Supplementary Material: Anchored Causal Inference in the Presence of Measurement Error

Basil Saeed MIT

Anastasiya Belyaeva MIT

Yuhao Wang MIT

Caroline Uhler MIT

A Proof of Lemma 1

Proof. By the strong law of large numbers, for all $i, j \in [p]$, it holds that

$$\mathbb{E}[\hat{\eta}_i] \xrightarrow{a.s.} \mathbb{E}[\eta_i] \quad \text{and} \quad \mathbb{E}[\hat{\eta}_{ij}] \xrightarrow{a.s.} \mathbb{E}[\eta_{ij}] \quad \text{as} \quad n \to \infty.$$

The functions g are continuous since they are continuously differentiable by (A2); therefore, for all $i \in [p]$, it holds that

$$\hat{\mu}_i = g_i(\mathbb{E}[\hat{\eta}_i]) \xrightarrow{a.s.} g_i(\mathbb{E}[\eta_i]) = \mu_i \quad \text{as} \quad n \to \infty;$$

similarly, for all $i, j \in [p]$, it holds that

$$\hat{\mu}_{ij} = g_{ij}(\mathbb{E}[\hat{\eta}_{ij}]) \xrightarrow{a.s.} g_{ij}(\mathbb{E}[\eta_{ij}]) = \mu_{ij} \quad \text{as} \quad n \to \infty.$$

Therefore,

$$(\hat{\Sigma}_{ij}) := \hat{\mu}_{ij} - \hat{\mu}_i\hat{\mu}_j \xrightarrow{a.s.} \mu_{ij} - \mu_i\mu_j = (\Sigma)_{ij} \quad \text{as} \quad n \to \infty$$

and the correlations

$$\hat{\rho}_{ij} = \frac{(\hat{\Sigma})_{ij}}{\sqrt{(\hat{\Sigma})_{ii}(\hat{\Sigma})_{jj}}} \xrightarrow{a.s.} \frac{(\Sigma)_{ij}}{\sqrt{(\Sigma)_{ii}(\Sigma)_{jj}}} = \rho_{ij} \quad \text{as} \quad n \to \infty.$$

Recursively applying a similar argument to equation (2) proves that $\hat{\rho}_{ij \cdot K}$ is consistent, thereby completing the proof. □

B Proof of Theorem 1

We start by defining the vectors of all correlations estimated from Algorithm 1 and all true correlations of Z as

$$\hat{\rho} := \begin{pmatrix} \hat{\rho}_{12} \\ \hat{\rho}_{13} \\ \cdot \\ \cdot \\ \cdot \\ \hat{\rho}_{(p-1)p} \end{pmatrix} \quad \text{and} \quad \rho := \begin{pmatrix} \rho_{12} \\ \rho_{13} \\ \cdot \\ \cdot \\ \cdot \\ \rho_{(p-1)p} \end{pmatrix}, \tag{S.1}$$

respectively. We use η to denote the vector obtained from concatenating all monomials in X_i and X_j that appear in η_i and η_{ij} for $i, j \in [p]$ in Assumption (A2), i.e.,

$$\eta := \begin{pmatrix} \eta_1^T & \eta_2^T & \dots & \eta_p^T & \dots & \eta_{11}^T & \eta_{12}^T & \eta_{pp}^T \end{pmatrix}.$$

We let

$$\hat{\eta} := \begin{pmatrix} \hat{\eta}_1^T & \hat{\eta}_2^T & \dots & \hat{\eta}_p^T & \dots & \hat{\eta}_{11}^T & \hat{\eta}_{12}^T & \hat{\eta}_{pp}^T \end{pmatrix}.$$

be the analogous concatenated vector with $\hat{\eta}_i$ and $\hat{\eta}_{ij}$ for $i, j \in [p]$ of sample monomials in $\hat{X}_i$ and $\hat{X}_j$ calculated from the data $\hat{X} = (\hat{X}^{(1)}, \hat{X}^{(2)}, \dots, \hat{X}^{(n)})$.

The following lemma is concerned with the asymptotic distribution of the correlation vector $\hat{\rho}$.

Lemma S.1. *Under Assumptions (A1) and (A2),*

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{D} \mathcal{N}_{|\rho|}\Big(0, A(\nu)\Big),$$

where ν is the vector of all first and second order moments of η and A is a continuous function of ν.

Proof. Assumption (A2) asserts that the covariance of η is finite. Hence, we can apply the Central Limit Theorem to obtain

$$\sqrt{n}(\mathbb{E}[\hat{\eta}] - \mathbb{E}[\eta]) \xrightarrow{D} \mathcal{N}_{|\eta|}(0, A_\eta(\nu)), \tag{S.2}$$

where A_η is the covariance matrix of η. The elements of the covariance matrix A_η can be written as a continuous function of the first-and second-order moments of η, i.e., they can be written as a continuous function of ν.

Assumptions (A1) and (A2) imply that we can write for all $i, j \in [p]$,

$$\rho_{ij} = \frac{g_{ij}(\mathbb{E}[\eta_{ij}]) - g_i(\mathbb{E}[\eta_i])g_j(\mathbb{E}[\eta_j])}{\sqrt{g_{ii}(\mathbb{E}[\eta_{ii}]) - g_i(\mathbb{E}[\eta_i])^2}\sqrt{g_{jj}(\mathbb{E}[\eta_{jj}]) - g_j(\mathbb{E}[\eta_j])^2}}. \tag{S.3}$$

We compute sample correlations $\hat{\rho}_{ij}$ in our algorithm as

$$\hat{\rho}_{ij} = \frac{g_{ij}(\mathbb{E}[\hat{\eta}_{ij}]) - g_i(\mathbb{E}[\hat{\eta}_i])g_j(\mathbb{E}[\hat{\eta}_j])}{\sqrt{g_{ii}(\mathbb{E}[\hat{\eta}_{ii}]) - g_i(\mathbb{E}[\hat{\eta}_i])^2}\sqrt{g_{jj}(\mathbb{E}[\hat{\eta}_{jj}]) - g_j(\mathbb{E}[\hat{\eta}_j])^2}}.$$

Based on equation (S.3) we can define a function $w : \mathbb{R}^{|\eta|} \to \mathbb{R}^{|\rho|}$ such that $w(\mathbb{E}[\eta]) = \rho$ and $w(\mathbb{E}[\hat{\eta}]) = \hat{\rho}$. Applying the Delta method to equation (S.2) with the function w, we get

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{D} \mathcal{N}_{|\rho|}\Big(0, A_\rho(\nu)\Big),$$

where $A_\rho(\nu) = \nabla w(\mathbb{E}[\eta])^T A_\eta(\nu) \nabla w(\mathbb{E}[\eta])$, since the elements of the mean vector $\mathbb{E}[\eta]$ are elements of ν. Notice that under the assumption that the variance of $X_i = F_i(Z_i)$ is non-zero, which we mention in Section 2, the denominator in (S.3) is non-zero, and therefore ρ is continuously differentiable in $g(\mathbb{E}[\eta])$, which is continuously differentiable in $\mathbb{E}[\eta]$ by Assumption (A2). Hence, $\nabla w(\mathbb{E}[\eta])$ is continuous in $\mathbb{E}[\eta]$, and therefore continuous in ν. Since $A_\rho(\nu)$ is a matrix product of functions continuous in ν, it is also continuous in ν, which completes the proof. $\square$

Lemma S.2. *If*

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{D} \mathcal{N}_{|\rho|}\Big(0, A_\rho(\nu)\Big),$$

where ν is the vector of all first- and second-order moments of η, and $A_\rho(\nu)$ is continuous in ν, then under Assumptions (A1) and (A2), for any $i, j \in [p]$ and $K \subseteq [p] \setminus \{i, j\}$,

$$\sqrt{n}(\hat{\rho}_{ij\cdot K} - \rho_{ij\cdot K}) \xrightarrow{D} \mathcal{N}_1(0, \tau_{ij\cdot K}(\nu)),$$

for some $\tau_{ij\cdot K}$ that is continuous in ν, where $\hat{\rho}_{ij\cdot K}$ are the partial correlations estimated by Algorithm 1.

Proof. We take any arbitrary but fixed $i, j \in [p]$ and subset $K \in [p]$, and we prove the lemma for $\hat{\rho}_{ij\cdot K}$. Let $k := |K| + 2$, where $|K|$ is the size of the conditioning set K. We begin by relabeling the variables of interest for clarity. We relabel i to 1, j to 2 and the elements of K to $S = \{3, \cdots, k\}$. Furthermore, we define the sets

$$S_m := \begin{cases} \{m, m+1, \cdots, k\} & 3 \le m \le k \\ \emptyset & m = k+1 \end{cases}.$$

Note that $S_3 = S$, and thereby, the partial correlation of interest is $\rho_{12\cdot S} = \rho_{12\cdot S_3}$. Now we define for $m \in \{3, \cdots, k+1\}$, the vectors

$$\hat{\rho}_m := \begin{pmatrix} \hat{\rho}_{1,2\cdot S_m} \\ \hat{\rho}_{1,3\cdot S_m} \\ \hat{\rho}_{2,3\cdot S_m} \\ \hat{\rho}_{1,4\cdot S_m} \\ \hat{\rho}_{2,4\cdot S_m} \\ \cdots \\ \hat{\rho}_{1,m-1\cdot S_m} \\ \hat{\rho}_{2,m-1\cdot S_m} \\ \cdots \\ \hat{\rho}_{m-2,m-1\cdot S_m} \end{pmatrix} \quad \text{and} \quad \rho_m := \begin{pmatrix} \rho_{1,2\cdot S_m} \\ \rho_{1,3\cdot S_m} \\ \rho_{2,3\cdot S_m} \\ \rho_{1,4\cdot S_m} \\ \rho_{2,4\cdot S_m} \\ \cdots \\ \rho_{1,m-1\cdot S_m} \\ \rho_{2,m-1\cdot S_m} \\ \cdots \\ \rho_{m-2,m-1\cdot S_m} \end{pmatrix}.$$

It follows from the definition of S_{k+1} that $\hat{\rho}_{k+1} = \hat{\rho}$ and $\rho_{k+1} = \rho$. In order to prove the lemma, we proceed by induction on m starting with the base case of $m = k+1$ and showing that for all m such that $3 \leq m \leq k$,

$$\sqrt{n}(\hat{\rho}_m - \rho_m) \xrightarrow{D} \mathcal{N}_{|\rho_m|}\Big(0, A_m(\nu)\Big) \tag{S.4}$$

for some A_m that is continuous in ν. Note that the base case is given by the hypothesis in the lemma. Moreover, the statement of the lemma is that the above holds for $m = 3$, and therefore, completing the inductive step proves the lemma.

