hat{\Theta}$, level α
Output: Matrix $D_{\min}$ of minimal measurement error

Define: $d_{\max} = \sup\{t \in [0,1) : \lambda_{\min}(\hat{S}_t) > 0\}$

for *each* $\eta \in [0, d_{\max}]$ **do**
  Define $\tilde{\pi}_{u,v}(\eta) = \max_{t \in [0,\eta]}\{\pi_{u,v}(t)\}$ from Eq. (4)
  Adjust $\tilde{\pi}_{u,v}(\eta)$ according to the BY procedure

for *each pair* $u < v$ **do**
  if $H_0^{(u,v,0)}$ *is rejected at level* α **then**
    $(D_{\min})_{uv} = \inf\left\{\eta : H_0^{(u,v,\eta)} \text{ is not rejected}\right\}$
  else
    $(D_{\min})_{uv} = 0$

return $D_{\min}$

The procedure is outlined in Alg. 2. For each pair $u, v \in V$, we compute the test statistic $T_{u,v}(d)$ and the corresponding p-values over a grid of d values. To control the FDR, we focus on edges that are initially significant when ignoring measurement error ($d = 0$) but may become insignificant as d increases. For each pair, the procedure outputs the minimum amount of measurement error that, if properly accounted for, would result in removing the estimated edge. The scientist can then assess whether the implied level of measurement error is plausible or whether the edge should be regarded as robustly estimated.

While the sensitivity analysis does not yield formal guarantees similar to the previously proposed false discovery control approach, it does yield an intuitive and interpretable measure of how robust each estimated edge is to measurement error. The procedure can be applied post hoc, complementing standard graphical model analyses that do not explicitly account for measurement error and helping researchers better understand the robustness of the estimated graph.

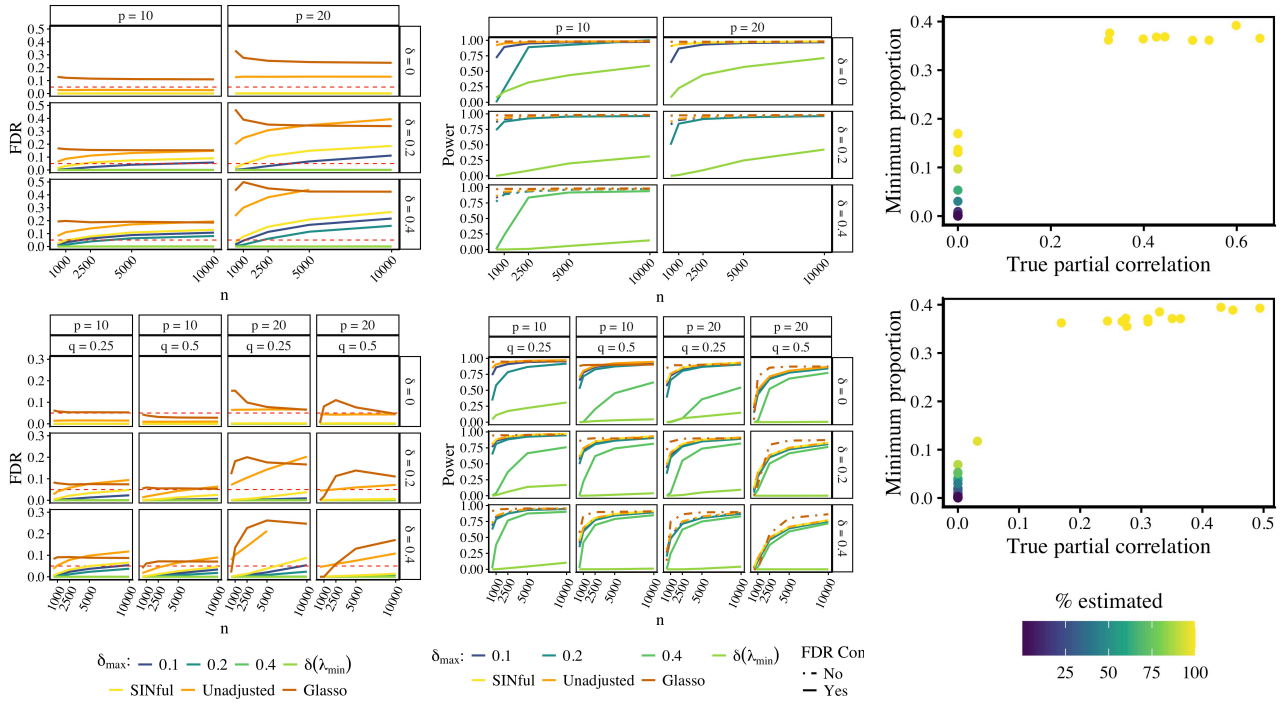
4 NUMERICAL RESULTS

We now evaluate the finite-sample performance of the proposed method in simulations and apply our methods to a real example of gene expression data.

