IDENTIFYING A CANONICAL LATENT CAUSAL MODEL FROM MULTI-ENVIRONMENT MIXED OBSERVATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Scientific variables are often observed through heterogeneous measurements rather than directly: some coordinates are continuously recorded after unknown monotone distortions, whereas others are ordinal or binary coarsenings. We study a linear Gaussian latent DAG shared across environments, with invariant measurement mechanisms and environment-varying independent structural-noise variances. We show that measurement invariance aligns latent marginal scales across environments and identifies a common-reference latent covariance family. When the structural noises have pairwise distinct cross-environment variance profiles, this family identifies a canonical latent structural causal model and the full latent DAG. The identified canonical model also determines hard-intervention laws in observed environments; for strictly increasing measurements that may be discontinuous, we further characterize the sharp set of physical measured responses and the exact condition for point identification. We develop a mixed-data covariance front end and a regularized coefficient estimator with fixed-dimensional finite-sample guarantees. Across eight graph topologies, three sample sizes, and strong and weak heterogeneity regimes, recovery improves with sample size and variance-profile separation; under weak separation, our estimator attains higher mean area under the precision–recall curve than same-front direct diagonalization and BACKSHIFT. A BRFSS 2015 application illustrates real-data behavior through held-out covariance fit and bootstrap edge-selection stability.

1 INTRODUCTION

Mixed continuous, ordinal, and binary measurements are commonly handled through latent-Gaussian or copula representations, which make it possible to recover latent dependence or causal structure despite heterogeneous marginal scales (Fan et al., 2017; Cui et al., 2016; Cai et al., 2022). Latent-variable causal methods further use observable indicators and measurement-model constraints to recover hidden variables or their causal structure (Silva et al., 2006; Kummerfeld & Ramsey, 2016; Xie et al., 2024). These lines of work show that heterogeneous or indirect measurements need not preclude causal analysis, but they do not by themselves determine how the numerical scales of the same latent coordinates should be aligned across multiple environments.

Multi-environment methods obtain additional causal information from distributional changes, for example through covariance shifts or changing source variances (Rothenhäusler et al., 2015; Montagna, 2026). BACKSHIFT exploits covariance changes when variables are available on a common numerical scale, whereas recent multi-environment identification results cover more general nonlinear settings under smooth invertibility conditions. We study this combination when the same latent Gaussian DAG is observed across environments through invariant but mixed measurements: a continuous coordinate may undergo an unknown strictly increasing, possibly discontinuous transformation, whereas an ordinal or binary coordinate is threshold-coarsened. Under these observations, within-environment dependence can be recovered on a standardized scale, but the cross-environment latent marginal scales needed for variance-based causal identification are not directly available. Our first task is therefore to recover a common-reference latent covariance family from the mixed observations.

Multiple environments turn this scale ambiguity into information when the measurement mechanisms are invariant. Cross-environment distributional comparisons align the relative latent marginal scales, while mixed-pair dependence identifies the corresponding latent correlations. Together these quan-

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.

