PROXY-HCD-IV: RECOVERING DIRECT EFFECTS AND HIDDEN CONFOUNDING FROM GROUPED PROXY VARIABLES

Anonymous authors
Paper under double-blind review

ABSTRACT

We study causal discovery when observed variables can be related by both direct effects and shared hidden causes, including cases where both occur simultaneously. Although the hidden causes are not observed, we assume that each is measured indirectly by a **group of proxy variables**, such as multiple noisy sensors or assays of the same underlying quantity. We propose **Proxy-HCD-IV**, which uses these grouped proxies to recover both direct causal relationships and relationships induced by shared hidden causes. The method learns a causal order and divides the proxies for each hidden cause into two independent groups, using one group as noisy regressors and the other as instruments for joint structural estimation. We establish identification and false-addition control under stated conditions, with extensions to nonlinear mechanisms. Experiments include FCI/RFCI, RCD, FoundCause, and matched proxy-aware controls. On an independent 40-graph equal-information benchmark with high proxy noise, Proxy-HCD-IV reduces graphs containing any false direct or hidden-effect selection from **30/40 to 3/40** relative to its matched OLS control, while directed F1 increases from **.898 to .913**, with lower bidirected F1 (**.914 to .733**). A semi-synthetic experiment based on real clinical feature distributions shows a similar trade-off. These results demonstrate how grouped proxy variables can provide additional information for recovering direct and hidden-confounding structure while substantially reducing false structural selections.

1 INTRODUCTION

Many causal systems contain two kinds of relationships among the variables we observe. First, one observed variable can directly affect another, written $X_i \to X_j$. Second, two observed variables can share an unobserved cause, represented by $X_i \leftrightarrow X_j$. Both relationships may hold for the same pair; this is a *bow*. Distinguishing these cases is important because a direct effect and hidden confounding have different causal interpretations.

This paper studies a setting with additional measurement information. For each hidden cause of interest, we observe several noisy *proxy variables*: repeated sensors, assays, or surrogate measurements that are intended to measure the same underlying quantity. The hidden cause itself is never observed. The proxy grouping is supplied by the scientific measurement design; the method does not infer which proxies belong together.

Goal. From the observed variables and grouped proxies, we seek (i) the directed graph among the observed variables and (ii) the pairs of observed variables that share a measured hidden cause. The output is an acyclic directed mixed graph (ADMG) in which bows are allowed. We do not assume access to the hidden variables, the true causal order, the true parent sets, or the structural coefficients.

Why grouped proxies can help. A proxy can be correlated with a node even when its hidden source does not directly affect that node. Conditioning on an observed variable can itself create such an association. We therefore do not infer hidden confounding from conditional proxy relevance. Instead, Proxy-HCD-IV uses two independently noisy views of each proxy group in an instrumental-variable (IV) construction. This lets the method target structural coefficients rather than predictive associations.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

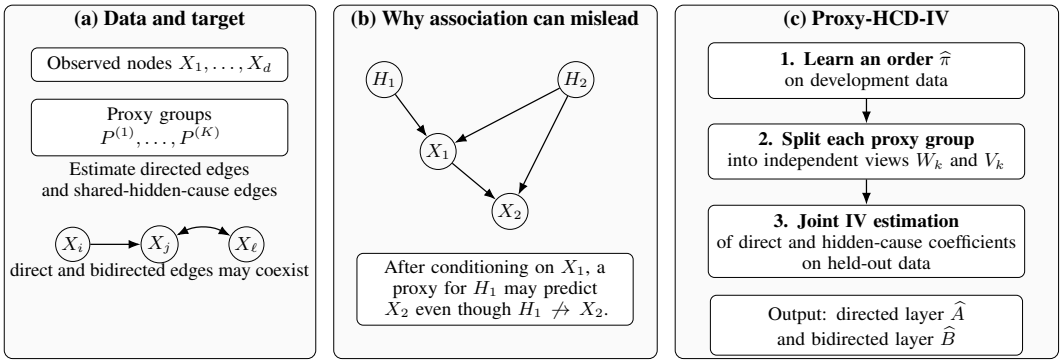


Figure 1: Problem and method. Grouped proxy variables provide repeated noisy measurements of hidden causes. Proxy-HCD-IV uses independent proxy views and held-out IV estimation to recover directed and shared-hidden-cause structure.

Method. Proxy-HCD-IV first learns an order of the observed variables on a development split. For a node X_j, every variable earlier in that order is treated as a candidate direct parent. Each proxy group is then divided into two disjoint groups. One group forms noisy regressors for the hidden causes; the other forms instruments. On an untouched inference split, the method jointly estimates the direct coefficients and the coefficients linking measured hidden causes to each node. Nonzero direct coefficients define directed edges. If the same hidden cause enters two nodes, the method adds a bidirected edge between them.

Contributions. Our contributions are: (1) a bow-capable structural target based on grouped proxy variables; (2) an IV estimator that jointly identifies direct and measured-hidden-cause support under explicit assumptions; (3) inference with asymptotic control of false structural additions, together with results connecting the learned order to the measurement model; and (4) controlled, nonlinear, and semi-synthetic experiments that separate equal-information ablations from comparisons with primary-only causal-discovery methods.

Related work and positioning

Latent-confounder discovery. FCI and RFCI return equivalence-class information when latent variables may be present (Spirtes et al., 2000; Zhang, 2008; Colombo et al., 2012). Stronger functional assumptions permit more orientation: RCD considers linear non-Gaussian models with latent confounders (Maeda & Shimizu, 2020); CAM-UV considers nonlinear additive models with unobserved variables (Maeda & Shimizu, 2021); differentiable and neural ADMG methods learn mixed graphs under their respective model classes (Bhattacharya et al., 2021; Ashman et al., 2023); and FoundCause is a recent pretrained primary-only method (Blöbaum et al., 2026). These methods use the primary variables alone, whereas our method additionally uses scientifically grouped proxy measurements.

Proxy information. Proxy-variable theory shows that hidden causal quantities can become identifiable when suitable proxy measurements satisfy rank or completeness conditions (Miao et al., 2018). Proxy-assisted conditional-independence tests also use such measurements for causal discovery (Liu et al., 2024). Our target is different: we recover an entire bow-capable directed/bidirected graph from grouped proxy measurements.

Order learning. LiNGAM identifies causal order in linear non-Gaussian acyclic models under causal sufficiency (Shimizu et al., 2006). We use proxy information to reduce hidden-root contributions before an ICA-LiNGAM-style order step, and we state explicitly how order errors enter the final guarantee. Related work studies identifiability of linear SCMs with confounding (Sanjaroonpouri & Ramazi, 2025) and efficient updates of Bayesian-network structure when new observed variables arrive (Kalantari et al., 2025). Here the extra measurements are not graph nodes; they are side information about hidden causes.