4.1 SIMULATIONS

We first conduct simulations where the true CIG for X is a path graph; i.e., $X_1 - X_2 - \ldots - X_p$ such that $X_u \perp\!\!\!\perp X_v \mid X_{V\setminus\{u,v\}}$ if $|u - v| > 1$. This model is particularly challenging in the presence of measurement error, because all variables are marginally dependent. Thus the CIG for W is a complete graph although the CIG for X is sparse. We consider a p-dimensional Gaussian graphical model with $p \in \{10, 20\}$. For each setting, the precision matrix Ω has nonzero entries only on the superdiagonal and subdiagonal, corresponding to adjacent nodes in the path, with values drawn from $\mathcal{U}[-1, 1]$. Diagonal entries are set as $\Omega_{uu} = \sum_{v \neq u} |\Omega_{uv}| + 0.1$, ensuring strict diagonal dominance and hence positive definiteness of Ω. The corresponding covariance matrix is given by $\Sigma = \Omega^{-1}$. Measurement error is incorporated through an additive diagonal matrix $D = \frac{\delta}{1-\delta}\Sigma_D$, where $\delta \in \{0, 0.2, 0.4\}$ denotes the proportion of observed variance attributable to measurement error. Observed data are sampled from a mean-zero Gaussian distribution with covariance $\Sigma + D$. Sample sizes are taken from $n \in \{500, 1000, 2500, 5000, 10000\}$. For a more general case, we also consider Erdös-Renyi random graphs where edges are added between each pair of nodes with probability $q \in \{0.25, 0.5\}$. Given the graph, we generate data in the same way as before. The results from the path graphs are shown in the top panels of Fig 3 and the results from the random graphs are shown in the bottom panels.

To examine the performance of the proposed FDR control procedure from Section 3.2 we generate 500 independent datasets for each combination of (p, δ, n). Each dataset is analyzed under eight methods: (i) a naive approach assuming no measurement error ($\delta_{\max} = 0$) and no multiple hypothesis testing adjustment (ii) a naive approach assuming no measurement error ($\delta_{\max} = 0$), (iii) our proposed procedure with $\delta_{\max} = 0.1$, (iv) our proposed procedure with $\delta_{\max} = 0.2$ (v) our proposed procedure with $\delta_{\max} = 0.4$, (vi) our proposed procedure with a maximal bound ($\delta(\lambda_{\min})$) based on the minimum eigenvalue constraint $\lambda_{\min}$, (vii) the SINful estimator of Drton and Perlman (2007), and (viii) Graphical lasso (Friedman et al., 2008). Glasso was implemented using the huge package (Zhao et al., 2012) to optimize its tuning parameters, assum-



(a) False Discovery Rate (b) Power (c) Sensitivity Analysis

Figure 3: Top panels correspond to the path graph setting and bottom panels to the random graph setting. (a) FDR as a function of the sample size n for different levels of prior knowledge on the measurement error proportion, $\delta_{\max}$ and other methods (SINful, unadjusted for multiple testing, and Glasso). Rows correspond to the true measurement error proportion $\delta \in \{0, 0.2, 0.4\}$ and columns to p or (p, q) configurations in the path and random graphs accordingly. The red dashed line indicates the nominal level $\alpha = 0.05$. (b) Power for each of the configurations, defined as the proportion of edges in the CIG for X that are correctly estimated. (c) Sensitivity analysis for a fixed $p = 10$ graph: average minimal assumed measurement error proportion required to remove each edge versus its true magnitude of partial correlation. Color indicates the estimation frequency under the naive model. The true measurement error proportion, δ, is 0.25 and 0.3 for the path and random graphs, respectively.

ing no measurement error and without multiple hypothesis testing adjustment.

Performance is evaluated in terms of edge recovery. In particular, we report the empirical average false discovery rate $\frac{|\hat{E} \setminus E|}{\max\{|\hat{E}|, 1\}}$ where $\hat{E}$ denotes the estimated edge set and E the true edge set. We additionally report power, defined as the proportion of true edges in the CIG for X that are correctly recovered.

Our procedure with $\delta_{\max} = 0$ coincides with the SINful method up to a small difference in the test statistic, as expected, they exhibit almost identical performance. Thus, for conciseness, the results from $\delta_{\max} = 0$ are omitted in Figure 3a. As shown in the top panel of Fig. 3a, the methods achieve FDR control at the user-specified level $\alpha = 0.05$ when the true $\delta = 0$; i.e., there is no measurement error. However, when $\delta = 0.2, 0.4$, both methods fail to control the FDR, especially as δ and p grow larger. Similarly, the graphical lasso (Glasso) and the naive approach assuming no measurement error ($\delta_{\max} = 0$) and no multiple hypothesis testing adjustment (Unadjusted) fail to control the FDR in the presence of measurement error. For our proposed procedure, FDR control is maintained provided that the assumed upper bound on the measurement error proportion is at least as large as the true δ. When the assumed bound is smaller than the true value, the procedure fails to control FDR, reflecting model misspecification.