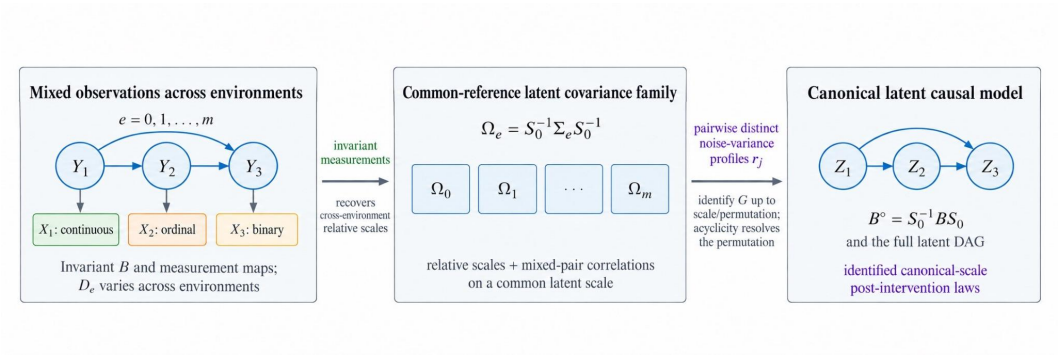


Figure 1: Population identification mechanism. Invariant measurements align relative latent scales across environments, producing the common-reference covariance family $\{\Omega_e\}$. Pairwise distinct structural-noise variance profiles identify the common source directions up to scale and permutation; acyclicity resolves the permutation and normalization yields B°, the full latent DAG, and the identified canonical-scale post-intervention laws.

tities place all environments on one common latent reference scale, yielding the covariance family needed by the causal backend.

Once the latent scales are aligned, cross-environment changes in structural-noise variances provide a source-separation signal. Sufficiently distinct variance profiles identify the common latent mixing directions up to scaling and permutation, and acyclicity resolves the remaining permutation. This yields a canonical coefficient representation and the full latent DAG. Figure 1 summarizes the population identification mechanism.

The identifiable target is a canonical structural causal model rather than the raw numerical coefficient matrix: positive rescalings of the latent coordinates alter raw coefficients but leave the canonical model and its DAG unchanged. This distinction also matters for interventions. Hard interventions on the canonical latent variables are identified in observed regimes, whereas a discontinuity in a continuous measurement map can leave a sharp set of possible physical measured responses at a directly targeted value.

We develop a mixed-data covariance estimator and a regularized coefficient estimator with separate graph thresholding, together with fixed-dimensional finite-sample guarantees. Empirically, recovery improves with sample size and variance-profile separation; in the weak simulation cells studied, the proposed estimator attains higher mean area under the precision–recall curve than same-front direct diagonalization and BACKSHIFT. A BRFSS 2015 application illustrates behavior on naturally occurring regional environments.

Contributions. Our contributions follow the identification chain. **(1)** Mixed continuous/ordinal observations with invariant measurement mechanisms identify a common-reference latent covariance family, including cross-environment relative scales. **(2)** Distinct structural-noise variance profiles identify the canonical latent structural model and full latent DAG. **(3)** The identified canonical model determines latent intervention laws in observed regimes, together with a sharp characterization of physical measured responses under possibly discontinuous monotone measurements. **(4)** A mixed-data estimator with separate graph thresholding admits fixed-dimensional covariance, coefficient, and support-recovery guarantees.

2 Related Work

Mixed measurements and latent measurement models. Semiparametric mixed-data methods recover latent Gaussian dependence or precision structure (Fan et al., 2017), while related work learns latent tree structure (Zhou et al., 2026) or causal graphs from mixed observations (Cui et al., 2016; Cai et al., 2022). A complementary measurement-model literature exploits tetrad, rank, and related constraints to locate latent factors or clusters from multiple indicators (Silva et al., 2006;

Kummerfeld & Ramsey, 2016; Sullivant et al., 2010); higher-order or non-Gaussian conditions such as Triad and GIN can further orient latent structure (Cai et al., 2019; Xie et al., 2024), and tensor-rank conditions address discrete latent structures (Chen et al., 2024). These results solve important measurement and latent-structure subproblems, but the statistical object needed here is different. Recovering a latent correlation matrix within each environment does not determine how latent marginal scales compare across environments. Our observation layer uses invariance of the coordinatewise measurement mechanism to identify those relative scales together with mixed-pair correlations, thereby constructing the common-reference covariance family $\{\Omega_e\}$ required for causal identification.