2 PROBLEM SETUP AND TARGET

For record t, let

$$O_t = (X_t, P_t^{(1)}, \ldots, P_t^{(K)}), \qquad t = 1, \ldots, n,$$

where $X = (X_1, \ldots, X_d)^\top$ are the primary observed variables. Proxy group $P^{(k)} = (P_{k1}, \ldots, P_{km_k})^\top$ contains m_k noisy measurements of hidden cause H_k. Thus, throughout the paper, *repeated measurements* means multiple proxy variables intended to measure the same hidden quantity.

Let $A^\star \in \{0,1\}^{d\times d}$ denote the directed graph: $A^\star_{ij} = 1$ when X_i is a direct parent of X_j. Let $M^\star \in \{0,1\}^{K\times d}$ denote hidden-cause incidence: $M^\star_{kj} = 1$ when H_k directly enters the structural equation for X_j. The bidirected layer induced by the measured hidden causes is

$$B^\star_{ij} = \mathbf{1}\{i \neq j,\ \exists k :\ M^\star_{ki} M^\star_{kj} = 1\}. \tag{1}$$

A bow occurs when $A^\star_{ij} = 1$ and $B^\star_{ij} = 1$. If the supplied proxy groups cover only some hidden causes, $B^\star$ refers only to the confounding covered by those groups; absence of a bidirected edge does not rule out an unmeasured common cause.

Measurement assumption. Each proxy group is *pure*: its variables measure one designated hidden cause rather than mixtures of several hidden causes or descendants of the primary variables. This grouping is external scientific information. The main linear theory further requires that each group can be split into two views with independent measurement errors; Section 4 states this precisely.

3 WHY TWO INDEPENDENT PROXY VIEWS ARE NEEDED

A proxy-node association is not, by itself, a structural hidden-cause edge. Figure 1(b) gives the simplest example. Both H_1 and H_2 affect X_1, and X_1 and H_2 affect X_2. There is no edge $H_1 \to X_2$. Nevertheless, conditioning on X_1 makes H_1 and H_2 statistically dependent, so a proxy for H_1 can become informative about X_2.

This motivates the IV construction. One proxy view is used as a noisy regressor for a hidden cause, and an independent view of the same hidden cause is used as an instrument. Under the assumptions below, the resulting moment equations target the structural coefficients directly. Appendix B gives the population calculation and shows both a false-association case and a cancellation case.

4 PROXY-HCD-IV

We first present the linear model used by the main estimator and theorem, then describe estimation, order learning, and nonlinear extensions.

4.1 LINEAR STRUCTURAL MODEL AND ASSUMPTIONS

Let π be an ordering of the observed variables and define

$$\mathrm{pred}_\pi(j) = \{i : \pi(i) < \pi(j)\}.$$

The order is *admissible* if every true directed edge points forward in the order. For an admissible order, write

$$X_j = \mu_j + \sum_{i \in \mathrm{pred}_\pi(j)} b_{ij} X_i + \sum_{k=1}^{K} \lambda_{kj} H_k + \varepsilon_j, \tag{2}$$

where $b_{ij} = 0$ for predecessor variables that are not true parents. Thus using all predecessors does not assert that they are all parents.

The main identification result uses the following assumptions.

Assumption 1 (Acyclicity and exogeneity)**.** *The directed graph among $X_1, \ldots, X_d$ is acyclic. The hidden causes $H_1, \ldots, H_K$, the primary disturbances $\varepsilon_1, \ldots, \varepsilon_d$, and the proxy measurement errors are centered and mutually independent across these three groups. Primary disturbances have positive finite variance.*

Assumption 2 (Two proxy views)**.** *Each proxy group can be divided into two disjoint summaries*

$$W_k = a_k H_k + \eta_k, \qquad V_k = c_k H_k + \nu_k, \qquad a_k c_k \neq 0, \tag{3}$$

where $\mathbb{E}[\nu\eta^\top] = 0$*. The loadings* a_k, c_k *are unknown. Independent measurement errors across the two view groups are sufficient.*

Assumption 3 (Rank)**.** *For each node* j*, the second-moment matrix of the vector containing an intercept, the predecessors* $X_{\mathrm{pred}_\pi(j)}$*, and the hidden causes* H *is positive definite.*

Assumptions at a glance.

Condition	Why it is needed	If it fails
Pure proxy groups	Connects each supplied group to one hidden cause.	A selected proxy coefficient need not correspond to a single hidden mechanism.
Independent proxy views	Makes V valid as an instrument for the noisy W view.	Cross-view moments can be endogenous; false hidden-cause support can appear.
Admissible order	Ensures all true observed parents are included before each child and no descendants enter its predecessor set.	Direct-effect support can be misspecified and the IV guarantee no longer applies.
Rank/relevance	Makes the cross-view moment matrix invertible.	The relevant coefficients are not identified from the available variation.

With five proxies per hidden cause, the implementation averages two prespecified proxies to form one view and the remaining three to form the other. More generally, the groups are chosen as evenly as possible. This reduces the larger of the two proxy-noise variances when the individual proxy errors have comparable variance.

For node j, define

$$C_j = (1, X_{\mathrm{pred}_\pi(j)}^\top)^\top, \quad R_j = (C_j^\top, W^\top)^\top, \quad Q_j = (C_j^\top, V^\top)^\top.$$

The coefficient vector is

$$\beta_j = \left(\mu_j, (b_{ij})_{i\in\mathrm{pred}_\pi(j)}, (\lambda_{kj}/a_k)_{k=1}^K\right)^\top.$$

Substituting $H_k = (W_k - \eta_k)/a_k$ into Eq. (2) yields

$$X_j = R_j^\top \beta_j + u_j, \qquad \mathbb{E}[Q_j u_j] = 0. \tag{4}$$

The V view is therefore an instrument for the noisy W view. This use of independently noisy indicators as instruments has classical precedent in latent-variable structural equations (Bollen, 1996).

Theorem 1 (Identification of direct and hidden-cause support)**.** *Under Assumptions 1–3 and an admissible order* π*, the matrix*

$$G_j = \mathbb{E}[Q_j R_j^\top]$$

is nonsingular and

$$\beta_j = G_j^{-1}\mathbb{E}[Q_j X_j].$$

Hence the nonzero direct coefficients b_{ij} *and hidden-cause coefficients* λ_{kj} *are identified. Consequently,* $A^\star$*,* $M^\star$*, and the bidirected layer* $B^\star$ *in Eq.* (1) *are identified for the measured hidden causes.*

The proof is in Appendix C. The theorem does not require Gaussianity, known proxy loadings, exact parent sets, or record-by-record reconstruction of the hidden causes.