As expected, when the specified upper bound is large, the power to detect "true edges" decreases due to the non-identifiability. In the extreme case when no prior knowledge is available and the maximal admissible bound is used, the power deteriorates substantially (Fig. 3b, top panel). However, we see that even specifying a loose upper bound $\delta_{\max} = .4$ results in good power while maintaining FDR control. One counter-intuitive finding is that FDR is better controlled in dense graphs; this is likely because there are fewer non-edges to be falsely discovered.

To demonstrate the sensitivity analysis, we fix a precision

matrix with $p = 10$ and true measurement error proportion $\delta = 0.25$. For $n = 5000$ observations, we generate 500 replications. For each edge, we compute the minimal measurement error proportion that must be assumed for the edge to disappear from the estimated graph. We would expect that true edges with a larger magnitude of true partial correlations require large values of δ to be eliminated. The results are displayed in Fig. 3c, where each point corresponds to a potential edge, with color indicating the proportion of replications in which the edge is selected under the naive model ($\delta_{\max} = 0$). The horizontal axis shows the absolute value of the true partial correlation, while the vertical axis represents the average minimal proportion of variance that must be attributed to measurement error for the edge to be removed. Again, the top panel corresponds to a path graph and the bottom panel corresponds to an Erdös-Renyi graph drawn randomly (with edge inclusion probability of .3), but fixed across all 500 replications.

As expected, true edges (i.e., those with non-zero partial correlations) are frequently estimated under the naive model and also require a larger measurement error proportion to be eliminated (always above $\delta = .3$). There are also a number of false edges (i.e, those with zero partial correlations) that are still selected frequently when measurement error is ignored (indicated by yellow points). However, these edges are eliminated once a small proportion of measurement error is accounted for (always below $\delta = .2$). Moreover, the minimum error required to remove true edges is well separated from the error required to remove false edges.

4.1.1 Violation of Assumption 2

To empirically explore the impact of violating Assumption 2, we generate 500 independent datasets with $p = 10$, $n = 5,000$, and an edge inclusion probability $q = 0.3$. The true proportion attributed to the measurement error of each variable, denoted δ_j for $j \in [p]$, is drawn from $\mathcal{U}[0.1, d_{up}]$ for $d_{up} \in \{0.2, 0.5\}$. The datasets were analyzed under two setups: our proposed procedure utilizing a grid search up to $\delta_{\max} \in \{0.1, 0.2, 0.4\}$, and the SINful estimator of Drton and Perlman (2007). As shown in Table 1, even though the error variance proportions differ across variables (violating Assumption 2), the FDR remains empirically controlled at the nominal 0.05 level, provided that the user-specified upper bound $\delta_{\max}$ is greater than or equal to d_{up}.

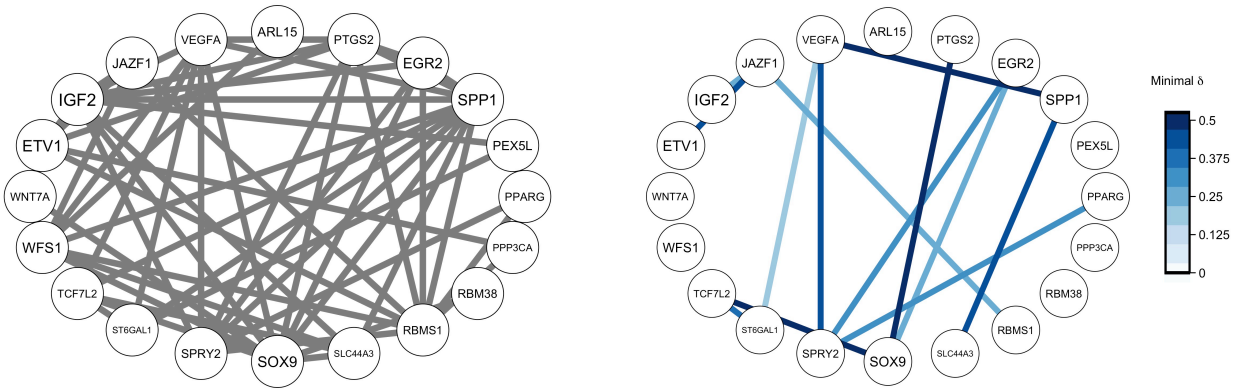
Table 1: Empirical FDR under violations of the equal error variance proportion assumption.

d_{up}	$\delta_{\max}$	FDR
0.2	0.1	0.017
0.2	0.2	0.001
0.2	0.4	0.000
0.2	SINful	0.110
0.5	0.1	0.118
0.5	0.2	0.061
0.5	0.4	0.002
0.5	SINful	0.188

4.2 GENE EXPRESSION DATA

We demonstrate the proposed method on a single-cell RNA sequencing dataset of 8569 human pancreatic cells (Baron et al., 2016) generated using the inDrop method (Klein et al., 2015). Gene expression data is typically zero-inflated due to limited mRNA capture efficiency and sampling variability inherent in capturing finite mRNA transcript counts (Klein et al., 2015). This introduces multiplicative noise contributing to the variability observed in the raw count data. Following the approach in (Baron et al., 2016), we normalized and log-transformed the count data, thereby justifying the application of our method, which assumes additive noise.