Heterogeneous environments and variance-based source separation. BACKSHIFT identifies observed-scale linear connectivity from covariance differences under unknown shift interventions, requiring sufficiently nonproportional intervention-variance changes (Rothenhäusler et al., 2015). Our backend instead separates sources from the *absolute* common-reference family $\{\Omega_e\}$ through reference-normalized structural-noise profiles. This can retain information lost by differencing: variance sequences $(1, 2, 3)$ and $(2, 3, 4)$ yield distinct normalized profiles $(2, 3)$ and $(3/2, 2)$ but identical difference signatures $(1, 2)$. Thus, despite shared simultaneous-diagonalization algebra, the two approaches use different information. Other methods exploit interventional distributions, target or invariance information, soft-intervention structure, bijective mechanisms, or cross-view dependence (Squires et al., 2020; Brouillard et al., 2020; Jaber et al., 2020; Jalaldoust et al., 2025; Heurtebise et al., 2026). Our contribution is an end-to-end identification chain enabling variance-based source separation after mixed, partly noninvertible measurements and recovering the canonical coefficient target induced by the remaining latent-scale symmetry.

Invertible multi-environment identification and intervention semantics. Montagna (2026) establishes graph identification for smooth invertible nonlinear acyclic SCMs under Gaussian source rescaling, faithfulness, and sufficient variability, resolving scaled-permutation ambiguity through acyclicity. Related work identifies functional mechanisms, counterfactuals, or latent causal structure under invertible transformations (Montagna & Locatello, 2026; von Kügelgen et al., 2023; Squires et al., 2023; Varici et al., 2024; 2025). Our setting instead allows ordinal threshold coarsenings and strictly increasing continuous-coordinate measurements without continuity or smoothness assumptions. From the mixed observation law, we identify a common-reference linear SCM and distinguish canonical latent interventions, which are identified, from directly targeted physical continuous responses, which may admit only the sharp interval $[L_j(z), U_j(z)]$.

3 Problem Setup and Canonical Target

Environments and latent SCM. Let $E \in \{0, \dots, m\}$ be the observed environment set, with $e = 0$ serving as the reference environment. For each environment e, let $Y^{(e)} = (Y_1^{(e)}, \dots, Y_p^{(e)})^\top \in \mathbb{R}^p$ denote the latent causal variables. Conditional on $E = e$, they follow the linear Gaussian structural equation model

$$Y^{(e)} = BY^{(e)} + \eta^{(e)}, \qquad \eta^{(e)} \sim \mathcal{N}(0, D_e), \qquad D_e = \mathrm{diag}(d_{e1}, \dots, d_{ep}) \succ 0. \tag{1}$$

Here $B \in \mathbb{R}^{p \times p}$ is invariant across environments, $B_{ii} = 0$, and $B_{ij} \neq 0$ represents the directed edge $j \to i$. We assume that this directed graph is acyclic. The vector $\eta^{(e)}$ contains the mutually independent structural noises in environment e, and $d_{ej} > 0$ is the variance of its jth component. Let I denote the $p \times p$ identity matrix. With $A := (I - B)^{-1}$, the latent covariance matrix is $\Sigma_e := \mathrm{Cov}(Y^{(e)} \mid E = e) = AD_eA^\top$.

Invariant mixed measurements. We observe $X^{(e)} = (X_1^{(e)}, \dots, X_p^{(e)})$, where each observed coordinate measures the corresponding latent coordinate and no observed coordinate mixes multiple latent variables. Throughout, $[p] := \{1, \dots, p\}$. Let $\mathcal{C}$ and $\mathcal{O}$ denote the continuous and ordinal coordinate sets, respectively, with binary coordinates included in $\mathcal{O}$. Let $[p] = \mathcal{C} \cup \mathcal{O}$ and $\mathcal{C} \cap \mathcal{O} = \varnothing$. For each $j \in \mathcal{O}$, let $K_j \geq 2$ be the number of ordered categories, with $K_j = 2$ corresponding to a binary coordinate. The observation is assumed to be generated by environment-invariant thresholds, using the fixed convention $X_j^{(e)} = k$ if and only if $\gamma_{j,k-1} < Y_j^{(e)} \leq \gamma_{jk}$ for $k = 1, \dots, K_j$, where $-\infty = \gamma_{j0} < \gamma_{j1} < \cdots < \gamma_{j,K_j-1} < \gamma_{jK_j} = +\infty$. For binary coordinates ($K_j = 2$), we

assume that the unique finite threshold is nonzero; for $K_j \geq 3$, strict threshold ordering ensures that at least one internal threshold is nonzero. We also assume $P(E = e) > 0$ for every observed environment(H1).