Corollary 1 (Fixed-basis nonlinear observed effects)**.** *Theorem 1 remains valid if the linear predecessor term is replaced by a prespecified finite feature vector* $\theta_j^\top \phi_j(X_{\mathrm{pred}_\pi(j)})$*, provided the corresponding feature/hidden-cause second-moment matrix is positive definite and the features are exogenous to* ε_j*. A direct edge* $i \to j$ *is represented by a nonzero coefficient group for the features assigned to* X_i*.*

Proposition 1 (Balanced proxy partition)**.** *Suppose a proxy group contains m measurements $P_r = H + \eta_r$ with independent errors of common variance σ^2. If r proxies are averaged into W and the remaining $m - r$ into V, then the two error variances are σ^2/r and $\sigma^2/(m - r)$. Their maximum is minimized when the two group sizes differ by at most one.*

This proposition motivates the balanced partition; it does not claim that the balanced split maximizes graph F1 in every finite sample.

Algorithm summary. The complete procedure is:

1. split the records into development and inference sets;
2. use the development set and proxy-group means to learn an order $\widehat{\pi}$;
3. split each proxy group into two balanced views W_k and V_k;
4. on the untouched inference set, fit Eq. (5) for every node using all predecessors in $\widehat{\pi}$, then apply one Holm correction and construct $\widehat{A}$, $\widehat{M}$, and $\widehat{B}$.

Only Step 2 uses the development rows; all reported coefficient tests use the untouched inference rows.

4.2 Estimation and inference

The data are split into development rows and untouched inference rows. The development split is used only to learn the order $\widehat{\pi}$. On the inference split, all predecessors under $\widehat{\pi}$ are included in C_j, and we estimate

$$\widehat{G}_j = N^{-1} \sum_{t=1}^{N} Q_{jt} R_{jt}^{\top}, \qquad \widehat{\beta}_j = \widehat{G}_j^{-1} N^{-1} \sum_{t=1}^{N} Q_{jt} X_{jt}. \tag{5}$$

HC1 sandwich standard errors give two-sided p-values for all nonintercept coefficients. A single Holm correction at level $\alpha = .10$ is applied to the full family of possible forward direct coefficients and hidden-cause coefficients. Retained direct coefficients define $\widehat{A}$; retained hidden-cause coefficients define $\widehat{M}$; and

$$\widehat{B}_{ij} = \mathbf{1}\Big\{ i \neq j,\ (\widehat{M}^{\top} \widehat{M})_{ij} > 0 \Big\}. \tag{6}$$

If the empirical moment matrix is singular, the procedure abstains for that equation rather than interpreting singularity as evidence of zero effect.

Theorem 2 (Control of false structural additions)**.** *Fix an admissible order independently of the inference rows. Under Assumptions 1–3, finite fourth moments, and nondegenerate coefficient variances, Eq.* (5) *is asymptotically normal with consistently estimated covariance. Holm therefore controls the asymptotic probability of at least one false addition to (A, M) at level α. For an independently learned order $\widehat{\pi}$,*

$$\Pr\{\textit{any false addition to } (A, M)\} \leq \alpha + \Pr(\widehat{\pi} \textit{ is inadmissible}) + o(1).$$

This is an asymptotic, fixed-dimensional result. It is not a weak-instrument guarantee. Correlated view errors, invalid proxy grouping, missing measured roots, or an inadmissible order can invalidate the conclusion.

What the guarantee covers. The familywise statement concerns false additions to the directed support A and the measured-root incidence support M. Because B is constructed from M, a false bidirected edge must be caused by at least one false incidence. The guarantee does not assert that all true edges will be recovered: weak or noisy proxy measurements can reduce power, which is exactly the precision–recall trade-off measured in Section 5. It also does not certify the absence of hidden causes for which no valid proxy group was supplied.

4.3 LEARNING THE ORDER

On the development split, each observed variable is first residualized on the proxy-group means. FastICA is then applied to the standardized residuals, followed by the usual LiNGAM-style normalization and an exact search for the order with minimum squared backward mass. In the experiments, $d = 8$, so this final order search is tractable. This is an ICA-LiNGAM-style adaptation (Shimizu et al., 2006); we do not assume that noisy proxy residualization automatically inherits the causal-sufficiency theorem of standard LiNGAM.

Proposition 2 (Consistency with increasingly informative proxy groups)**.** *Assume Eq.* (2)*, independent non-Gaussian primary disturbances, and m independent affine proxy measurements per hidden cause with bounded error variance and nonvanishing average loading. As $m \to \infty$, the proxy-group means recover each hidden cause up to scale in L_2. If the development-sample regression and ICA estimators are consistent under this vanishing perturbation, the residual system converges to the LiNGAM model $(I - B_0^\top)^{-1}\varepsilon$, and the learned order is admissible with probability tending to one as $m, n \to \infty$.*

Appendix C gives the proof and a finite-sample order-margin certificate. The experiments use five proxies per hidden cause, so order admissibility is also reported directly rather than inferred from Proposition 2.

Finite-sample order certificate. Let B_0 be the true directed coefficient matrix and $\widehat{B}$ the matrix used by the exact backward-mass order search. Put $M_d = d(d-1)/2$ and $\beta_{\min} = \min_{B_{0,ij}\neq 0} |B_{0,ij}|$. Appendix C proves the following deterministic certificate.

Proposition 3 (Finite-sample order margin)**.** *If $\|\widehat{B} - B_0\|_\infty \le \epsilon$ and $(\beta_{\min} - \epsilon)^2 > M_d\epsilon^2$, every minimizing order is admissible. Hence any bound $\Pr(\|\widehat{B} - B_0\|_\infty > \epsilon) \le \delta$ implies $\Pr(\widehat{\pi}$ inadmissible$) \le \delta$ under the same margin condition.*

The certificate separates the causal-order requirement from the estimator-specific FastICA error bound; we do not insert a distribution-free ICA constant that the analysis does not provide.

4.4 NONLINEAR EXTENSIONS

The linear formulation is the main executable estimator. Corollary 1 already permits fixed nonlinear features of observed predecessors. For nonlinear hidden effects and nonlinear proxy channels, identification requires a stronger cross-view condition.