Our analysis focuses on estimating the undirected graph, i.e., gene co-expression network, for a set of 20 genes known to be involved in Type II Diabetes (Sharma et al., 2018). As shown in Fig. 4a, the graph estimated without accounting for measurement error is dense. Since the noise variance is unknown and no direct estimate is available, we conduct a sensitivity analysis to account for measurement error up to 50%. Fig. 4b visualizes the edges that are eliminated once measurement error is accounted for. The color intensity of each edge represents the minimal proportion of measurement error required for that edge to be removed from the estimated edge set. Lighter edges are eliminated when only a small amount of measurement error is assumed. For example, the edge between ST6GAL1 and VEGFA disappears when up to 20% of the observed variance is allowed to be noise. In contrast, darker edges require a larger proportion of measurement error to be removed. The edge between VEGFA and SPP1, for instance, is only removed when the assumed noise proportion reaches approximately 45%. Thus, darker edges correspond to conditional dependencies that are more robust to measurement error. By aggregating the graphs estimated at various levels of measurement error, a researcher can fix a specific threshold ($\delta_{\max}$) and obtain a single graph at that level of measurement error. We set 50% as a conservative upper bound on the error variance proportion since this implies that the variance of the measurement error does not exceed the variance of the true process. Edges shown in Fig. 4a but not Fig. 4b are robust under noise levels satisfying $\delta \leq 0.5$.



(a) Full gene co-expression network

(b) Removed edges

Figure 4: Estimated gene co-expression networks for 20 Type II Diabetes related genes from the human pancreatic single-cell RNA-seq dataset. (a) An undirected network estimated from the normalized and log-transformed expression data without accounting for measurement error. (b) Edges removed after incorporating measurement error through sensitivity analysis. Edge color indicates the minimal noise proportion (δ) required to eliminate the corresponding edge, with darker shades representing edges that are more robust to noise.

5 DISCUSSION

Our proposed methods address the challenge of estimating a GGM when the sample is contaminated with additive noise. Traditional graph estimation methods assume access to clean data and may produce many spurious edges when measurement error is present. On the other hand, previous methods that explicitly account for measurement error require replicates or exact knowledge of the measurement error covariance. We propose a viable alternative when such information is not available. Specifically, under a partial identification approach, we may control the false discovery rate of estimated edges given only an upper bound on the measurement error variance. Although a natural consequence of the weaker side information is reduced power to detect true edges, the weak assumptions allow the practitioner to be much more confident in the edges that are ultimately estimated. We also propose a sensitivity analysis framework that interpretably quantifies the minimum noise level needed to invalidate each estimated edge, offering practitioners a valuable tool to assess robustness.

Several interesting extensions could be addressed in future work. While our framework greatly relaxes the typical assumption that the covariance of the measurement error is known, it does require a simplifying assumption that measurement error variance is a fixed (but unknown) proportion of the observed variance. Generalizing the approach to allow varying proportions across variables so that $D_{jj} = \delta_j \Theta_{jj}$ and $\delta_j \leq \delta$ is theoretically straightforward; however, it presents computational challenges which would be interesting future work. Accommodating a more flexible class of measurement error models, including multiplicative measurement errors or dependent measurement errors, would also be an interesting extension. Another future direction would be extending the framework to accommodate high-dimensional settings, which would be valuable in many applications. Additionally, our current testing approach assumes the null hypothesis of "no edge." As noted by Malinsky (2024), a "cautious" approach to graph selection in which the null hypothesis posits an edge may be preferred in certain contexts. Extending our frameworks to this approach could be fruitful. Finally, extensions beyond the Gaussian framework, to accommodate non-Gaussian settings or discrete distributions, would be valuable and enable applicability across a wider range of fields.

Acknowledgements

This research was supported by a gift to the LinkedIn-Cornell Bowers Strategic Partnership (SM). This material is based upon work supported by the National Science Foundation under Award No. DMS-2542851 (YSW).

References

Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics. Wiley.

Armour, C., Fried, E. I., Deserno, M. K., Tsai, J., and Pietrzak, R. H. (2017). A network analysis of dsm-5 posttraumatic stress disorder symptoms and correlates in us military veterans. *Journal of anxiety disorders*, 45:49–59.