To complete the inductive step, assume that for m such that $3 \leq m < k+1$, we have

$$\sqrt{n}(\hat{\rho}_{m+1} - \rho_{m+1}) \xrightarrow{D} \mathcal{N}_{|\rho_{m+1}|}\Big(0, A_{m+1}(\nu)\Big). \tag{S.5}$$

Note that for any $\alpha, \beta \in [p]$, the recursive formula for the partial correlations

$$\rho_{\alpha\beta\cdot S_m} = \frac{\rho_{\alpha\beta\cdot S_{m+1}} - \rho_{\alpha m\cdot S_{m+1}}\rho_{\beta m\cdot S_{m+1}}}{\sqrt{1-\rho^2_{\alpha m\cdot S_{m+1}}}\sqrt{1-\rho^2_{\beta m\cdot S_{m+1}}}} \tag{S.6}$$

implies that the vector ρ_m can be written as a function of ρ_{m+1}. Let $f_m : \mathbb{R}^{|\rho_{m+1}|} \to \mathbb{R}^{|\rho_m|}$ be this function, then we have

$$f_m(\rho_{m+1}) = f_m\left(\begin{pmatrix} \rho_{1,2\cdot S_{m+1}} \\ \rho_{1,3\cdot S_{m+1}} \\ \rho_{2,3\cdot S_{m+1}} \\ \rho_{1,4\cdot S_{m+1}} \\ \rho_{2,4\cdot S_{m+1}} \\ \cdots \\ \rho_{1,m\cdot S_{m+1}} \\ \rho_{2,m\cdot S_{m+1}} \\ \cdots \\ \rho_{m-1,m\cdot S_{m+1}} \end{pmatrix}\right) = \begin{pmatrix} \rho_{1,2\cdot S_m} \\ \rho_{1,3\cdot S_m} \\ \rho_{2,3\cdot S_m} \\ \rho_{1,4\cdot S_m} \\ \rho_{2,4\cdot S_m} \\ \cdots \\ \rho_{1,(m-1)\cdot S_m} \\ \rho_{2,(m-1)\cdot S_m} \\ \cdots \\ \rho_{m-2,(m-1)\cdot S_m} \end{pmatrix} = \rho_m.$$

Note that this implies $f_m(\hat{\rho}_{m+1}) = \hat{\rho}_m$ since our procedure uses this recursive formula to estimate the partial correlations. Applying the Delta method to (S.5) with the function f_m gives

$$\sqrt{n}(\hat{\rho}_m - \rho_m) \xrightarrow{D} \mathcal{N}_{|\rho_m|}\Big(0, A_m(\nu, \rho_{m+1})\Big),$$

where $A_m(\nu, \rho_{m+1}) := \nabla f_m(\rho_{m+1}) A_{m+1}(\nu) \nabla f_m(\rho_{m+1})^T$. The matrix $\nabla f_m(\rho_{m+1})$ can be computed to be the following matrix

$$D := \nabla f_m(\rho_{m+1}) = \begin{pmatrix} a_{12} & 0 & 0 & \cdot & 0 & b_{12,1} & b_{12,2} & 0 & 0 & \cdot & 0 \\ 0 & a_{13} & 0 & \cdot & 0 & b_{13,1} & 0 & b_{13,3} & 0 & \cdot & 0 \\ 0 & 0 & a_{23} & \cdot & 0 & 0 & b_{23,2} & 0 & b_{23,3} & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & a_{\frac{(m-1)(m)}{2}} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix},$$

where

$$a_{xy} = \frac{1}{\sqrt{1-\rho^2_{x,m\cdot S_{m+1}}}\sqrt{1-\rho^2_{y,m\cdot S_{m+1}}}},$$

$$b_{xy,x} = \frac{\rho_{x,y\cdot S_{m+1}}\rho_{x,m\cdot S_{m+1}} - \rho_{y,m\cdot S_{m+1}}}{\sqrt{(1-\rho^2_{x,m\cdot S_{m+1}})^3}\sqrt{1-\rho^2_{y,m\cdot S_{m+1}}}},$$

and

$$b_{xy,y} = \frac{\rho_{x,y\cdot S_{m+1}}\rho_{y,m\cdot S_{m+1}} - \rho_{x,m\cdot S_{m+1}}}{\sqrt{1-\rho^2_{x,m\cdot S_{m+1}}}\sqrt{(1-\rho^2_{y,m\cdot S_{m+1}})^3}}.$$

To simplify indexing, we define the index function

$$I(x,y) = x + \frac{(y-2)(y-1)}{2}.$$

Then, the element $a_{x,y}$ will be on the $I(x,y)^{th}$ row and column of the Jacobian D. We can now compute the elements of the matrix A_m in terms of the elements of A_{m+1}. Namely, defining $d := \frac{(m-2)(m-1)}{2}$ and using the notation $M_{[x,y]}$ to denote the entry in the x^{th} row and y^{th} column of M, we can compute the element in the $I(x,y)^{th}$ row and $I(z,w)^{th}$ column of A_m to be

$$\begin{aligned}
A_{m[I(x,y),I(z,w)]} &= \sum_{p=1}^{I(1,m)} \sum_{q=1}^{I(1,m)} D_{m[I(x,y),p]} A_{m+1[p,q]} D^T_{m[q,I(z,w)]} \\
&= \sum_{p=1}^{I(1,m)} D_{m[I(x,y),p]} \sum_{q=1}^{I(1,m)} A_{m+1[p,q]} D_{m[I(z,w),q]} \\
&= \sum_{p=1}^{I(1,m)} D_{m[I(x,y),p]} \Big(a_{z,w} A_{m+1[p,I(z,w)]} + b_{zw,z} A_{m+1[p,d+z]} + b_{zw,w} A_{m+1[p,d+w]} \Big) \\
&= a_{x,y} \big(a_{z,w} A_{m+1[I(x,y),I(z,w)]} + b_{zw,z} A_{m+1[I(x,y),d+z]} + b_{zw,w} A_{m+1[I(x,y),d+w]} \big) \\
&\quad + b_{xy,x} \big(a_{z,w} A_{m+1[d+x,I(z,w)]} + b_{zw,z} A_{m+1[d+x,d+z]} + b_{zw,w} A_{m+1[d+x,d+w]} \big) \\
&\quad + b_{xy,y} \big(a_{d+z,d+w} A_{m+1[d+y,I(z,w)]} + b_{zw,z} A_{m+1[d+y,d+z]} + b_{zw,w} A_{m+1[d+y,d+w]} \big).
\end{aligned} \tag{S.7}$$

Note that equation (S.6) shows that ρ_{m+1} is a continuously differentiable function of ρ_{m+2} since it is a composition of continuously differentiable functions. Hence, D is continuous in ρ_{m+2}. Furthermore, ρ_{m+2} is continuous in ρ, which can be seen by applying a similar argument recursively. Therefore, ρ_{m+1} is continuous in ρ. By Assumption (A2), for all $i, j \in [p]$,

$$\rho_{ij} = \frac{g_{ij}(\mathbb{E}[\eta_{ij}]) - g_i(\mathbb{E}[\eta_i]) g_j(\mathbb{E}[\eta_j])}{\sqrt{g_{ii}(\mathbb{E}[\eta_{ii}]) - g_i(\mathbb{E}[\eta_i])^2} \sqrt{g_{jj}(\mathbb{E}[\eta_{jj}]) - g_j(\mathbb{E}[\eta_j])^2}}. \tag{S.8}$$

Hence, ρ is continuous in ν. Therefore, ρ_{m+1} is continuous in ν. A similar argument shows that each of $a_{xy}, b_{xy,x}$ and $b_{xy,y}$ are continuous in ν. Finally, the $I(x,y), I(z,w)$ entry of A_m is continuous in the elements of $A_{m+1}(\nu)$ for arbitrary x, y, z and w. Therefore, we can reparameterize $A_m(\nu, \rho_{m+1}) = \tilde{A}_m(\nu)$, and the inductive step follows for $m \in \{3, \cdots, k\}$. Specifically, for $m = 3$, we have the desired statement:

$$\sqrt{n}(\hat{\rho}_3 - \rho_3) \xrightarrow{D} \mathcal{N}_1(0, \tilde{A}_3(\nu))$$

for $\tilde{A}_3(\nu)$ continuously differentiable in ν. Relabeling back to i, j and K, and defining $\tau_{ij\cdot K}(\nu) := \tilde{A}_3(\nu)$, we have

$$\sqrt{n}(\hat{\rho}_{ij\cdot K} - \rho_{ij\cdot K}) \xrightarrow{D} \mathcal{N}_1(0, \tau_{ij\cdot K}(\nu)),$$

which completes the proof. □

Note that it was not necessary to find the form of the elements of A_m explicitly to argue that it was continuous. However, the proof of this lemma gives us a recursive formula (S.7) to compute the elements of A_3. Furthermore, this recursive formula is independent of the choice of noise functions F and the associated functions g. Hence, this recursion can be used for all noise models, as long as the base case is derived for that noise model, i.e., as long as the elements of the matrix $A(\nu)$ in Lemma S.1 can be found.