Theorem 3 (Nonlinear cross-view identification)**.** *Suppose $X_j = f_j(C_j) + \sum_{k=1}^{K} g_{kj}(H_k) + \varepsilon_j$ with $\mathbb{E}[\varepsilon_j \mid C_j, H] = 0$. For each k, let U_k, V_k be conditionally independent proxy views given H_k. Assume there are square-integrable bridges h_{kj} satisfying $\mathbb{E}[h_{kj}(U_k) \mid H_k] = g_{kj}(H_k)$, and assume the joint additive conditional-expectation operator is injective: $\mathbb{E}[\sum_k d_k(U_k) \mid V, C_j] = 0$ implies $d_k = 0$ almost surely for every k in the bridge class. Then the conditional moment*

$$\mathbb{E}\left[X_j - f_j(C_j) - \sum_k h_{kj}(U_k) \mid V, C_j\right] = 0$$

identifies the bridges. If the bridge map is injective at zero, $g_{kj} \equiv 0$ iff $h_{kj} \equiv 0$, so hidden-cause support is identified.

This theorem states sufficient conditions; it does not claim that arbitrary nonlinear proxy channels satisfy them. Appendix C gives the proof and the finite-sieve stress tests.

5 EXPERIMENTS

The experiments address three questions. First, how does Proxy-HCD-IV compare with published latent-confounder methods when each method receives its natural input? Second, when information is held fixed, what is the effect of IV and the all-predecessor design? Third, does the same behavior persist when the exogenous distributions are calibrated to real data? Metric definitions and additional stress tests are in Appendix D.

5.1 Comparison with published latent-confounder methods

Table 1 reports the available results on 20 linear-proxy bow graphs. FCI-S, RFCI-Core, RCD-S+DAG, and FoundCause receive only the primary variables X. The proxy-aware methods additionally receive the grouped proxy variables. Because the information sets differ, this table provides reference context rather than an equal-information ranking. FCI/RFCI return PAGs, so quantities not defined by their native output are reported as N/A rather than converted to unsupported exact ADMG scores.

Table 1: Published-method reference comparison on 20 linear-proxy bow graphs. Primary-only methods receive X; proxy-aware methods additionally receive the two five-view proxy blocks and their grouping. Because the information sets differ, this table provides methodological context rather than an equal-input ranking.

Information	Method	Pair F1	Dir. F1	$\leftrightarrow$ F1	Bow F1	SHD $\downarrow$
Primary X only	FCI-S	.835	N/A	N/A	N/A	N/A
	RFCI-Core	.835	N/A	N/A	N/A	N/A
	RCD-S+DAG	.827	.643	.282	.000	7.00
	FoundCause	.725	.722	.229	.183	11.50
X + linear proxy blocks	Sparse-order + OLS	.977	.977	1.000	1.000	.25
	Sparse-order + IV	.966	.977	.967	.990	.45
	Proxy-HCD-OLS	.996	.996	1.000	1.000	.05
	Proxy-HCD-IV	**1.000**	**1.000**	**1.000**	**1.000**	**.00**

FCI-S, RFCI-Core, and RCD-S+DAG denote the local/source-derived implementations used in this benchmark. FoundCause is scored from saved predictions using the prespecified .5 confounder cutoff. The strong proxy-aware rows reflect the accurate linear proxy measurements in this benchmark; Table 2 provides the main equal-information evaluation. N/A means that the method's native PAG output does not directly define the scored structural quantity.

The strong proxy-aware results in Table 1 occur under accurate linear proxy measurements. The independent equal-information experiment below is therefore the main evaluation of the final design.

5.2 Independent equal-information benchmark

The confirmation cohort contains 40 graphs: 20 random mixed graphs and 20 graphs containing the parent-collider pattern of Figure 1(b). Each graph has $d = 8$ observed variables, $K = 3$ hidden causes, $n = 1000$ records, and five proxies per hidden cause. Direct-only, hidden-only, bow, and absent pairs all occur. Each graph is evaluated at proxy-noise SD $.35$ and 1.5. The proxy partition was fixed before this cohort was run.

All rows in Table 2 receive the same observed variables, raw proxy groups, grouping, and 750/250 development/inference split. *Proxy-HCD-OLS* uses the same order and all-predecessor design as Proxy-HCD-IV but replaces IV with OLS. *Sparse-order+IV/OLS* uses the same order learner but selects a sparse predecessor set by BIC before the final regression. These are matched controls, not published methods.

Table 2: Independent equal-information benchmark: 40 graphs (20 random mixed and 20 collider mixed) at each proxy-noise level. Every row receives the same primary variables, five-view proxy blocks, grouping, and 750/250 development/inference split. "Any FP" counts graphs with at least one false directed coefficient or latent incidence. Lower SHD and Any FP are better.

Sensor SD	Method	Pair F1	Dir. F1	↔ F1	Bow F1	SHD ↓	Any FP(A, M) ↓
0.35	Sparse-order + OLS	.985	.985	.974	.771	.550	13/40
	Sparse-order + IV	.988	.985	.986	.779	.425	11/40
	Proxy-HCD-OLS	.985	.988	.975	.800	.475	8/40
	Proxy-HCD-IV	**.987**	**.992**	**.983**	**.817**	**.325**	**4/40**
1.50	Sparse-order + OLS	.905	.767	.889	.464	5.775	40/40
	Sparse-order + IV	.922	.767	.844	.435	5.700	40/40
	Proxy-HCD-OLS	**.952**	.898	**.914**	**.657**	**2.925**	30/40
	Proxy-HCD-IV	.863	**.913**	.733	.558	3.100	**3/40**

All runs use the balanced cross-view partition. At high noise, relative to the matched all-predecessor OLS control, Proxy-HCD-IV reduces graphs with any false support addition from 30/40 to 3/40 while increasing directed F1 from .898 to .913. Bidirected F1 decreases from .914 to .733 and bow F1 from .657 to .558; the mean SHD difference is .175. Sparse-order denotes the ICA-LiNGAM-style matched control described in the text.

At high proxy noise, Proxy-HCD-IV reduces graphs with at least one false direct or hidden-cause selection from $30/40$ to $3/40$ relative to Proxy-HCD-OLS, while directed F1 increases from $.898$ to $.913$. Bidirected F1 decreases from $.914$ to $.733$, showing that the reduction in false structural additions is obtained at the cost of hidden-confounding recall. The paired difference in the false-support indicator is $-.675$ with descriptive 95% interval $[-.825, -.525]$. The learned order is admissible in all 40 low-noise graphs and 38 of 40 high-noise graphs.

5.3 Semi-synthetic benchmark with real feature distributions

To reduce dependence on idealized simulation marginals, we construct a semi-synthetic benchmark from the Wisconsin Diagnostic Breast Cancer (WDBC) data (Wolberg et al., 1993). WDBC contains 569 records and 30 continuous image-derived nuclear features. Standardized empirical feature columns provide bootstrap distributions for hidden causes, primary disturbances, and proxy errors. The graph, coefficients, hidden-cause incidences, and proxy assignment remain controlled, so the exact directed, bidirected, and bow ground truth is known. This is a real-distribution calibration, not a causal analysis of breast cancer.