Banerjee, O., El Ghaoui, L., and d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *The Journal of Machine Learning Research*, 9:485–516.

Baron, M., Veres, A., Wolock, S. L., Faust, A. L., Gaujoux, R., Vetere, A., Ryu, J. H., Wagner, B. K., Shen-Orr, S. S., Klein, A. M., et al. (2016). A single-cell transcriptomic map of the human and mouse pancreas reveals inter-and intra-cell population structure. *Cell systems*, 3(4):346–360.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pages 1165–1188.

Blanchard, G. and Roquain, E. (2008). Two simple sufficient conditions for fdr control. *Electronic Journal of Statistics*, 2:963–992.

Brennecke, P., Anders, S., Kim, J. K., Kołodziejczyk, A. A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S. A., Marioni, J. C., et al. (2013). Accounting for technical noise in single-cell rna-seq experiments. *Nature methods*, 10(11):1093–1095.

Byrd, M., Nghiem, L. H., and McGee, M. (2021). Bayesian regularization of gaussian graphical models with measurement error. *Computational Statistics & Data Analysis*, 156:107085.

Cai, T., Liu, W., and Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: a modern perspective*. Chapman and Hall/CRC.

Casanellas, M., Garrote-López, M., and Zwiernik, P. (2024). Identifiability in robust estimation of tree structured models. *Bernoulli*, 30(1):1–21.

Chen, W., Drton, M., and Wang, Y. S. (2019). On causal discovery with an equal-variance assumption. *Biometrika*, 106(4):973–980.

Cinelli, C. and Hazlett, C. (2019). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(1):39–67.

Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., and Wynder, E. L. (1959). Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer institute*, 22(1):173–203.

Dai, H., Spirtes, P., and Zhang, K. (2022). Independence testing-based approach to causal discovery under measurement error and linear non-Gaussian models. *Advances in Neural Information Processing Systems*, 35:27524–27536.

Drton, M. and Perlman, M. D. (2007). Multiple testing and error control in gaussian graphical model selection. *Statistical Science*, 22(3):430–449.

Franks, A., DAmour, A., and Feller, A. (2020). Flexible sensitivity analysis for observational studies without observable implications. *Journal of the American Statistical Association*.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Fuller, W. A. (2009). *Measurement error models*. John Wiley & Sons.

Ghoshal, A. and Honorio, J. (2017). Learning identifiable Gaussian Bayesian networks in polynomial time and sample complexity. In *Advances in Neural Information Processing Systems*, pages 6457–6466.

Graybill, F. A. (1969). *Introduction to Matrices with Applications in Statistics*. Wadsworth Publishing Company.

Grün, D., Kester, L., and Van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. *Nature methods*, 11(6):637–640.

Imbens, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review*, 93(2):126–132.

Janková, J. and van de Geer, S. (2015). Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics*, 9(1).

Jin, Z., Pasquato, M., Davis, B. L., Deleu, T., Luo, Y., Cho, C., Lemos, P., Perreault-Levasseur, L., Bengio, Y., Kang, X., et al. (2025). Causal discovery in astrophysics: Unraveling supermassive black hole and galaxy coevolution. *The Astrophysical Journal*, 979(2):212.

Katiyar, A., Hoffmann, J., and Caramanis, C. (2019). Robust Estimation of Tree Structured Gaussian Graphical Models. In *Proceedings of the 36th International Conference on Machine Learning*, pages 3292–3300. PMLR. ISSN: 2640-3498.

Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D. A., and Kirschner, M. W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, 161(5):1187–1201.

Kline, B. and Tamer, E. (2023). Recent developments in partial identification. *Annual Review of Economics*, 15:125–150.

Koka, T., Machkour, J., and Muma, M. (2024). False discovery rate control for gaussian graphical models via neighborhood screening. In *2024 32nd European Signal Processing Conference (EUSIPCO)*, pages 2482–2486. IEEE.

Lauritzen, S. L. (1996). *Graphical models*, volume 17. Clarendon Press.

Lee, S., Sobczyk, P., and Bogdan, M. (2019). Structure learning of gaussian markov random fields with false discovery rate control. *Symmetry*, 11(10):1311.

Li, J. and Maathuis, M. H. (2021). GGM knockoff filter: False discovery rate control for gaussian graphical models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83(3):534–558.

Liu, W. (2013). Gaussian graphical model estimation with false discovery rate control. *The Annals of Statistics*, 41(6):2948–2978.

Loh, P.-L. and Bühlmann, P. (2014). High-dimensional learning of linear causal networks via inverse covariance estimation. *Journal of Machine Learning Research*, 15:3065–3105.

Malinsky, D. (2024). A cautious approach to constraint-based causal model selection. *arXiv preprint arXiv:2404.18232*.

Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., Allison, K. R., Kellis, M., Collins, J. J., et al. (2012). Wisdom of crowds for robust gene network inference. *Nature methods*, 9(8):796–804.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the Lasso. *The Annals of Statistics*, 34(3):1436 – 1462.