Reference scale and canonical target. For each coordinate and environment, let $s_{ej} := \sqrt{\Sigma_{e,jj}} > 0$ denote the latent marginal standard deviation, and define $S_0 := \mathrm{diag}(s_{01}, \ldots, s_{0p})$ from the reference environment. We express all latent coordinates in units of their reference-environment standard deviations. The resulting reference-scale latent variables, covariance matrices, and structural coefficient matrix are

$$Z^{(e)} := S_0^{-1} Y^{(e)}, \qquad \Omega_e := S_0^{-1} \Sigma_e S_0^{-1}, \qquad B^\circ := S_0^{-1} B S_0. \tag{2}$$

Thus $\Omega_e = \mathrm{Cov}(Z^{(e)} \mid E = e)$ and $\mathrm{diag}(\Omega_0) = \mathbf{1}$. The matrix B° expresses the same latent DAG as B, but with every latent coordinate measured on the common reference scale. Positive coordinatewise rescaling of the raw latent variables changes B by positive diagonal similarity while leaving B° unchanged.

Cross-environment variance profiles. For $e = 1, \ldots, m$ and $j \in [p]$, define the reference-normalized noise variance as $r_{ej} := d_{ej}/d_{0j}$ and collect the nonreference values for coordinate j into the profile

$$r_j := (r_{1j}, \ldots, r_{mj})^\top = (d_{1j}/d_{0j}, \ldots, d_{mj}/d_{0j})^\top \in \mathbb{R}^m. \tag{3}$$

H2 (variance-profile separation). The profiles are pairwise distinct: $r_j \neq r_k, \text{ for every } j \neq k.$

Thus no two structural-noise coordinates have the same pattern of variance changes relative to the reference environment. Section 4.2 shows how H2 separates the latent source directions once the common-reference covariance family $\{\Omega_e\}_{e=0}^m$ has been identified.

4 Population Identification

4.1 Mixed observations identify a common reference scale

We write $P(X, E)$ for the joint observational law, where $X = X^{(E)}$ is the observed mixed vector in the realized environment. Let Φ denote the standard normal CDF and define $F_{0j}(x) := P(X_j \leq x \mid E = 0)$ for each continuous coordinate $j \in \mathcal{C}$. The common-reference covariance family is identified in three steps. First, for $j \in \mathcal{C}$, the reference marginal CDF fixes the latent reference scale: the transformation $h_j := \Phi^{-1} \circ F_{0j}$ satisfies $h_j(X_j^{(e)}) = Y_j^{(e)}/s_{0j}$ a.s. Thus the relative scale $\alpha_{ej} := s_{ej}/s_{0j}$ is identified by $\alpha_{ej}^2 = \mathrm{Var}\{h_j(X_j^{(e)})\}$. Second, for an ordinal coordinate $j \in \mathcal{O}$, define $\delta_{ejk} := \Phi^{-1}\{P(X_j \leq k \mid E = e)\} = \gamma_{jk}/s_{ej}, k = 1, \ldots, K_j - 1$. Any nonzero invariant threshold γ_{jk} then identifies $\alpha_{ej} = \delta_{0jk}/\delta_{ejk}$. Such a threshold exists by the binary scale-visibility condition in Section 3 when $K_j = 2$, and by strict ordering of the internal thresholds when $K_j \geq 3$. Third, each latent correlation $\rho_{e,j\ell} := \mathrm{Corr}(Y_j^{(e)}, Y_\ell^{(e)})$, $j \neq \ell$, is identified by inverting the corresponding strictly monotone bivariate Gaussian bridge for continuous–continuous, continuous–ordinal, or ordinal–ordinal pairs. Combining these correlations with the identified relative scales gives $\Omega_{e,jj} = \alpha_{ej}^2, \Omega_{e,j\ell} = \alpha_{ej}\alpha_{e\ell}\rho_{e,j\ell}$, placing all environments on a common reference scale. Appendix B provides the bridge formulas and proof.