Table 3: WDBC-calibrated semi-synthetic benchmark (20 mixed graphs). Exogenous roots, disturbances, and proxy errors are independent bootstrap draws from standardized empirical WDBC feature marginals; the mixed graph and coefficients are controlled so exact ground truth is known.

Proxy SD	Method	Pair F1	Dir. F1	↔ F1	Bow F1	SHD ↓
.35	Proxy-HCD-OLS	.962	.930	.919	.593	**2.00**
	Proxy-HCD-IV	.960	.930	.912	.585	2.10
1.50	Proxy-HCD-OLS	**.958**	.830	**.932**	.486	**2.90**
	Proxy-HCD-IV	.846	**.869**	.752	**.558**	3.50

At high proxy noise, Proxy-HCD-IV raises directed F1 from $.830$ to $.869$ and reduces graphs with any false direct/hidden-cause selection from $17/20$ to $5/20$ relative to the matched OLS control. Bidirected F1 decreases from $.932$ to $.752$ and SHD increases from 2.90 to 3.50. Thus the same false-addition/recall trade-off appears when the exogenous distributions are taken from real clinical feature marginals.

5.4 STRESS TESTS AND ABLATIONS

The graph-family decomposition in Table 4 separates generic random mixed graphs from graphs containing the conditioning pattern in Figure 1(b). At high proxy noise, Proxy-HCD-IV has one graph with any false support addition in the collider family, compared with 14/20 for Proxy-HCD-OLS; in that family it also improves bow F1 and SHD. On the random family, IV is more conservative and loses more bidirected recall.

Table 4: High-noise ($\sigma = 1.5$) independent confirmation decomposed by graph family. Each family contains 20 new graphs.

Family	Method	Dir. F1	$\leftrightarrow$ F1	Bow F1	SHD	Any FP
Random	Proxy-HCD-OLS	.868	.954	.488	2.75	16/20
	Proxy-HCD-IV	**.906**	.654	.243	3.40	**2/20**
Collider	Proxy-HCD-OLS	**.928**	.874	.825	3.10	14/20
	Proxy-HCD-IV	.920	.811	**.873**	**2.80**	**1/20**

Table 5 evaluates the order learner as the number of proxies per hidden cause varies. Across 240 fits, admissibility ranges from 90% to 100%; the experiment is not monotone in the number of views and is not presented as a proof of Proposition 2.

Table 5: Order-admissibility stress test at proxy-noise SD 1.5: 40 new linear mixed graphs (20 random, 20 collider) for each number m of repeated measurements per root. The same primary graph/data are reused across m within a seed.

Views m	2	3	5	10	20	50
Admissible orders	36/40	40/40	37/40	39/40	39/40	39/40
Mean reversed true edges	.175	.000	.125	.025	.025	.025

Appendix D gives the matched high-noise ablation and nonlinear observed-mechanism and nonlinear-proxy stress tests. Appendix B reports a negative control in which correlated errors across the two proxy views violate the IV moment condition.

6 DISCUSSION AND LIMITATIONS

Proxy-HCD-IV addresses a specific causal-discovery setting: hidden causes are not observed directly, but each measured hidden cause has a scientifically defined group of proxy variables. Under independent proxy views and an admissible order, these measurements provide IV moments that separate direct observed-variable effects from measured hidden-cause effects. Shared hidden-cause support then induces the bidirected layer, so direct and bidirected edges can coexist.

The experiments show a consistent operating characteristic. With strong proxies, Proxy-HCD-IV is highly accurate. With noisy proxies, it is more conservative than the matched OLS estimator: it makes far fewer false structural additions but can miss true hidden-confounding edges. The method is therefore most appropriate when the proxy grouping and independent-view design are credible and false structural claims are especially costly.

The assumptions also delimit the conclusions. The main theorem uses linear hidden contributions and affine proxy summaries; the nonlinear extension requires bridge existence and completeness. Correlated view errors, invalid proxy groups, weak instruments, unmeasured hidden causes without proxy groups, or model misspecification can break identification. The order result is asymptotic, and finite-sample order accuracy depends on the non-Gaussianity and conditioning of the ICA problem. Finally, the primary-only baselines in Table 1 receive less information than Proxy-HCD-IV and should not be interpreted as an equal-input performance ranking.

REFERENCES

Matthew Ashman, Chao Ma, Agrin Hilmkil, Joel Jennings, and Cheng Zhang. Causal reasoning in the presence of latent confounders via neural ADMG learning. In *The Eleventh International Conference on Learning Representations*, 2023.

Rohit Bhattacharya, Tushar Nagarajan, Daniel Malinsky, and Ilya Shpitser. Differentiable causal discovery under unmeasured confounding. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 2314–2322. PMLR, 2021.

Patrick Blöbaum, Krishnakumar Balasubramanian, and Shiva Prasad Kasiviswanathan. FoundCause: Causal discovery with latent confounders from observational data, 2026. Preprint.

Kenneth A. Bollen. An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika*, 61(1):109–121, 1996. doi: 10.1007/BF02296961.

Diego Colombo, Marloes H. Maathuis, Markus Kalisch, and Thomas S. Richardson. Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, 40(1):294–321, 2012. doi: 10.1214/11-AOS940.

Luigi Gresele, Paul K. Rubenstein, Arash Mehrjou, Francesco Locatello, and Bernhard Schölkopf. The incomplete rosetta stone problem: Identifiability results for multi-view nonlinear ICA. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence*, volume 115 of *Proceedings of Machine Learning Research*, pp. 217–227. PMLR, 2020.

Hamid Kalantari, Russell Greiner, and Pouria Ramazi. Extendable and iterative structure learning strategy for bayesian networks. In *The Thirteenth International Conference on Learning Representations*, 2025.

Mingzhou Liu, Xinwei Sun, Yu Qiao, and Yizhou Wang. Causal discovery via conditional independence testing with proxy variables. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 31866–31889. PMLR, 2024.

Takashi Nicholas Maeda and Shohei Shimizu. RCD: Repetitive causal discovery of linear non-gaussian acyclic models with latent confounders. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 735–745. PMLR, 2020.

Takashi Nicholas Maeda and Shohei Shimizu. Causal additive models with unobserved variables. In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pp. 97–106. PMLR, 2021.

Wang Miao, Zhi Geng, and Eric J. Tchetgen Tchetgen. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993, 2018. doi: 10.1093/biomet/asy038.

Vahideh Sanjaroonpouri and Pouria Ramazi. Linear scm identification in the presence of confounders and gaussian noise. In *The Thirteenth International Conference on Learning Representations*, 2025.