Nikolakakis, K. E., Kalogerias, D. S., and Sarwate, A. D. (2019). Learning tree structures from noisy data. In Chaudhuri, K. and Sugiyama, M., editors, *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 1771–1782. PMLR.

Olkin, I. (1976). Asymptotic distribution of functions of a correlation matrix. *Essays in provability and statistics: A volume in honor of Professor Junjiro Ogawa*.

Ortiz, A., Munilla, J., Álvarez-Illán, I., Górriz, J. M., Ramírez, J., and Initiative, A. D. N. (2015). Exploratory graphical models of functional and structural connectivity patterns for alzheimer's disease diagnosis. *Frontiers in computational neuroscience*, 9:132.

Peters, J., Mooij, J. M., Janzing, D., and Schölkopf, B. (2014). Causal discovery with continuous additive noise models. *Journal of Machine Learning Research*, 15:2009–2053.

Rosenbaum, P. R. and Rubin, D. B. (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Methodological)*, 45(2):212–218.

Roverato, A. and Whittaker, J. (1998). The isserlis matrix and its application to non-decomposable graphical gaussian models. *Biometrika*, 85(3):711–725.

Saeed, B., Belyaeva, A., Wang, Y., and Uhler, C. (2020a). Anchored Causal Inference in the Presence of Measurement Error. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 619–628. PMLR. ISSN: 2640-3498.

Saeed, B., Belyaeva, A., Wang, Y., and Uhler, C. (2020b). Anchored causal inference in the presence of measurement error. In *Conference on uncertainty in artificial intelligence*, pages 619–628. PMLR.

Sharma, A., Halu, A., Decano, J. L., Padi, M., Liu, Y.-Y., Prasad, R. B., Fadista, J., Santolini, M., Menche, J., Weiss, S. T., et al. (2018). Controllability in an islet specific regulatory network identifies the transcriptional factor nfatc4, which regulates type 2 diabetes associated genes. *NPJ systems biology and applications*, 4(1):25.

Shi, W., Ghosal, S., and Martin, R. (2021). Bayesian estimation of sparse precision matrices in the presence of gaussian measurement error. *Electronic Journal of Statistics*, 15(2):4545–4579.

Tandon, A., Han, A., and Tan, V. (2021). Sga: A robust algorithm for partial recovery of tree-structured graphical models with noisy samples. In Meila, M. and Zhang, T., editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 10107–10117. PMLR.

Uhler, C. (2018). Gaussian graphical models. In *Handbook of graphical models*, pages 217–238. CRC Press.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge University Press.

VanderWeele, T. J. and Ding, P. (2017). Sensitivity analysis in observational research: introducing the e-value. *Annals of internal medicine*, 167(4):268–274.

Yang, Y., Ghassami, A., Nafea, M., Kiyavash, N., Zhang, K., and Shpitser, I. (2022). Causal discovery in linear latent variable models subject to measurement error. *Advances in Neural Information Processing Systems*, 35:874–886.

Yu, L., Kaufmann, T., and Lederer, J. (2021). False discovery rates in biological networks. In *International Conference on Artificial Intelligence and Statistics*, pages 163–171. PMLR.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35.

Zhang, H., Chen, K., Lin, N., Yang, A., Hao, Z., and Chen, Z. (2025). Conditional independent test in the presence of measurement error with causal structure learning. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pages 9112–9120.

Zhang, K., Gong, M., Ramsey, J. D., Batmanghelich, K., Spirtes, P., and Glymour, C. (2018). Causal discovery with linear non-gaussian models under measurement error: Structural identifiability results. In *UAI*, pages 1063–1072.

Zhao, T., Liu, H., Roeder, K., Lafferty, J., and Wasserman, L. (2012). The huge package for high-dimensional undirected graph estimation in r. *The Journal of Machine Learning Research*, 13(1):1059–1062.

Robust estimation of graphical models with measurement error: False discovery control and sensitivity analysis (Supplementary Material)

Shira Mingelgrin[1] **Y. Samuel Wang**[1]

[1]Department of Statistics and Data Science, Cornell University, Ithaca, New York, USA

A PROOFS

A.1 PROOF OF LEMMA 1

Proof. Let $\hat{\Theta}$ denote the sample covariance matrix of a random vector $W \in \mathbb{R}^p$ with mean zero and covariance Θ. Further, let $\mathrm{vech}(\cdot)$ denote the half-vectorization form, an $m \times (m+1)/2$ column vector obtained by vectorizing only the lower triangular part of a symmetric $m \times m$ matrix, then as shown by (Olkin, 1976),

$$\sqrt{n-1}\big(\mathrm{vech}(\hat{\Theta}) - \mathrm{vech}(\Theta)\big) \xrightarrow{d} \mathcal{N}\big(0, \mathrm{Iss}(\Theta)\big).$$