Proof of Theorem 1. Follows directly from combining Lemma S.1 and Lemma S.2. □

C Proof of Corollary 1

Proof. By Theorem 1 and the hypothesis of the corollary,

$$\sqrt{n}(\hat{\rho}_{ij\cdot K} - \rho_{ij\cdot K}) \xrightarrow{D} \mathcal{N}_1\big(0, \tilde{\tau}_{ij\cdot K}(\rho_{ij\cdot K})\big).$$

By application of the Delta method with

$$z_{ij\cdot K}(\rho) = \int \frac{1}{\sqrt{\tilde{\tau}_{ij\cdot K}(\rho)}} d\rho + C,$$

we obtain

$$\sqrt{n}(z_{ij\cdot K}(\hat{\rho}_{ij\cdot K}) - z_{ij\cdot K}(\rho_{ij\cdot K})) \xrightarrow{D} \mathcal{N}_1\left(0, \left(z'_{ij\cdot K}(\rho)\right)^2 \tilde{\tau}_{ij\cdot K}(\rho_{ij\cdot K})\right) = \mathcal{N}_1(0,1).$$

Note that the condition imposed on C, $z_{ij\cdot K}(0) = 0$, by the corollary is not required to prove the result, but is only needed to prove Theorem 2. □

D Proof of Corollary 2

Proof. By the Law of Large Numbers, $\hat{\nu} \xrightarrow{a.s.} \nu$ as $n \to \infty$. Therefore $\tau_{ij\cdot K}(\hat{\nu}) \xrightarrow{a.s.} \tau_{ij\cdot K}(\nu)$ since $\tau_{ij\cdot K}$ is continuous in ν by Theorem 1. Combining this with the convergence result of Theorem 1 gives

$$\sqrt{n}(\hat{\rho}_{i,j\cdot K} - \rho_{i,j\cdot K}) \xrightarrow{D} \mathcal{N}_1\left(0, \tau_{ij\cdot K}(\hat{\nu})\right),$$

and hence

$$\sqrt{n}\Big(\frac{\hat{\rho}_{i,j\cdot K}}{\sqrt{\tau_{ij\cdot K}(\hat{\nu})}} - \frac{\rho_{i,j\cdot K}}{\sqrt{\tau_{ij\cdot K}(\hat{\nu})}}\Big) \xrightarrow{D} \mathcal{N}_1\left(0,1\right).$$

If we define

$$\zeta_{ij\cdot K}(y, \hat{\nu}) := \frac{y}{\sqrt{\tau_{ij\cdot K}(\hat{\nu})}},$$

we obtain

$$\sqrt{n}\Big(\zeta_{ij\cdot K}(\hat{\rho}_{ij\cdot K}, \hat{\nu}) - \zeta_{ij\cdot K}(\rho_{ij\cdot K}, \hat{\nu})\Big) \xrightarrow{D} \mathcal{N}_1(0,1),$$

which completes the proof. □

E Proof of Theorem 2

We rely on the consistency of the causal discovery algorithm that our procedure uses such as PC (Spirtes et al., 2000) or GSP (Solus et al., 2017) in the oracle setting, i.e., when the conditional independence statements of the underlying graph are known. Hence, to prove consistency of our procedure, it is sufficient to show that the conditional independence statements that our procedure estimates from the observed data converges to the true set of conditional independence statements under the faithfulness assumption in (A1).

First, recall that our procedure estimates the CI statements implied by $\mathbb{P}$ through declaring $X_i \perp\!\!\!\perp X_j | X_K$ if

$$|T(\hat{\rho}_{ij\cdot K})| \leq \Phi^{-1}(1 - \frac{\alpha}{2}), \tag{S.9}$$

where T could be one of two statistics:

(i) T is chosen as in Corollary 1 to be

$$T(\hat{\rho}_{ij\cdot K}) = \sqrt{n}\, z_{ij\cdot K}(\hat{\rho}_{ij\cdot K}) := \sqrt{n}\Big(\int \frac{1}{\tau_{ij\cdot K}(\hat{\rho}_{ij\cdot K})} d\hat{\rho}_{ij\cdot K} + C\Big) \tag{S.10}$$

with C chosen such that $z_{ij\cdot K}(0) = 0$ if the conditions of Corollary 1 are satisfied,

(ii) or T is chosen as in Corollary 2 to be

$$T(\hat{\rho}_{ij\cdot K}) = \sqrt{n}\, \zeta_{ij\cdot K}(\hat{\rho}_{ij\cdot K}, \hat{\nu}) := \sqrt{n}\, \frac{\hat{\rho}_{ij\cdot K}}{\tau_{ij\cdot K}(\hat{\rho}_{ij\cdot K}, \hat{\nu})}. \tag{S.11}$$

The first step in proving the theorem is the following lemma.

Lemma S.3. *As $n \to \infty$, the CI statements that our procedure estimates from the observations of X converge to the CI statements implied by $\mathbb{P}$.*

Proof. Take any arbitrary $i, j \in [p]$ and $K \subseteq [p] \setminus \{i, j\}$. First, note that in both settings of $T(\hat{\rho}_{ij \cdot K})$ in (S.10) and (S.11), $T(\hat{\rho}_{ij \cdot K})$ is monotonic and continuous in $\hat{\rho}_{ij \cdot K}$. In the first setting it is the anti-derivative of a strictly positive function of $\hat{\rho}_{ij \cdot K}$ and in the second, it is linear in $\hat{\rho}_{ij \cdot K}$ with positive slope. Monotonicity and the definitions of $z_{ij \cdot K}$ and $\zeta_{ij \cdot K}$ imply that for $n \neq 0$, $T(\rho_{ij \cdot K}) = 0$ if and only if $\rho_{ij \cdot K} = 0$. Continuity and Lemma 1 imply that

$$T(\hat{\rho}_{ij \cdot K}) \xrightarrow{a.s.} T(\rho_{ij \cdot K}) \quad \text{as} \quad n \to \infty.$$

Let H_α be the event that $X_i \not\perp\!\!\!\perp X_j | X_K$ was declared by the test in (S.9). Let H be the event that $X_i \not\perp\!\!\!\perp X_j | X_K$ according to the measure $\mathbb{P}$. Let H' be the event that $X_i \perp\!\!\!\perp X_j | X_K$ according to $\mathbb{P}$. We analyze the limits of the probability of declaring a CI statement correctly, $\mathbb{P}(H_\alpha | H)$, and the limits of declaring a CI statement incorrectly, $\mathbb{P}(H_\alpha | H')$. First, for all $\alpha \in (0, 1]$,

$$\begin{aligned}
\mathbb{P}(H_\alpha | H) &= \mathbb{P}(|T(\hat{\rho}_{ij \cdot K})| > \Phi^{-1}(1 - \frac{\alpha}{2}) \,|\, \rho_{ij \cdot K} \neq 0) \\
&\to \mathbb{P}(|T(\rho_{ij \cdot K})| > \Phi^{-1}(1 - \frac{\alpha}{2}) \,|\, \rho_{ij \cdot K} \neq 0) \\
&\to 1 \quad \text{as} \quad n \to \infty,
\end{aligned} \tag{S.12}$$

where to obtain (S.12), we used that $T(\rho_{ij \cdot K}) \neq 0$ since $\rho_{ij \cdot K} \neq 0$. Hence, $|T(\rho_{ij \cdot K})| = |\sqrt{n} \cdot c| \to \infty$ for $c \neq 0$ as $n \to \infty$. Moreover,

$$\begin{aligned}
&\mathbb{P}(H_\alpha | H') \\
&\quad = \mathbb{P}\Big(|T(\hat{\rho}_{ij \cdot K})| > \Phi^{-1}(1 - \frac{\alpha}{2}) \,\Big|\, \rho_{ij \cdot K} = 0\Big) \\
&\quad = \mathbb{P}\Bigg(T(\hat{\rho}_{ij \cdot K}) > \Phi^{-1}(1 - \frac{\alpha}{2}) | T(\rho_{ij \cdot K}) = 0\Bigg) + \mathbb{P}\Bigg(T(\hat{\rho}_{ij \cdot K}) < \Phi^{-1}(\frac{\alpha}{2}) \Bigg| T(\rho_{ij \cdot K}) = 0\Bigg) \\
&\quad \to \alpha \quad \text{as} \quad n \to \infty,
\end{aligned} \tag{S.13}$$

where (S.13) follows from Corollaries 1 and 2 that assert the asymptotic normality of T in both settings. Hence, for any $\epsilon > 0$, we can set $\alpha_\epsilon = \epsilon/2$ and we will obtain $\mathbb{P}(H_{\alpha_\epsilon} | H') \to \alpha_\epsilon < \epsilon$ as $n \to \infty$. Therefore both errors in estimating the CI statements implied by $\mathbb{P}$ vanish asymptotically, implying that the set of CI statements obtained from observations X converge to those implied by $\mathbb{P}$, thereby completing the proof. □

Proof of Theorem 2. Under faithfulness, the CI statements implied by $\mathbb{P}$ are those implied by $\mathcal{G}$. Hence, by Lemma S.3, the set of CI statements obtained from X as $n \to \infty$ converge to those implied by $\mathcal{G}$. Therefore, if the causal discovery algorithm used in step 6 of Algorithm 1 is consistent in the oracle setting, then Algorithm 1 is consistent. □

F Derivation of the transforms for the dropout model

In this section, we derive the transforms for the dropout model. Recall, in the dropout model introduced in Section 3, we consider an anchored causal model where $Z \sim \mathcal{N}(\mu, \Sigma)$ satisfies (A1). In Example 3.2, the corrupted observation vector X is modeled as

$$X_i = F_i(Z_i) = \begin{cases} Z_i & w.p \quad q_i \\ 0 & w.p \quad 1 - q_i \end{cases} \quad \text{for all} \quad i \in [p], \tag{S.14}$$

with $q_i \in (0, 1]$. Note that Assumption (A2) is satisfied since each X_i is independent of all other variables given its parent Z_i. We can find the moments of Z in terms of the moments of X:

$$\mathbb{E}[X_i] = q_i \mu_i, \quad \mathbb{E}[X_i^2] = q_i \mu_{ii}, \quad \mathbb{E}[X_i X_j] = q_i q_j \mu_{ij} \tag{S.15}$$

for all $i, j \in [p]$ with $i \neq j$, where we defined $\mu_{ij} := \mathbb{E}[Z_i Z_j]$. From this, we can see that Assumption (A2) is satisfied with

$$\eta_i := X_i, \quad \eta_{ii} := X_i^2, \quad \eta_{ij} := X_i X_j, \tag{S.16}$$

and

$$g_i(y) := \frac{y}{q_i}, \quad g_{ii}(y) := \frac{y}{q_i}, \quad g_{ij}(y) := \frac{y}{q_i q_j}. \tag{S.17}$$

F.1 Derivation of the Dropout Stabilizing Transform

In this section, we derive the *dropout stabilizing transform* under the assumption that $\mu_i = 0$ for all $i \in [p]$ and $K = \emptyset$, i.e., we find a variance stabilizing transformation $z_{ij} = z_{ij\cdot\emptyset}$ for the correlations $\rho_{ij} = \rho_{ij\cdot\emptyset}$. We first show that $\tau_{ij}(\nu)$ can be reparameterized as $\tilde{\tau}_{ij}(\rho_{ij})$ and then solve for the dropout stabilizing transform $z_{ij}(\rho)$. We follow the proof of Lemma S.1 and later impose the $\mu = 0$ assumption.