Proposition 1 (Common-reference covariance identification)**.** *Under the Gaussian latent SEM and invariant mixed-measurement assumptions of Section 3, including binary scale visibility and $P(E = e) > 0$ for every observed environment, $P(X, E)$ uniquely determines $\{\Omega_e\}_{e=0}^m$. Moreover, $\{\Sigma_e\}_{e=0}^m$ is identified only up to one positive diagonal congruence shared by all environments.*

The proposition is the observation-to-causal interface: measurement invariance upgrades environment-specific dependence information to an *aligned covariance family*. Absolute latent units remain arbitrary, but the cross-environment relative scales required by the causal backend are identified.

4.2 Main identification theorem: from variance signatures to the full DAG

Define the reference-scale mixing matrix $G := S_0^{-1}(I - B)^{-1}$, so that $\Omega_e = G D_e G^\top$. Let $\Omega_0^{-1/2}$ denote the symmetric positive-definite inverse square root of Ω_0, and define $O := \Omega_0^{-1/2} G D_0^{1/2}$.

Because $\Omega_0 = GD_0G^\top$, the matrix O is orthogonal. Whitening the covariance family therefore gives

$$H_e := \Omega_0^{-1/2}\Omega_e\Omega_0^{-1/2} = O\ \mathrm{diag}(r_{e1}, \ldots, r_{ep})\, O^\top. \tag{4}$$

Thus the profile r_j defined in Section 3 is the cross-environment variance signature of structural noise j. Under H2, the signatures are pairwise distinct, so the common eigendirections are unique up to signed permutation. Equivalently, the columns of G are determined up to permutation and nonzero scaling. The remaining issue is causal labeling: which recovered source column belongs to which structural equation.

Theorem 2 (Canonical latent causal-model identification)**.** *Assume the model in Section 3 and H2, i.e., the normalized structural-noise variance profiles $\{r_j\}_{j=1}^p$ are pairwise distinct. Then $P(X, E)$ identifies G up to column permutation and nonzero diagonal scaling, and acyclicity uniquely removes the source permutation. Consequently, $P(X, E)$ uniquely identifies $B^\circ = S_0^{-1}BS_0$ and $\mathrm{DAG}(B) = \mathrm{DAG}(B^\circ)$. Any two observationally equivalent models in this class have structural matrices related by positive diagonal similarity.*

Write $G_0 := \mathrm{DAG}(B) = \mathrm{DAG}(B^\circ)$ for the true directed graph. Figure 1 summarizes the resulting identification chain. Acyclicity is used only to align the recovered source columns with structural equations. Among row permutations of an inverse-mixing representative, the correct permutation is the unique one for which all diagonal entries are nonzero; any competing permutation would induce a directed cycle. Row normalization to unit diagonal then recovers $I - B^\circ$. Positive coordinatewise latent rescaling is the remaining and sharp observational symmetry, so B°—rather than the raw numerical matrix B—is the canonical exact coefficient target. Appendix C gives the full source-separation, permutation-removal, and sharpness arguments.