Define the diagonally adjusted covariance matrix as a function of the noise proportion $d \in [0, 1)$,

$$S_d = \Theta - d\,\Theta_D, \qquad \hat{S}_d = \hat{\Theta} - d\,\hat{\Theta}_D,$$

where Θ_D and $\hat{\Theta}_D$ denote the diagonal parts of Θ and $\hat{\Theta}$, respectively. Then

$$\mathrm{vech}(S_d) = M\,\mathrm{vech}(\Theta), \qquad \mathrm{vech}(\hat{S}_d) = M\,\mathrm{vech}(\hat{\Theta}),$$

where M is the diagonal $p(p+1)/2 \times p(p+1)/2$ matrix with entries $M_{ij,ij} = \begin{cases} 1, & i < j \\ 1-d, & i = j \end{cases}$. Consequently,

$$\sqrt{n-1}\big(\mathrm{vech}(\hat{S}_d) - \mathrm{vech}(S_d)\big) \xrightarrow{d} \mathcal{N}\big(0, M\mathrm{Iss}(\Theta)M\big).$$

Consider the matrix inversion mapping $h(S_d) = \mathrm{vech}(S_d^{-1})$, which is continuously differentiable on the space of positive definite matrices. We constrain the values of d such that $\hat{S}_d$ is positive definite, and apply the delta method (Van der Vaart, 2000). The Jacobian of h at S_d is a $p(p+1)/2$ by $p(p+1)/2$ matrix with entries

$$\frac{\partial S_d^{-1}}{\partial (S_d)_{pq}} = -S_d^{-1}\Delta_{pq}S_d^{-1},$$

where Δ_{pq} is the symmetric matrix with ones in positions (p, q) and (q, p) and zeros elsewhere (Graybill, 1969). As shown in Roverato and Whittaker (1998), the Jacobian can be expressed in terms of the Isserlis matrix as

$$\nabla h(S_d) = -\,\mathrm{Iss}(S_d^{-1})\,\mathrm{Iss}(I)^{-1}.$$

Applying the delta method yields

$$\sqrt{n-1}\big(h(\hat{S}_d) - h(S_d)\big) \xrightarrow{d} \mathcal{N}\big(0, V_S\big),$$

with

$$V_S = \mathrm{Iss}(S_d^{-1})\mathrm{Iss}(I)^{-1}M\mathrm{Iss}(\Theta)M\mathrm{Iss}(I)^{-1}\mathrm{Iss}(S_d^{-1}),$$

□

A.2 PROOF OF THEOREM 1

Proof. We follow similar steps as the derivation of the Fisher's $z-$ transformation in Anderson (2003). Let S_d denote the population adjusted covariance matrix and $\hat{S}_d$ its sample analogue. By Lemma 1,

$$\sqrt{n-1}\big(\hat{S}_d^{-1} - S_d^{-1}\big) \xrightarrow{d} \mathcal{N}(0, V_S)$$

where V_S is given in Lemma 1. Define, for $i \leq j \in V$,

$$u_{ij} = \frac{(\hat{S}_d^{-1})_{ij}}{\sqrt{(S_d^{-1})_{ii}(S_d^{-1})_{jj}}}, \qquad \mathrm{p}_{ij} = \frac{(S_d^{-1})_{ij}}{\sqrt{(S_d^{-1})_{ii}(S_d^{-1})_{jj}}}$$

Since the denominators are deterministic, u_{ij} are asymptotically jointly normal with means $\mathrm{p}_{ij} = \frac{(S_d^{-1})_{ij}}{\sqrt{(S_d^{-1})_{ii}(S_d^{-1})_{jj}}}$ and covariances given by $\mathrm{cov}(u_{ij}, u_{kl}) = \frac{1}{\sqrt{(S_d^{-1})_{ii}(S_d^{-1})_{jj}}} \frac{1}{\sqrt{(S_d^{-1})_{kk}(S_d^{-1})_{ll}}} (V_S)_{ij,kl}$, for $i \leq j, k \leq l \in V$.

The true partial correlation between variables i and j given $V \setminus \{i, j\}$ is

$$\rho_{ij \cdot V \setminus \{i,j\}} = \frac{\mathrm{p}_{ij}}{\sqrt{\mathrm{p}_{ii}\mathrm{p}_{jj}}}$$

and the sample partial correlation is

$$\hat{\rho}_{ij \cdot V \setminus \{i,j\}} = \frac{u_{ij}}{\sqrt{u_{ii}u_{jj}}}$$