We take any arbitrary, but fixed distinct nodes $i, j \in [p]$ and define

$$\eta := \left(\eta_i \; \eta_j \; \eta_{ii} \; \eta_{jj} \; \eta_{ij}\right)^T$$

as the vector of monomials in X_i and X_j from (S.16). Similarly, we define

$$\hat{\eta} = \left(\hat{\eta}_i \; \hat{\eta}_j \; \hat{\eta}_{ii} \; \hat{\eta}_{jj} \; \hat{\eta}_{ij}\right)^T$$

as the analogous vector of monomials in $\hat{X}_i$ and $\hat{X}_j$ estimated from the observed data.

Then, applying the Central Limit Theorem gives

$$\sqrt{n}(\mathbb{E}[\hat{\eta}] - \mathbb{E}[\eta]) \xrightarrow{D} \mathcal{N}_5(0, A_5(\nu)),$$

where $A_5(\nu)$ is the matrix

$$\begin{pmatrix} \mathrm{Cov}(X_i, X_i) & \mathrm{Cov}(X_j, X_i) & \mathrm{Cov}(X_i^2, X_i) & \mathrm{Cov}(X_j^2, X_i) & \mathrm{Cov}(X_iX_j, X_i) \\ \mathrm{Cov}(X_i, X_j) & \mathrm{Cov}(X_j, X_j) & \mathrm{Cov}(X_i^2, X_j) & \mathrm{Cov}(X_j^2, X_j) & \mathrm{Cov}(X_iX_j, X_j) \\ \mathrm{Cov}(X_i, X_i^2) & \mathrm{Cov}(X_j, X_i^2) & \mathrm{Cov}(X_i^2, X_i^2) & \mathrm{Cov}(X_j^2, X_i^2) & \mathrm{Cov}(X_iX_j, X_i^2) \\ \mathrm{Cov}(X_i, X_j^2) & \mathrm{Cov}(X_j, X_j^2) & \mathrm{Cov}(X_i^2, X_j^2) & \mathrm{Cov}(X_j^2, X_j^2) & \mathrm{Cov}(X_iX_j, X_j^2) \\ \mathrm{Cov}(X_i, X_iX_j) & \mathrm{Cov}(X_j, X_iX_j) & \mathrm{Cov}(X_i^2, X_iX_j) & \mathrm{Cov}(X_j^2, X_iX_j) & \mathrm{Cov}(X_iX_j, X_iX_j) \end{pmatrix},$$

and ν is the vector of all first and second order moments of η. Now, define $w : \mathbb{R}^5 \to \mathbb{R}^1$ as

$$w \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{pmatrix} = \frac{a_5 - a_1 a_2}{\sqrt{a_3 - a_1^2}\sqrt{a_4 - a_2^2}}.$$

Note that we have $w(\mathbb{E}[\eta]) = \rho_{ij}$ and $w(\mathbb{E}[\hat{\eta}]) = \hat{\rho}_{ij}$. Applying the Delta method with w gives

$$\sqrt{n}(\hat{\rho}_{ij} - \rho_{ij}) \xrightarrow{D} \mathcal{N}_1(0, \tau_{ij}(\nu)), \tag{S.18}$$

where

$$\tau_{ij}(\nu) = \nabla w(\mathbb{E}[\eta])^T A_5(\nu) \nabla w(\mathbb{E}[\eta]). \tag{S.19}$$

Carrying out the multiplication gives the asymptotic variance $\tau_{ij}(\nu)$ parameterized by elements of ν. In the case of the dropout model, any moments of X are linear in moments of Z, for example,

$$\mathbb{E}[X_i X_k^2 X_j] = q_i q_k q_j \, \mathbb{E}[Z_i Z_k Z_j].$$

Furthermore, any moments of Z, which is a Gaussian random variable, can be written as polynomials in the first and second order moments of Z, i.e., the elements of μ and Σ. Hence, after imposing the constraint that $\mu = 0$, we can reparameterize $\tau_{ij}(\nu)$ in terms of Σ as

$$\bar{\tau}_{ij}(\Sigma) = \frac{1}{q_i q_j} + \frac{2(\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}})^2}{q_i q_j} - \frac{9(\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}})^2}{4q_j} - \frac{9(\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}})^2}{4q_i} + \frac{(\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}})^2}{2} + (\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}})^4,$$

where $\sigma_{kl} = (\Sigma)_{kl}$. The details of the computation are included in the Supplementary Mathematica Notebook. Now, using $\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}$ we can reparameterize $\bar{\tau}_{ij}(\Sigma)$ once more to obtain

$$\tilde{\tau}_{ij}(\rho_{ij}) = \frac{1}{q_i q_j} + \frac{2\rho_{ij}^2}{q_i q_j} - \frac{9\rho_{ij}^2}{4q_j} - \frac{9\rho_{ij}^2}{4q_i} + \frac{\rho_{ij}^2}{2} + \rho_{ij}^4.$$

Hence, in the $\mu = 0$ case, we can rewrite (S.18) as

$$\sqrt{n}(\hat{\rho}_{ij} - \rho_{ij}) \xrightarrow{D} \mathcal{N}_1(0, \tilde{\tau}_{ij}(\rho_{ij})).$$

In order to find a variance stabilizing transform for ρ_{ij}, we can now solve

$$z_{ij}(\rho) = \int \frac{1}{\sqrt{\tilde{\tau}_{ij}(\rho)}} d\rho + C$$

with C chosen such that $z_{ij}(0) = 0$. Then, by Corollary 1, we have

$$\sqrt{n}\Big(z_{ij}(\hat{\rho}_{ij}) - z_{ij}(\rho_{ij})\Big) \xrightarrow{D} \mathcal{N}_1(0,1).$$

There is no closed form for $z_{ij}(\rho)$ in this case. However, it can written as

$$z_{ij}(\rho) = -\mathbf{i}\frac{\sqrt{1-\frac{8q_iq_j\rho^2}{z_+}}\sqrt{1+\frac{8q_iq_j\rho^2}{z_-}}\int_0^{\mathbf{i}\arcsin 2\rho\sqrt{\frac{2q_iq_j}{z_+}}}(1+\frac{z_+}{z_-}\sin^2\theta)^{-\frac{1}{2}}d\theta}{\sqrt{(\frac{2q_iq_j}{z_+})(\frac{4+(8-9q_i-9q_j+2q_iq_j)\rho^2+4q_iq_j\rho^4}{q_iq_j})}}, \tag{S.20}$$

where $\mathbf{i} = \sqrt{-1}$, and

$$z_+ = +8 - 9q_i - 9q_j + 2q_iq_j + \sqrt{-64q_iq_j + (8 - 9q_i - 9q_j + 2q_iq_j)^2},$$
$$z_- = -8 + 9q_i + 9q_j - 2q_iq_j + \sqrt{-64q_iq_j + (8 - 9q_i - 9q_j + 2q_iq_j)^2}.$$

The integral that appears in the expression of $z_{ij}(\rho)$ is the elliptic integral of the first kind, and can be computed numerically.

F.2 Conditions for the Dropout Stabilizing Transform

As mentioned in Section 4, the dropout stabilizing transform only exists when $\mu = 0$ and $K = \emptyset$. If the derivation was done with non-zero means, it would not have been possible to reparameterize the asymptotic variance of the correlations $\tau_{ij}(\nu)$ in terms of only the correlation ρ_{ij} to satisfy the conditions of Corollary 1. In Figure S.1, we demonstrate the dependence of $\tau_{ij}(\nu)$ from equation (S.19) for fixed ρ_{ij} on $\sigma := \sigma_{ii} = \sigma_{jj}$, which are elements of ν, when $\mu \neq 0$. This can be additionally verified through the Supplementary Mathematica notebook. Figures S.1 (a)-(e) show that τ_{ij} is still dependent on elements of ν, even for a fixed correlation ρ_{ij} for $q \neq 1$ when $\mu \neq 0$, and hence a transform of the kind in Corollary 1 does not exist for $\mu \neq 0$. For $q = 1$, i.e. no dropout, the dropout model reduces to the measurement-error-free Gaussian, and τ_{ij} no longer depends on μ and σ for fixed ρ_{ij}. In this case, a transform of the kind in Corollary 1 does exist and as shown in Lehmann (1998), it is the Fisher's z-transform.

F.3 Derivation of the Dropout Normalizing Transform

In this section we give a way to compute the *dropout normalizing transform* corresponding to Corollary 2 under the dropout model. We begin by showing how to compute the asymptotic variance of the partial correlations, $\tau_{ij\cdot K}(\nu)$.