Shohei Shimizu, Patrik O. Hoyer, Aapo Hyvärinen, and Antti Kerminen. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7:2003–2030, 2006.

Peter Spirtes, Clark N. Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT Press, 2 edition, 2000.

William Wolberg, Olvi Mangasarian, Nick Street, and W. Street. Breast cancer wisconsin (diagnostic). UCI Machine Learning Repository, 1993. Dataset.

Jiji Zhang. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16–17):1873–1896, 2008. doi: 10.1016/j.artint.2008.08.001.

REPRODUCIBILITY, ETHICS, AND AI ASSISTANCE

The archive contains predictions, p-values, coefficient matrices, learned orders, split indices, graph truths stored separately for evaluation, FoundCause prediction files and rescoring checks, and scripts used to build the reported tables. FCI/RFCI/RCD results retain their implementation provenance. The WDBC experiment is semi-synthetic and does not support clinical conclusions. Generative AI assisted mathematical checking, experiment implementation, manuscript restructuring, and typesetting; the authors remain responsible for the claims, code, citations, and final content.

A APPENDIX GUIDE

The appendix has three parts. Section B gives the population example that explains why proxy association after conditioning need not equal a structural hidden-cause effect. Section C gives the proofs and the nonlinear and finite-order extensions. Section D provides metric definitions, ablations, stress tests, the semi-synthetic protocol, and baseline implementation details.

B WHY PROXY ASSOCIATION CAN DIFFER FROM HIDDEN-CAUSE SUPPORT

B.1 A PARENT-COLLIDER EXAMPLE

Consider first the four-node mechanism used in Figure 1(b):

$$X_1 = H_1 + H_2 + \varepsilon_1, \qquad X_2 = aX_1 + \delta H_1 + H_2 + \varepsilon_2, \tag{7}$$

with independent standard-normal roots and disturbances. There is no structural incidence $H_1 \to X_2$ when $\delta = 0$, yet conditioning on X_1 couples H_1 and H_2. With an exact H_1 measurement, the coefficient of H_1 after adjustment for X_1 is $\delta - \frac{1}{2}$: $\delta = 0$ therefore creates a nonzero conditional association without a structural incidence, while $\delta = \frac{1}{2}$ cancels a genuine incidence.

More generally, let $H = (H_1, \dots, H_K)^\top$ have covariance $\Sigma_H \succ 0$ and suppose

$$X_j = a_j^\top C_j + \lambda_j^\top H + \varepsilon_j,$$

where C_j is an observed adjustment context and ε_j is independent of (C_j, H). The best linear coefficient of H after adjustment is governed by

$$V_j \lambda_j, \qquad V_j = \Sigma_H - \Sigma_{HC_j} \Sigma_{C_j C_j}^{-1} \Sigma_{C_j H}.$$

If C_j contains a collider or descendant that mixes hidden roots, V_j need not be diagonal. A nonincident root can therefore acquire a nonzero adjusted coefficient, while a true incidence can be attenuated or cancelled. This is the population mechanism behind Eq. (7); it is a property of the conditioning target itself rather than a failure of a particular residual-testing algorithm.

B.2 CROSS-VIEW MOMENTS

For repeated measurements

$$W_k = a_k H_k + \eta_k, \qquad V_k = c_k H_k + \nu_k,$$

with nonzero loadings and $\mathbb{E}[\nu\eta^\top] = 0$, substitute $H_k = (W_k - \eta_k)/a_k$ into the structural equation. The resulting regression error contains the W-view measurement error, so OLS on W is generally biased. The independent V view, however, supplies instruments satisfying $\mathbb{E}[Q_j u_j] = 0$ under the stated exogeneity assumptions. The identifying moments are therefore unconditional cross-view moments rather than conditional proxy-relevance statements. The complete joint-support proof is given next.

B.3 NEGATIVE CONTROL: CORRELATED PROXY-VIEW ERRORS

The independent-view requirement is substantive. In the controlled negative experiment, deliberately correlating the two view-error groups causes the nonexistent target incidence to be selected in $190/200$ runs at $n = 1000$ and $200/200$ at $n = 4000$. This failure shows that the independent-view assumption is part of the identification argument, not merely a regularization choice.

C PROXY-HCD-IV: PROOFS AND LIMITATIONS

C.1 ADMISSIBLE PREDECESSOR CONTEXTS

Let B_0 denote the observed directed coefficient matrix in source-by-target convention, so $(B_0)_{ij} = b_{ij}$. An order π is *admissible* if $b_{ij} \neq 0$ implies $\pi(i) < \pi(j)$. For node j, let $P_j = \mathcal{P}_\pi(j)$ and $C_j^\pi = (1, X_{P_j}^\top)^\top$. Every true observed parent of X_j is contained in P_j, and no descendant of X_j is contained in P_j. Consequently C_j^π is measurable with respect to the roots and primary disturbances other than ε_j, so

$$\mathbb{E}[C_j^\pi \varepsilon_j] = 0, \qquad \mathbb{E}[H\varepsilon_j] = 0.$$

With pure measurement errors independent of the roots and primary disturbances, these imply the unconditional exogeneity conditions required below.

C.2 PROOF OF THEOREM 1

Pad the true observed coefficients by zeros for predecessor nonparents, so Eq. (2) holds exactly over all predecessors. Substituting $H_k = (W_k - \eta_k)/a_k$ into Eq. (2) gives

$$X_j = R_j^\top \beta_j + u_j, \qquad u_j = \varepsilon_j - \sum_k (\lambda_{kj}/a_k)\eta_k.$$

By the exogeneity identities,

$$\mathbb{E}[Q_j u_j] = 0, \qquad \mathbb{E}[Q_j X_j] = G_j \beta_j, \qquad G_j = \mathbb{E}[Q_j R_j^\top].$$

It therefore suffices to show that G_j is nonsingular.