Define the mapping $f : \mathbb{R}^3 \to \mathbb{R}$ by

$$f(u_{ii}, u_{jj}, u_{ij}) = \frac{u_{ij}}{\sqrt{u_{ii}u_{jj}}}$$

Applying the delta method to $\boldsymbol{u} = (u_{ii}, u_{jj}, u_{ij})^\top$ yields

$$\sqrt{n}\big(\hat{\rho}_{ij \cdot V \setminus \{i,j\}} - \rho_{ij \cdot V \setminus \{i,j\}}\big) \xrightarrow{d} \mathcal{N}(0, \sigma_{ij}^2),$$

where $\sigma_{ij}^2 = \nabla f(\mathbf{p})^\top V_{ij} \nabla f(\mathbf{p})$ and V_{ij} is the covariance matrix of (u_{ii}, u_{jj}, u_{ij}). The gradient of f with respect to $\boldsymbol{u}$ is given by

$$\begin{aligned}
\frac{\partial f}{\boldsymbol{u}_1} &= -\frac{\boldsymbol{u}_3}{2\boldsymbol{u}_1\sqrt{\boldsymbol{u}_1\boldsymbol{u}_2}} \\
\frac{\partial f}{\boldsymbol{u}_2} &= -\frac{\boldsymbol{u}_3}{2\boldsymbol{u}_2\sqrt{\boldsymbol{u}_1\boldsymbol{u}_2}} \\
\frac{\partial f}{\boldsymbol{u}_3} &= \frac{1}{\sqrt{\boldsymbol{u}_1\boldsymbol{u}_2}}
\end{aligned}$$

which under the null hypothesis $\rho_{ij \cdot V \setminus \{i,j\}} = 0$, simplifies to

$$\nabla f(\mathbf{p}) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

since by definition $\mathrm{p}_{ii} = 1$, $\mathrm{p}_{jj} = 1$ and $\rho_{ij \cdot V \setminus \{i,j\}} = 0$ implies that $\mathrm{p}_{ij} = 0$. Hence

$$\sigma_{ij}^2 = \frac{(V_S)_{ij,ij}}{(S_d^{-1})_{ii}(S_d^{-1})_{jj}}$$

Using the degrees-of-freedom correction for the partial correlations $n - 1 \mapsto n - p + 1$ since we condition on $(p - 2)$ variables (Anderson, 2003), we obtain

$$\sqrt{n-p+1}\,\frac{\hat{\rho}_{ij \cdot V \setminus \{i,j\}}}{\sqrt{\hat{\sigma}_{ij}^2}} \xrightarrow{d} \mathcal{N}(0, 1).$$

□

A.3 PROOF OF THEOREM 2

Proof. By Theorem 1, $T_{u,v}(d = \delta)$ is asymptotically $\mathcal{N}(0,1)$ so that $\pi_{u,v}(d)$ is asymptotically $\mathcal{U}(0,1)$. Since, $\pi^{\star}_{u,v}$ is calculated using $\min_{d\in[0,\delta_{\max}]} T_{u,v}(d)$, we have that

$$\pi^{\star}_{u,v} \geq \pi_{u,v}(d).$$

Thus, for any $t \in (0,1)$,

$$\lim_{n\to\infty} P(\pi^{\star}_{u,v} \leq t) \leq \lim_{n\to\infty} P(\pi_{u,v}(d) \leq t) \leq t,$$

so that $\pi^{\star}_{u,v}$ is stochastically lower bounded by a uniform distribution. Thus, the Benjamini and Yekutieli (2001) applied to $\pi^{\star}_{u,v}$ controls the false discovery rate; see e.g. Blanchard and Roquain (2008).

□

B EXAMPLE FOR FIGURE 1

Consider the following numerical example corresponding to the CIG in Fig. 1, where the true latent variables follow the path $X_1 - X_2 - X_3$. Suppose the precision matrix of the latent variables X is given by:

$$\Omega \coloneqq \begin{pmatrix} 1 & 0.8 & 0 \\ 0.8 & 1 & 0.3 \\ 0 & 0.3 & 1 \end{pmatrix}, \tag{8}$$

where $\Omega_{1,3} = 0$ indicating the absence of an edge between X_1 and X_3. Furthermore, suppose the true measurement error level is $\delta = .2$. If δ was known, we could use the population covariance of the measurement errors D to identify the population covariance of the latent process of interest as $\Sigma = \Theta - D$. Subsequently we could identify the partial correlations $\rho_{u,v\cdot V} = -\Omega_{u,v}/\sqrt{\Omega_{u,u}\Omega_{v,v}}$.

When D is unknown but constrained to a specific set, different D (with the same fixed Θ) will yield different partial correlations. For instance, if δ is only known to lie in the interval $(0, 0.3)$, meaning prior knowledge provides an upper bound $\delta_{\max} = 0.3$, the partial identification set for the partial correlation $\rho_{1,3\cdot\{2\}}$ is $[-0.29, 0.1]$. Since the presence or absence of an edge in the graph is equivalent to the partial correlation being non-zero or zero, respectively, an identification set containing both zero and non-zero values implies that the edge between X_1 and X_3 (and consequently the graph) is only partially identifiable.