In the proof of Lemma S.2, we showed that if we know the continuous function $A_\rho(\nu)$ such that

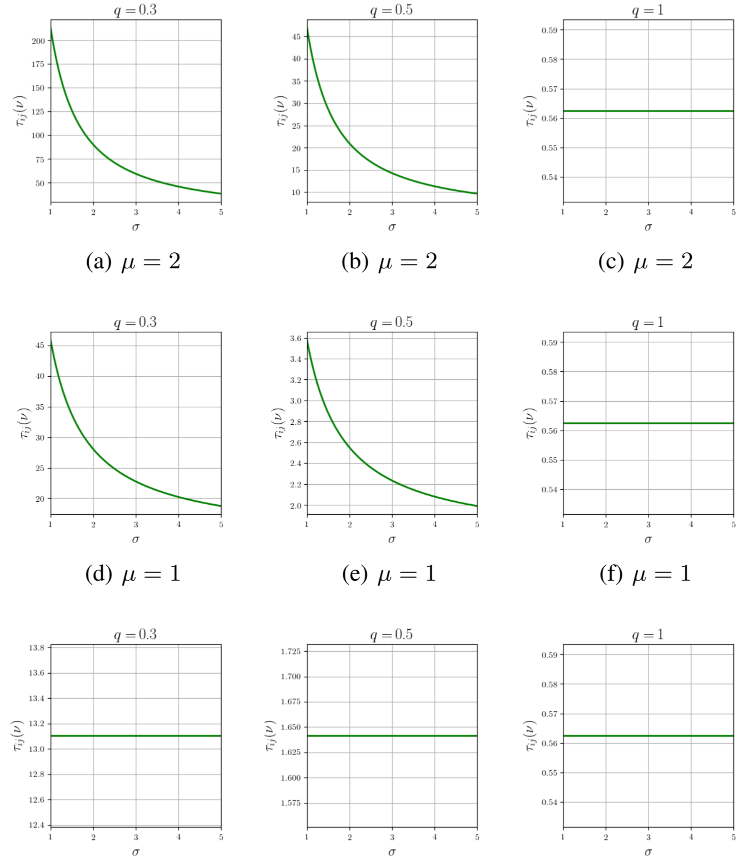
$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{D} \mathcal{N}_{|\rho|}\Big(0, A_\rho(\nu)\Big), \tag{S.21}$$

then we can recursively compute the function $\tau_{ij\cdot K}(\nu)$ beginning with the matrix $A_\rho(\nu)$. Hence, to give a way to compute $\tau_{ij\cdot K}$ for the dropout model, it is sufficient to describe the elements of the matrix $A_\rho(\nu)$ and thus we find a formula for each element of the $A_\rho(\nu)$ matrix. First, recall that the elements of $A_\rho(\nu)$ correspond to the covariances of the sample correlations of the latent variables Z estimated in step 3 of Algorithm 1. That is, each element of $A_\rho(\nu)$ will correspond to the asymptotic covariance of $\sqrt{n}\hat{\rho}_{ab}$ and $\sqrt{n}\hat{\rho}_{cd}$ for some $a, b, c, d \in [p]$ such that $a \neq b$ and $c \neq d$. There are three different cases for each entry in $A_\rho(\nu)$, corresponding to different cases of a, b, c, d:

(a) $\{a, b\} = \{c, d\}$ are distinct, and the element is along the diagonal, corresponding to the asymptotic variance of $\sqrt{n}\hat{\rho}_{ab}$,

(b) $a \notin \{c, d\}$ and $b \in \{c, d\}$,

(c) all of a, b, c, d are distinct.



(a) $\mu = 2$ (b) $\mu = 2$ (c) $\mu = 2$

(d) $\mu = 1$ (e) $\mu = 1$ (f) $\mu = 1$

(g) $\mu = 0$ (h) $\mu = 0$ (i) $\mu = 0$

Figure S.1: Plots of $\tau_{ij \cdot \emptyset}$ when ρ_{ij} is fixed to 0.5, $\mu \in \{0, 1, 2\}$, with σ allowed to vary. This shows that we cannot reparameterize τ_{ij} as a function of only ρ_{ij} for non-zero mean, unless $q = 1$. For $q = 1$ the transform corresponding to Corollary 1 is the Fisher's z-transform.

To analyze all three cases, it is sufficient to take four arbitrary, but fixed distinct $i, j, k, l \in [p]$. We begin by noting that for the dropout model, we can write

$$\begin{aligned}
\hat{\rho}_{i,j} &= \frac{\hat{\mu}_{ij} - \hat{\mu}_i \hat{\mu}_j}{\sqrt{\hat{\mu}_{ii} - \hat{\mu}_i^2}\sqrt{\hat{\mu}_{jj} - \hat{\mu}_j^2}} \\
&= \frac{\frac{1}{q_i q_j}\mathbb{E}[\hat{\eta}_{ij}] - \frac{1}{q_i}\mathbb{E}[\hat{\eta}_i]\frac{1}{q_j}\mathbb{E}[\hat{\eta}_j]}{\sqrt{\frac{1}{q_i}\mathbb{E}[\hat{\eta}_{ii}] - (\frac{1}{q_i}\mathbb{E}[\hat{\eta}_i])^2}\sqrt{\frac{1}{q_j}\mathbb{E}[\hat{\eta}_{jj}] - (\frac{1}{q_j}\mathbb{E}[\hat{\eta}_j])^2}} \\
&= \frac{\mathbb{E}[\hat{\eta}_{ij}] - \mathbb{E}[\hat{\eta}_i]\mathbb{E}[\hat{\eta}_j]}{\sqrt{q_i \mathbb{E}[\hat{\eta}_{ii}] - \mathbb{E}[\hat{\eta}_i]^2}\sqrt{q_j \mathbb{E}[\hat{\eta}_{jj}] - \mathbb{E}[\hat{\eta}_j]^2}}.
\end{aligned} \tag{S.22}$$

Define

$$\hat{\rho} = (\hat{\rho}_{ij}\ \hat{\rho}_{ik}\ \hat{\rho}_{il}\ \hat{\rho}_{jk}\ \hat{\rho}_{jl}\ \hat{\rho}_{kl})^T,$$

as the vector of estimated correlations of X_i, X_j, X_k, X_l obtained from $\mathbb{E}[\hat{\eta}]$ by (S.22). Similarly, let

$$\rho = (\rho_{ij}\ \rho_{ik}\ \rho_{il}\ \rho_{jk}\ \rho_{jl}\ \rho_{kl})^T$$

be the analogous vector of true correlations. In the next part of the derivation, we will apply the Delta method to the vectors of moments of the monomials in X_i, X_j X_k, and X_l of Assumption (A2), to obtain the asymptotic distribution of the vector $\hat{\rho}$, as in the proof of Lemma S.1, We begin by defining the vector of relevant monomials in X_i, X_j, X_k, and X_l

$$\eta = \left(\eta_i\ \eta_j\ \eta_k\ \eta_l\ \eta_{ii}\ \eta_{ij}\ \eta_{ik}\ \eta_{il}\ \eta_{jj}\ \eta_{jk}\ \eta_{jl}\ \eta_{kk}\ \eta_{kl}\ \eta_{ll}\right)^T$$

where the components are defined for our model in equation (S.16). Then, by the Central Limit Theorem, we have

$$\sqrt{n}(\mathbb{E}[\hat{\eta}] - \mathbb{E}[\eta]) \xrightarrow{D} \mathcal{N}_{14}\left(0, A_{14}(\nu)\right) \tag{S.23}$$

where $A_{14}(\nu)$ is the covariance matrix of the vector η, and ν is the vector of all first and second order moments of η. To obtain the convergence result stated in Lemma S.1, we define the function $w_\rho : \mathbb{R}^{14} \to \mathbb{R}^6$ based on (S.22) such that $w_\rho(\mathbb{E}[\eta]) = \rho$ and $w_\rho(\mathbb{E}[\hat{\eta}]) = \hat{\rho}$. Then

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{D} \mathcal{N}_6\Big(0, \nabla w_\rho(\mathbb{E}[\eta])^T A_{14}(\nu) \nabla w_\rho(\mathbb{E}[\eta])\Big). \tag{S.24}$$

Since the moments in η are included in the vector of moments ν, we can define

$$A_6(\nu) := \nabla w_\rho(\mathbb{E}[\eta])^T A_{14}(\nu) \nabla w_\rho(\mathbb{E}[\eta]). \tag{S.25}$$

The explicit form of $A_6(\nu)$ can be found by carrying out the matrix multiplication in (S.25). Before performing the matrix multiplication, we note that for the dropout model, we can write any moment of X as a linear function of a moment of Z, for example,

$$\mathbb{E}[X_i X_k^2 X_j] = q_i q_k q_j \, \mathbb{E}[Z_i Z_k Z_j].$$

Furthermore, since Z is a Gaussian random vector and all moments of a Gaussian random vector can be written in terms of its first and second order moments, we can parameterize the asymptotic covariance with the moments of the Gaussian Z as

$$\bar{A}(\mu, \Sigma) := A_6(\eta).$$

For each entry in $A_6(\eta)$, we list the three cases mentioned previously in terms of the parameterization as $\bar{A}(\mu, \Sigma)$. The full computation is carried out in the Supplementary Mathematica notebook. We use the notation $\sigma_{ij} := (\Sigma)_{ij}$ to denote the elements of Σ. For $a, b, c, d \in \{i, j, k, l\}$,

(a) If the element corresponds to the asymptotic covariance of $\sqrt{n}\hat{\rho}_{ab}$ and $\sqrt{n}\hat{\rho}_{cd}$ with $\{a, b\} = \{c, d\}$, then it is equal to

$$\begin{aligned}
&\frac{1}{\sigma_{aa}\sigma_{bb}}\Big(\frac{\sigma_{aa}\sigma_{bb}}{q_a q_b} \\
&+ \Big(-\mu_a^2\sigma_{bb} - \mu_a^2\mu_b^2 - 4\mu_a\mu_b\sigma_{ab} + \frac{\mu_b^4\sigma_{ab}^2}{4\sigma_{bb}^2} + \frac{\mu_b^2\sigma_{ab}^2}{2\sigma_{bb}}\Big)\frac{1}{q_b} \\
&+ \Big(-\mu_b^2\sigma_{aa} - \mu_a^2\mu_b^2 - 4\mu_a\mu_b\sigma_{ab} + \frac{\mu_a^4\sigma_{ab}^2}{4\sigma_{aa}^2} + \frac{\mu_a^2\sigma_{ab}^2}{2\sigma_{aa}}\Big)\frac{1}{q_a} \\
&+ \big(\mu_a^2\sigma_{bb} + \mu_b^2\sigma_{aa} + \mu_a^2\mu_b^2 + 4\mu_a\mu_b\sigma_{ab} + 2\sigma_{ab}^2\big)\frac{1}{q_a q_b} \\
&+ \Big(-\frac{\mu_a^4\sigma_{ab}^2}{4\sigma_{aa}^2} - \frac{\mu_a^2\sigma_{ab}^2}{2\sigma_{aa}} - \frac{\mu_b^4\sigma_{ab}^2}{4\sigma_{bb}^2} - \frac{\mu_b^2\sigma_{ab}^2}{2\sigma_{bb}} + \frac{\sigma_{ab}^4}{\sigma_{aa}\sigma_{bb}} + \mu_a^2\mu_b^2 + 4\mu_a\mu_b\sigma_{ab} + \frac{\sigma_{ab}^2}{2}\Big) \\
&- \frac{9\sigma_{ab}^2}{4}\Big(\frac{1}{q_a} + \frac{1}{q_b}\Big)\Big);
\end{aligned}$$