Let

$$S_j = ((C_j^\pi)^\top, H^\top)^\top, \qquad \Gamma_j = \mathbb{E}[S_j S_j^\top].$$

By Assumption 3, $\Gamma_j \succ 0$. Define the diagonal loading matrices

$$D_a = \mathrm{diag}(I_{1+|P_j|}, a_1, \ldots, a_K), \qquad D_c = \mathrm{diag}(I_{1+|P_j|}, c_1, \ldots, c_K).$$

Then

$$R_j = D_a S_j + (0, \eta^\top)^\top, \qquad Q_j = D_c S_j + (0, \nu^\top)^\top.$$

All cross terms involving measurement errors vanish by centering, exogeneity, and $\mathbb{E}[\nu\eta^\top] = 0$, so

$$G_j = D_c \Gamma_j D_a^\top.$$

Both D_a and D_c are nonsingular because $a_k c_k \neq 0$, and $\Gamma_j \succ 0$, hence G_j is nonsingular. Therefore

$$\beta_j = G_j^{-1} \mathbb{E}[Q_j X_j],$$

which identifies both the direct-coefficient support and the latent-incidence support. □

C.3 PROOF OF COROLLARY 1

Replace the raw predecessor coordinates in C_j^π by the fixed finite feature vector $\phi_j(X_{\mathcal{P}_\pi(j)})$. Because these features are functions of predecessors, they remain independent of the current disturbance ε_j under an admissible order. The same substitution of $H_k = (W_k - \eta_k)/a_k$ gives the same moment equation, now with an enlarged feature/root vector. Positive definiteness of its second-moment matrix again makes G_j nonsingular. A direct edge is represented by a nonzero coefficient group assigned to that predecessor. □

C.4 PROOF OF PROPOSITION 1

For $W = r^{-1}\sum_{s=1}^{r} P_s$ and $V = (m-r)^{-1}\sum_{s=r+1}^{m} P_s$, conditional independence of the measurement errors gives

$$\mathrm{Var}(W - H) = \sigma^2/r, \qquad \mathrm{Var}(V - H) = \sigma^2/(m-r).$$

The larger of $1/r$ and $1/(m-r)$ is minimized when the two group sizes are as equal as possible. For $m = 5$, this is a balanced partition. □

C.5 Why the matched OLS control is not equivalent

Even with the correct predecessor context, ordinary regression on noisy proxy summaries does not target the same coefficients. Let C be a centered predecessor vector and $Z = H + e$ a noisy proxy summary with independent error covariance Σ_e. If

$$Y = C^\top \theta + H^\top \lambda + \varepsilon,$$

then population OLS of Y on (C, Z) yields latent coefficient

$$\gamma_Z = (V + \Sigma_e)^{-1} V \lambda, \qquad V = \Sigma_{HH} - \Sigma_{HC}\Sigma_{CC}^{-1}\Sigma_{CH},$$

and observed coefficient

$$\theta_Z = \theta + \Sigma_{CC}^{-1}\Sigma_{CH}(\lambda - \gamma_Z).$$

Hence measurement error can distort both latent-incidence and direct-effect coefficients. This explains why Proxy-HCD-OLS can have good aggregate accuracy while still making many graph-level false support additions.

C.6 Proof of Theorem 2

Fix an admissible order independently of the inference rows. The estimator obeys

$$\widehat{\beta}_j - \beta_j = \widehat{G}_j^{-1} \frac{1}{N} \sum_{t=1}^{N} Q_{jt} u_{jt}.$$

Under finite fourth moments, a law of large numbers gives $\widehat{G}_j \to G_j$, and a multivariate central limit theorem gives

$$\sqrt{N}(\widehat{\beta}_j - \beta_j) \Rightarrow N(0, G_j^{-1}\Omega_j G_j^{-\top}), \qquad \Omega_j = \mathbb{E}[Q_j Q_j^\top u_j^2].$$

The HC1 sandwich covariance is consistent in fixed dimension, so each two-sided normal p-value is asymptotically valid for a fixed coefficient null. There are

$$m = d(d-1)/2 + Kd$$

nonintercept hypotheses across the system. Holm's procedure therefore controls the asymptotic probability of any false addition to (A, M) at level α, without requiring dependence assumptions across equations. Since every false addition to B must arise from a false latent incidence, the same bound also protects the induced bidirected layer.

For an independently learned order $\widehat{\pi}$, condition on that order and split the probability into admissible and inadmissible events. This yields

$$\Pr\{\text{false addition to } (A, M)\} \le \alpha + \Pr(\widehat{\pi} \text{ inadmissible}) + o(1).$$

This completes the proof. □

C.7 Nonlinear cross-view extension

Theorem 3 is stated in the main paper. We give its proof here.

C.8 Proof of Theorem 3

Let

$$r_j = X_j - f_j(C_j) - \sum_k h_{kj}(U_k).$$

By the structural equation, conditional view independence, and the bridge equations,

$$\mathbb{E}[r_j \mid V, C_j] = \sum_k \mathbb{E}[g_{kj}(H_k) - h_{kj}(U_k) \mid V, C_j] + \mathbb{E}[\varepsilon_j \mid V, C_j] = 0.$$

Suppose another bridge collection $\widetilde{h}_{kj}$ satisfies the same conditional moment. Their difference $d_k = h_{kj} - \widetilde{h}_{kj}$ obeys

$$\mathbb{E}\left[\sum_k d_k(U_k) \mid V, C_j\right] = 0.$$

By the stated injectivity condition for the joint additive operator, the only solution is $d_k = 0$ almost surely for every k. Thus the bridges are unique. If the bridge map is injective at zero, $h_{kj} = 0$ almost surely exactly when $g_{kj} = 0$ almost surely, which identifies incidence support. The theorem does not assert that these conditions hold for arbitrary proxy channels; it states the conditions required for the nonlinear extension. □

C.9 Proof of Proposition 2

Let the m measurements of root H_k be $P_{kr} = a_{kr}H_k + \eta_{kr}$, with conditionally independent centered errors of uniformly bounded variance and average loading $\bar{a}_{k,m} = m^{-1}\sum_r a_{kr} \to a_k \neq 0$. Then the block mean obeys

$$\bar{P}_{k,m} = \bar{a}_{k,m}H_k + \bar{\eta}_{k,m}, \qquad \mathbb{E}[\bar{\eta}^2_{k,m}] = O(m^{-1}),$$

so $\bar{P}_{k,m} - a_k H_k \to 0$ in L_2. In the linear structural model, writing $T = (I - B_0^\top)^{-1}$ gives

$$X = T\mu + T\Lambda^\top H + T\varepsilon.$$

Population linear projection of X on the limiting scaled roots removes $T\Lambda^\top H$, leaving $T\varepsilon$. Consistency of the development-sample regression and the assumed stability of the ICA estimator under the vanishing proxy perturbation therefore make the empirical residual matrix converge to the standard LiNGAM mixing model. Independent non-Gaussian disturbances identify that mixing matrix up to scaling and permutation, and the LiNGAM diagonal normalization/triangularization recovers an admissible order under its usual identifiability conditions (Shimizu et al., 2006). Hence $\Pr(\widehat{\pi} \text{ admissible}) \to 1$. □

C.10 Finite-sample order-margin certificate

Proposition 3 is stated in the main paper.