(b) If the element corresponds to the asymptotic covariance of $\sqrt{n}\hat{\rho}_{ab}$ and $\sqrt{n}\hat{\rho}_{cd}$ with $a \notin \{c, d\}$ and $b = d$, then the element is equal to

$$\begin{aligned}
&\Big(2q_b\sigma_{bb}\sigma_{bc}(\sigma_{ab}\sigma_{ac}^2\sigma_{bb} - 2\sigma_{aa}\sigma_{ac}\sigma_{bb}\sigma_{bc} + \sigma_{aa}\sigma_{ab}\sigma_{bc}^2) \\
&- \sigma_{cc}(1 - q_b)\mu_b^4\sigma_{aa}\sigma_{ab}\sigma_{bc} \\
&+ 2(1 - q_b)\mu_b^2\sigma_{aa}\sigma_{bb}\sigma_{cc}\sigma_{ab}\sigma_{bc} \\
&- (1 + q_b)\sigma_{bb}^2\sigma_{cc}\sigma_{aa}\sigma_{ab}\sigma_{bc} \\
&+ \sigma_{cc}\sigma_{bb}^2\sigma_{aa}(\sigma_{ab}\sigma_{bc} - 4\sigma_{ac}\sigma_{bb}) \\
&- 4\sigma_{bb}^2\sigma_{ac}\sigma_{cc}(q_b\sigma_{ab}^2 + (q_b - 1)\mu_b^2\sigma_{aa} + \sigma_{bb}\sigma_{aa})\Big)\frac{1}{4q_b\sigma_{bb}^3\sqrt{(\sigma_{aa}\sigma_{cc})^3}};
\end{aligned}$$

(c) If the element corresponds to the asymptotic covariance of $\sqrt{n}\hat{\rho}_{ab}$ and $\sqrt{n}\hat{\rho}_{cd}$ with a, b, c, d distinct, then the element is equal to

$$\begin{aligned}
&\Big(\sigma_{cc}\sigma_{cd}(\sigma_{ab}\sigma_{ad}^2\sigma_{bb} - 2\sigma_{aa}\sigma_{ad}\sigma_{bb}\sigma_{bd} + \sigma_{aa}\sigma_{ab}\sigma_{bd}^2)\\
&+ \sigma_{aa}\sigma_{bb}(\sigma_{ad}\sigma_{bc}\sigma_{cc} + \sigma_{ac}\sigma_{bd}\sigma_{cc} - \sigma_{ac}\sigma_{bc}\sigma_{cc} - \sigma_{ad}\sigma_{bc}\sigma_{cd})\sigma_{dd}\\
&- 2\sigma_{ac}\sigma_{ad}\sigma_{bb]}\sigma_{cc}\sigma_{dd}\sigma_{ab}\\
&+ \sigma_{ac}^2\sigma_{bb}\sigma_{cd}\sigma_{dd}\sigma_{ab}\\
&+ \sigma_{aa}\sigma_{bc}\sigma_{dd}\sigma_{ab}(\sigma_{bc}\sigma_{cd} - 2\sigma_{bd}\sigma_{cc})\Big)\frac{1}{2\sqrt{(\sigma_{aa}\sigma_{bb}\sigma_{cc}\sigma_{dd})^3}}.
\end{aligned}$$

These expressions can be used as the base case of the recursion (S.7), which allows us to indirectly find $\tau_{ij\cdot K}$ for any $i, j \in [p]$ and $K \subseteq [p] \setminus \{i, j\}$ under the dropout model.

The dropout normalizing transform is then computed using

$$\zeta_{ij\cdot K}(\hat{\rho}, \hat{\nu}) := \hat{\rho}/\sqrt{\tau_{ij\cdot K}(\hat{\nu})}.$$

F.4 Derivation of the Dropout Normalizing Transform with Shrinkage

In this section, we derive the dropout normalizing transform when the partial correlations are estimated from the shrinkage matrix $\hat{\Lambda}$ for a fixed shrinkage coefficient α. The derivation closely follows Section F.3. First recall that from Section 4

$$\hat{\Lambda} = (1-\alpha)\hat{\Sigma} + \alpha\hat{S}$$

where $\hat{S} = \sum_{i=1}^{n} X^{(i)}X^{(i)T}$ is covariance matrix of the observations of X. Then, denoting $\hat{\lambda}_{ij} = (\hat{\Lambda})_{ij}$, we can express correlations as

$$\hat{\rho}_{ij} = \frac{\hat{\lambda}_{ij}}{\sqrt{\hat{\lambda}_{ii}}\sqrt{\hat{\lambda}_{jj}}} \tag{S.26}$$

where the elements of $\hat{\Lambda}$ are

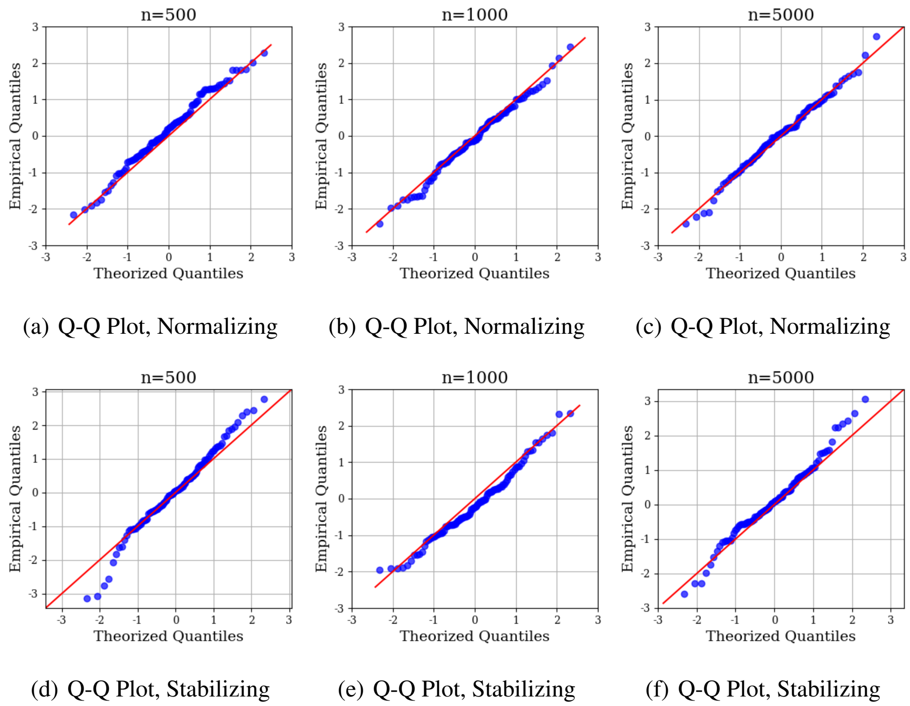
$$\begin{aligned}
\hat{\lambda}_{ij} &= (1-\alpha)\Big(\frac{1}{q_i q_j}\,\mathbb{E}[\hat{\eta}_{ij}] - \frac{1}{q_i}\,\mathbb{E}[\hat{\eta}_i]\frac{1}{q_j}\,\mathbb{E}[\hat{\eta}_j]\Big) + \alpha\Big(\mathbb{E}[\hat{\eta}_{ij}] - \mathbb{E}[\hat{\eta}_i]\,\mathbb{E}[\hat{\eta}_j]\Big),\\
\hat{\lambda}_{ii} &= (1-\alpha)\Big(\frac{1}{q_i}\,\mathbb{E}[\hat{\eta}_{ii}] - (\frac{1}{q_i}\,\mathbb{E}[\hat{\eta}_i])^2\Big) + \alpha\Big(\mathbb{E}[\hat{\eta}_{ii}] - \mathbb{E}[\hat{\eta}_i]^2\Big),\\
\hat{\lambda}_{jj} &= (1-\alpha)\Big(\frac{1}{q_j}\,\mathbb{E}[\hat{\eta}_{jj}] - (\frac{1}{q_j}\,\mathbb{E}[\hat{\eta}_j])^2\Big) + \alpha\Big(\mathbb{E}[\hat{\eta}_{jj}] - \mathbb{E}[\hat{\eta}_j]^2\Big).
\end{aligned} \tag{S.27}$$

We can define the function w_ρ of equation (S.24) based on equations (S.26) and (S.27) and proceed as in Section F.3 to derive the corresponding elements of the asymptotic covariance matrix. The derivation of the dropout normalizing transform with shrinkage, in addition to the result is shown in the Supplementary Mathematica notebook. Note that in this case, the elements of the asymptotic covariance matrix of $\sqrt{n}\hat{\rho}$ will be functions of α.

G Experiments

We include additional simulation results for varying $p \in \{10, 30\}$, $n \in \{1000, 2000, 10000, 50000\}$ and $d \in \{2, 3, 5\}$. Specifically, we evaluate the estimated skeleton as well as the CPDAG in recapitulating the true DAG $\mathcal{G}$ using ROC curves and SHD. For the majority of the settings, the dropout stabilizing transform outperforms the naive Gaussian CI test applied on the corrupted data. As pointed out in Section 5.1, both dropout transforms tend to outperform the Gaussian CI test when the number of samples is high. In plotting the ROC curve for the CPDAG for $p \in \{10, 30\}$, we consider an undirected edge in the CPDAG a true positive if a directed edge exists in $\mathcal{G}$ in either direction, and a false positive otherwise. We consider a directed edge in the CPDAG a true positive if a directed edge of the same direction exists in $\mathcal{G}$, and a false positive otherwise.

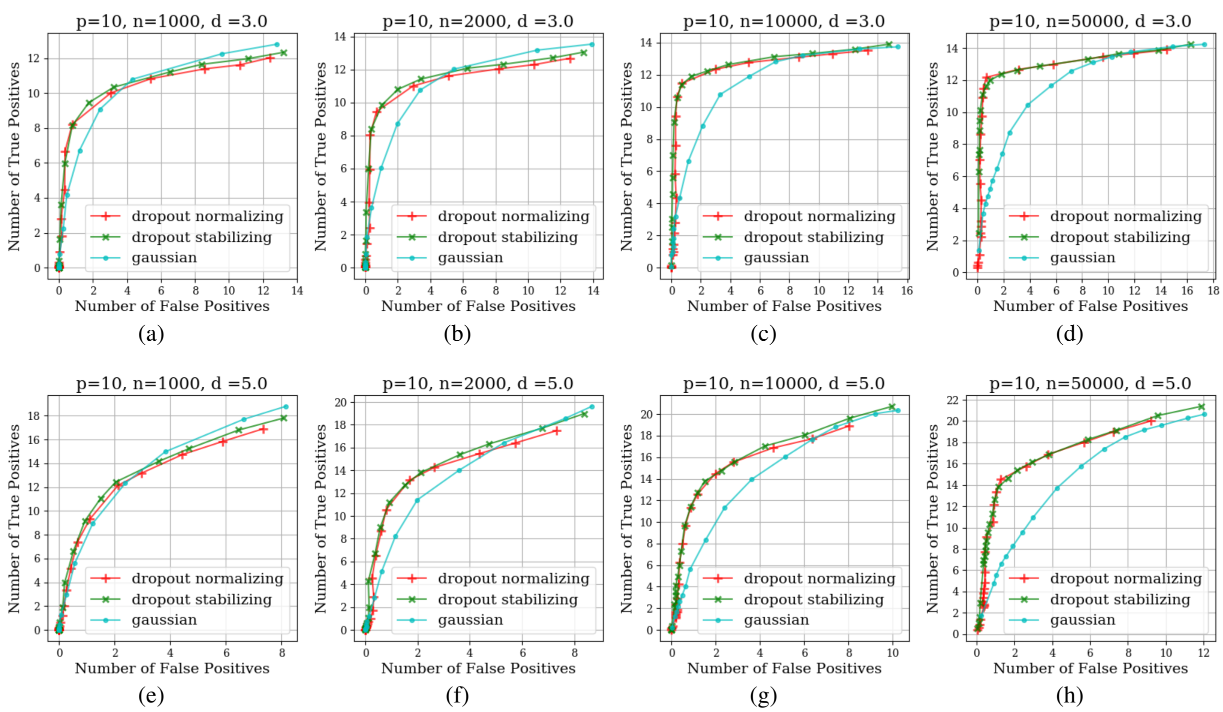
We also include the inferred gene regulatory network for the pancreatic type II diabetes data set, collected with inDrop single-cell RNA-seq technology. We use the dropout stabilizing transform and Algorithm 1 to obtain causal relationships between latent genes.



(a) Q-Q Plot, Normalizing (b) Q-Q Plot, Normalizing (c) Q-Q Plot, Normalizing

(d) Q-Q Plot, Stabilizing (e) Q-Q Plot, Stabilizing (f) Q-Q Plot, Stabilizing

Figure S.2: Q-Q Plots for empirical distributions of the statistic computed from the dropout normalizing and dropout stabilizing transforms under the null hypothesis



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure S.3: ROC curves for evaluating the estimated skeleton of the true DAG using dropout stabilizing transform, dropout normalizing transform, and Gaussian CI test in simulations with $p = 10$ and $n \in \{1000, 2000, 10000, 50000\}$ and $d \in \{3, 5\}$.

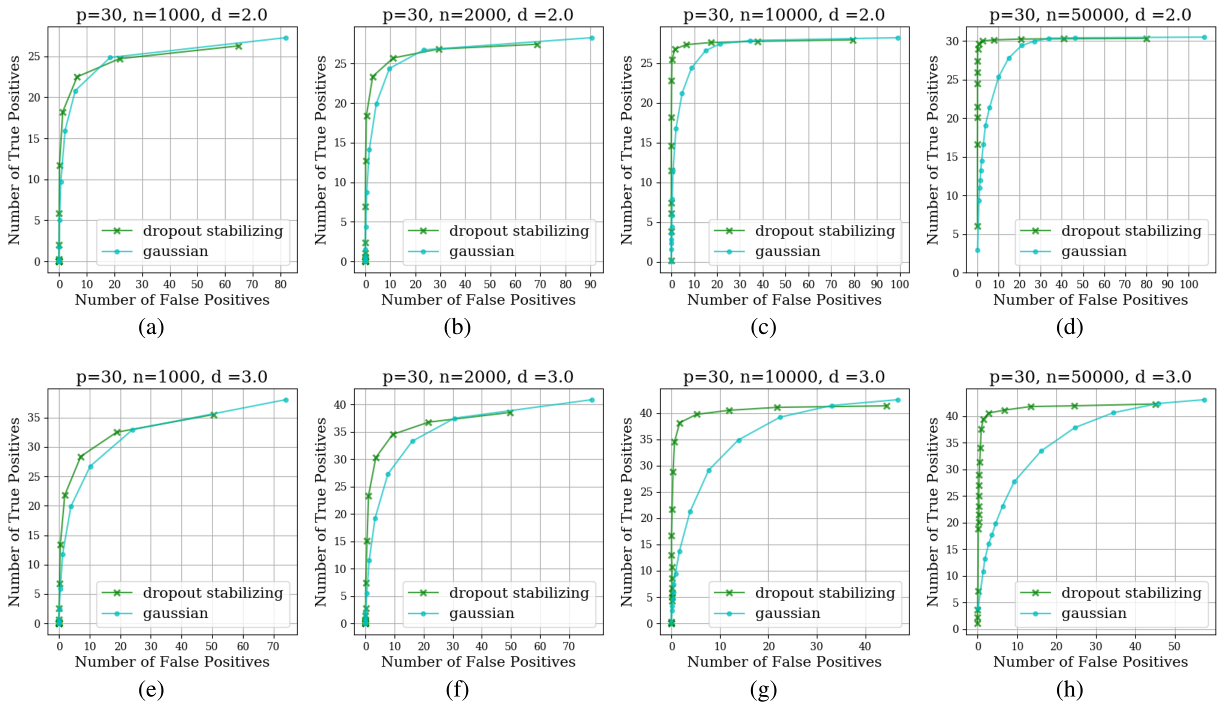


Figure S.4: ROC curves for evaluating the estimated skeleton of the true DAG using dropout stabilizing transform and Gaussian CI test in simulations with $p = 30$ and $n \in \{1000, 2000, 10000, 50000\}$ and $d \in \{2, 3\}$.

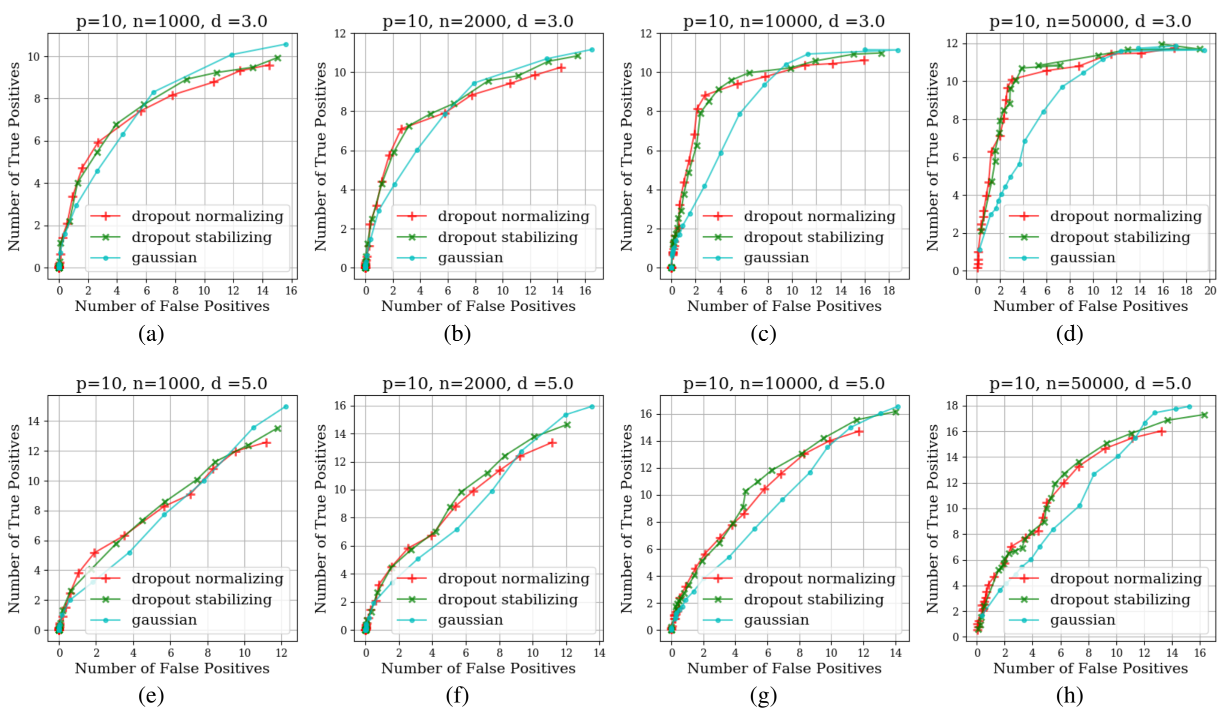


Figure S.5: ROC curves for evaluating the estimated CPDAG of the true DAG using dropout stabilizing transform, dropout normalizing transform, and Gaussian CI test in simulations with $p = 10$ and $n \in \{1000, 2000, 10000, 50000\}$ and $d \in \{3, 5\}$.