C.11 Proof of Proposition 3

Let $L_{\widehat{B}}(\pi)$ be the sum of squared entries of $\widehat{B}$ that point backward under order π. For any admissible true order $\pi^\star$, every corresponding entry of B_0 is zero, so under $\|\widehat{B} - B_0\|_\infty \le \epsilon$,

$$L_{\widehat{B}}(\pi^\star) \le M_d\epsilon^2.$$

Any inadmissible order reverses at least one true directed edge. For that edge, the corresponding estimated coefficient has magnitude at least $\beta_{\min} - \epsilon$, hence

$$L_{\widehat{B}}(\pi) \ge (\beta_{\min} - \epsilon)^2.$$

The stated strict inequality therefore makes every inadmissible order more costly than an admissible one, so every global minimizer is admissible. Taking probabilities immediately gives the second statement. □

The certificate is intentionally modular. A sharper finite-sample result can substitute any valid concentration inequality for the chosen ICA/residualization estimator. A distribution-free numerical δ is not claimed because finite-sample ICA accuracy depends on source non-Gaussianity and mixing geometry.

C.12 Scope and limits

The linear theorem is conditional on an admissible order; Theorem 3 broadens the measurement and latent-mechanism class only when bridge existence and completeness hold. Correlated view errors, invalid proxy grouping, hidden roots without measured blocks, weak instruments, or information-destroying nonlinear channels can still break identification. Proposition 3 gives a finite-sample ordering certificate once a coefficient-matrix error radius is available, but it does not manufacture a distribution-free FastICA error radius.

D ADDITIONAL EXPERIMENTS

D.1 METRICS AND SCORING CONVENTIONS

For a scored binary item set, $\mathrm{F1} = 2\mathrm{TP}/(2\mathrm{TP} + \mathrm{FP} + \mathrm{FN})$, with zero assigned to an empty denominator. Pair F1 scores unordered entries of $A \vee A^\top \vee B$; directed F1 scores ordered entries of A; bidirected F1 scores unordered entries of B; bow F1 scores ordered entries of $A \odot B$. Mixed SHD is

$$\sum_{i \neq j} |A^\star_{ij} - \widehat{A}_{ij}| + \sum_{i<j} |B^P_{ij} - \widehat{B}_{ij}|,$$

so a reversal costs two. "Any FP(A, M)" counts graphs with at least one falsely selected direct coefficient or latent incidence. FCI/RFCI native PAGs do not directly define exact directed, bidirected, bow, or mixed-SHD quantities, so those cells are reported as N/A. Whole-graph bootstrap intervals preserve method pairing and are descriptive.

D.2 EQUAL-INFORMATION MIXED-GRAPH DESIGN

The independent confirmation benchmark uses 40 graphs: 20 random mixed graphs and 20 graphs embedding the parent-collider motif. Every dataset has $d = 8$ observed variables, $K = 3$ hidden roots, $n = 1000$ records, and five pure repeated measurements per root. Each graph is observed at proxy-noise SD .35 and 1.5. Direct-only, hidden-only, bow, and absent relations all occur, and node labels are permuted before fitting. All equal-information procedures receive the same primary variables, raw proxy blocks, block grouping, and development/inference split.

D.3 GRAPH-FAMILY AND ORDER-LEARNING STRESS TESTS

The graph-family decomposition and the 240-fit order/view-count study are reported in main-text Tables 4 and 5. Their seed-level outputs are included in the reproducibility archive.

D.4 NONLINEAR STRESS TESTS

Table 6: Nonlinear observed-parent stress test for the fixed-basis extension: 40 new random mixed graphs with direct mechanisms $b_{ij}X_i + q_{ij}\tanh(X_i)$ and the prespecified basis $(X_i, \tanh X_i)$. Proxy-HCD-IV uses the balanced cross-view partition.

Sensor SD	Pair F1	Dir. F1	↔ F1	Bow F1	SHD	Order valid
.35	.961	.947	.987	.642	.700	39/40
1.50	.774	.729	.714	.185	4.375	33/40

The fixed-basis experiment changes observed direct mechanisms to $b_{ij}X_i + q_{ij}\tanh(X_i)$ and supplies the estimator the basis $(X_i, \tanh X_i)$, as allowed by Corollary 1. We additionally generated 20 graphs with nonlinear latent effects $\lambda_1 H + \lambda_2 \tanh H$ and nonlinear monotone proxy channels. With a fixed two-dimensional cross-view dictionary, directed recovery remains strong but hidden-incidence power degrades under noisy nonlinear measurements; the full seed-level results are included in the reproducibility archive. This is a finite-sieve stress test, not a claim that arbitrary nonlinear channels satisfy completeness.

D.5 WDBC-CALIBRATED SEMI-SYNTHETIC PROTOCOL

The semi-synthetic benchmark uses the 569-record, 30-feature Wisconsin Diagnostic Breast Cancer dataset (Wolberg et al., 1993) only to supply empirical exogenous distributions. Each real feature is standardized. Fixed disjoint feature columns are assigned to hidden-root marginals, primary-disturbance marginals, and proxy-error marginals. For every simulated record, these exogenous quantities are sampled independently with replacement from the corresponding empirical feature columns; this preserves non-Gaussian real-data marginal shapes while maintaining the independence assumptions of the structural model. The directed graph, root–node incidence matrix, coefficients,

and proxy assignment are generated independently of WDBC, so exact mixed-graph truth is known. Twenty graph seeds are evaluated at proxy SD .35 and 1.5. We call this *WDBC-calibrated semi-synthetic*, not a causal analysis of breast cancer.

The seed-level file `results/semisynthetic/wdbc_semisynthetic_perseed.csv` and the exact generator script are included with the submission. At high noise, Proxy-HCD-IV has directed F1 .869 versus .830 for the matched OLS control and reduces graphs with any false direct/incidence support from $17/20$ to $5/20$, while bidirected F1 decreases from .932 to .752.

D.6 FoundCause uncertainty and baseline implementation details

The full cross-method comparison is reported in main-text Table 1. FCI/RFCI/RCD retain their stated implementation provenance, and PAG outputs are not converted into unsupported exact ADMG scores. FoundCause is evaluated on the same primary-only benchmark. We rescore its saved prediction matrices with the common metric code. Table 7 reports the resulting means and graph-bootstrap uncertainty; per-seed predictions, thresholds, and hashes are included in the reproducibility archive.

Table 7: FoundCause rescored from all 20 uploaded prediction matrices. Percentile intervals use 10,000 whole-graph bootstrap resamples.

Metric	Mean	95% interval
Pair F1	0.725	[0.672, 0.780]
Directed F1	0.722	[0.666, 0.776]
Bidirected F1	0.229	[0.143, 0.337]
Bow F1	0.183	[0.092, 0.283]
Mixed SHD	11.500	[9.100, 13.900]
Bidirected average precision	0.427	[0.315, 0.548]
Bidirected AUROC	0.811	[0.738, 0.874]