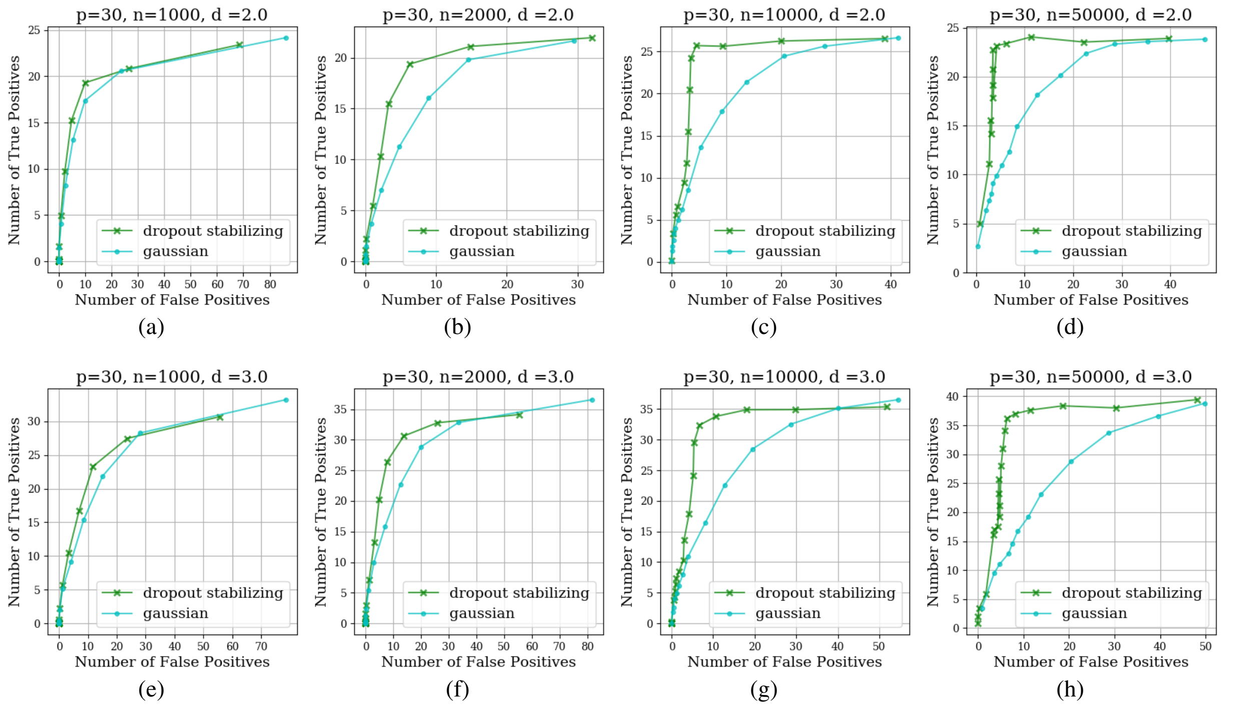


Figure S.6: ROC curves for evaluating the estimated CPDAG of the true DAG using dropout stabilizing transform and Gaussian CI test in simulations with $p = 30$ and $n \in \{1000, 2000, 10000, 50000\}$ and $d \in \{2, 3\}$.

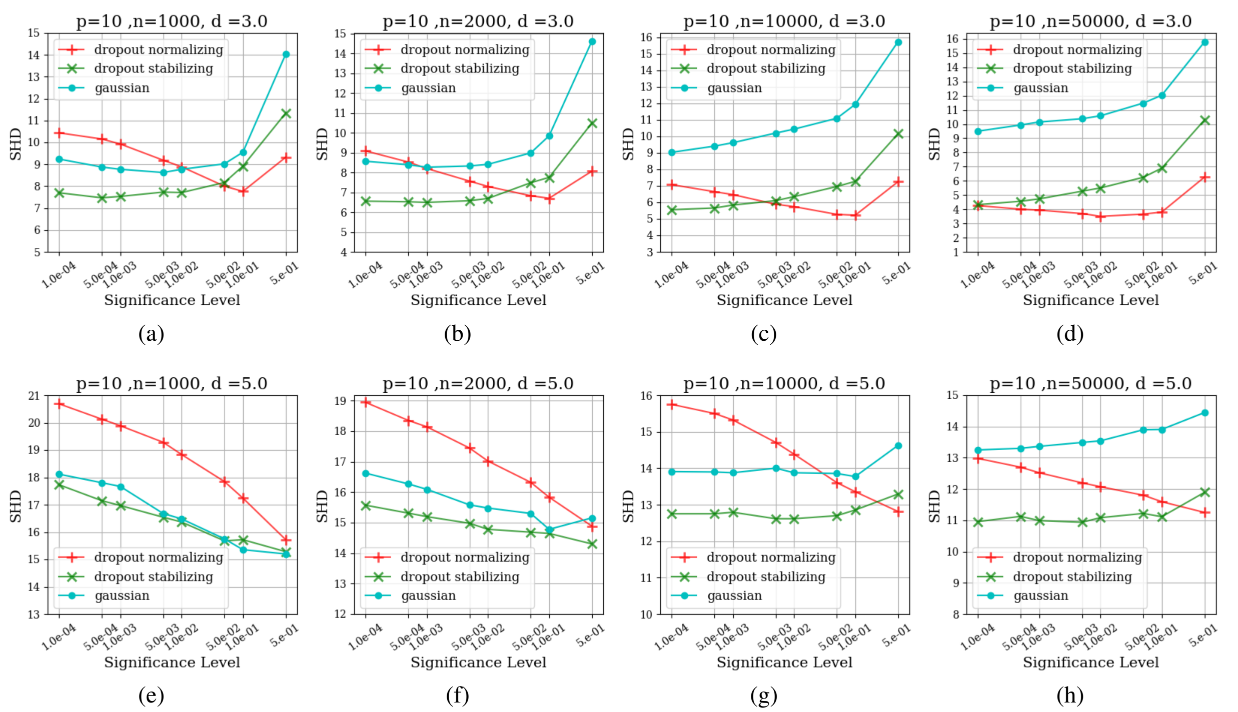


Figure S.7: SHD for evaluating the estimated skeleton of the true DAG using dropout stabilizing transform, dropout normalizing transform, and Gaussian CI test in simulations with $p = 10$ and $n \in \{1000, 2000, 10000, 50000\}$ and $d \in \{3, 5\}$.

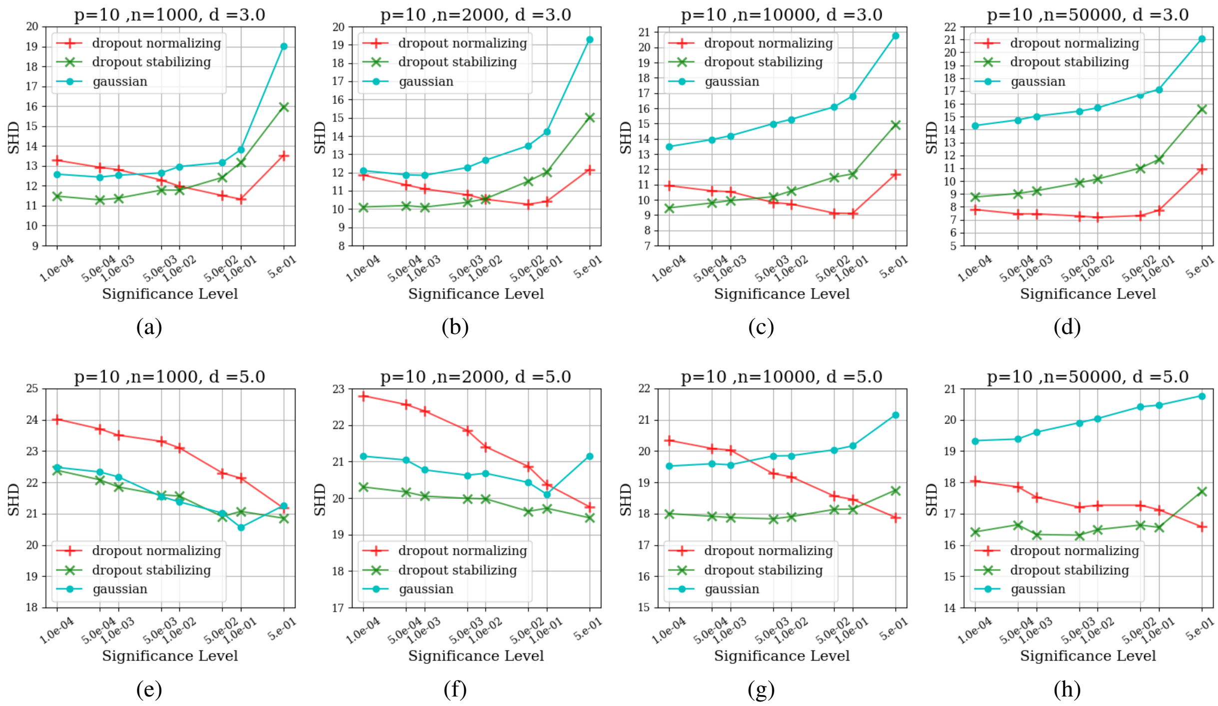


Figure S.8: SHD for evaluating the estimated CPDAG of the true DAG using dropout stabilizing transform, dropout normalizing transform, and Gaussian CI test in simulations with $p = 10$ and $n \in \{1000, 2000, 10000, 50000\}$ and $d \in \{3, 5\}$.

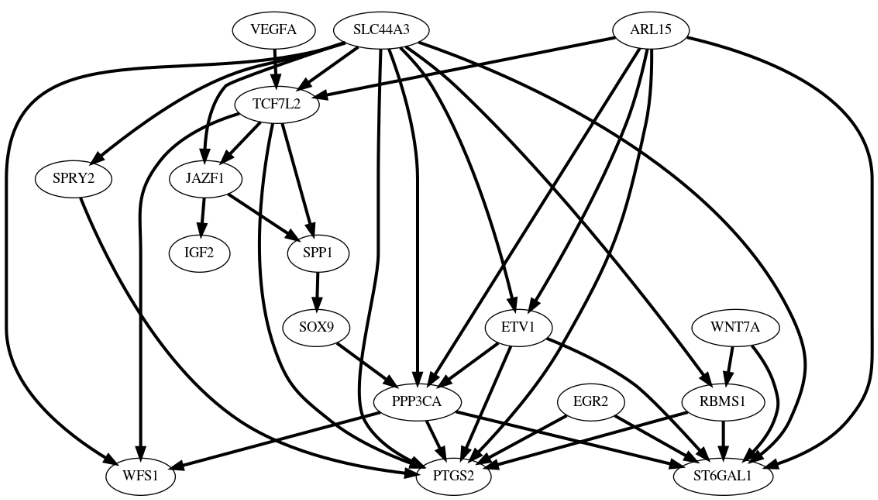


Figure S.9: Gene regulatory network inferred from the pancreas data set collected with inDrop. Dropout stabilizing transform was used to learn the causal edges between latent error-free genes.

References

E. L. Lehmann. *Elements of Large-Sample Theory*. Springer Texts in Statistics, 1998.

L. Solus, Y. Wang, L. Matejovicova, and C. Uhler. Consistency guarantees for permutation-based causal inference algorithms. *Preprint arXiv:1702.03530*, 2017.

P. Spirtes, C. N. Glymour, and R. Scheines. *Causation, Prediction, and Search*. MIT press, 2000.