4.3 Canonical interventions and physical measured responses

Canonical latent interventions. The identified pair (B°, Ω_e) determines the regime-specific canonical structural-noise covariance and SCM:

$$D_e^\circ := (I - B^\circ)\Omega_e(I - B^\circ)^\top, \qquad Z^{(e)} = B^\circ Z^{(e)} + \epsilon^{(e),\circ}, \quad \epsilon^{(e),\circ} \sim \mathcal{N}(0, D_e^\circ). \tag{5}$$

Let $\mathcal{S} \subseteq [p]$ be a target set, $\mathcal{R} := [p] \setminus \mathcal{S}$ its complement, and $a_\mathcal{S} \in \mathbb{R}^{|\mathcal{S}|}$ the target values. Subscripts such as $B^\circ_{\mathcal{R}\mathcal{R}}$ and $B^\circ_{\mathcal{R}\mathcal{S}}$ denote the corresponding row/column blocks. In an observed regime e, the hard intervention $do(Z_\mathcal{S} = a_\mathcal{S})$ gives

$$Z_\mathcal{R} = (I - B^\circ_{\mathcal{R}\mathcal{R}})^{-1}\big(B^\circ_{\mathcal{R}\mathcal{S}}a_\mathcal{S} + \epsilon^{(e),\circ}_\mathcal{R}\big), \tag{6}$$

so the full latent Gaussian post-intervention law is point identified. The same holds for a specified mixture of observed regimes; an unseen regime additionally requires its structural-noise covariance to be specified.

Physical measured responses. Physical measurement introduces a different identification question. A continuously distributed latent coordinate assigns zero observational probability to any single point. Consequently, if a strictly increasing measurement map has a jump, different values of the map at that latent point can be observationally indistinguishable even though a hard intervention places mass exactly there. To state the resulting boundary, we use an *intervention-invariant measurement mechanism* (IIMM): interventions replace the targeted latent structural equations while leaving all measurement maps and thresholds unchanged. IIMM is an intervention-semantics assumption only and is not required for observational identification.

For $j \in \mathcal{C}$, define the canonical measurement map $m_j(z) := g_j(s_{0j}z)$ and

$$L_j(z) := \inf\{x : F_{0j}(x) \geq \Phi(z)\}, \qquad U_j(z) := \inf\{x : F_{0j}(x) > \Phi(z)\}. \tag{7}$$

For any functional ψ, let $\mathrm{IdSet}\{\psi \mid P(X, E)\}$ denote the set of values of ψ over all models in the assumed class that induce the same observational law $P(X, E)$.

For $j \in \mathcal{O}$, recall $\delta_{0jk} = \gamma_{jk}/s_{0j}$ and define $\kappa_j(a)$ as the unique $k \in \{1, \ldots, K_j\}$ satisfying $\delta_{0,j,k-1} < a \leq \delta_{0jk}$. Thus a hard target $do(Z_j = a)$ has deterministic ordinal response $X_j = \kappa_j(a)$; in particular, $a = \delta_{0jk}$ is assigned to category k under the fixed endpoint convention.

Theorem 3 (Sharp physical response set)**.** *For any $j \in \mathcal{C}$ and finite z, with m_j strictly increasing,*

$$\mathrm{IdSet}\{m_j(z) \mid P(X, E)\} = [L_j(z), U_j(z)].$$

Hence the physical value is point identified if and only if $L_j(z) = U_j(z)$. Under IIMM, the full physical mixed-X post-intervention law is uniquely identified if and only if this interval collapses for every directly targeted continuous coordinate. Under the fixed ordinal endpoint convention above, ordinal targets and untargeted coordinates require no analogous collapse.

Thus canonical latent interventions and physical measured responses have different identification boundaries under noninvertible measurement. The lower-quantile version $q_j(z) := F_{0j}^{-1}(\Phi(z)) = L_j(z)$ is always an identified canonical observational version; the interval above is the exact physical-response identified set when pointwise measurement ambiguity remains. Appendix D proves sharpness by admissible point modifications on Gaussian-null sets.

5 Estimation and Finite-Sample Guarantees

For each environment e, suppose we observe n_e i.i.d. samples $X_1^{(e)}, \ldots, X_{n_e}^{(e)}$, and define $n_{\min} := \min_{0 \le e \le m} n_e$. Let $\widehat{\alpha}_{ej}$ denote the estimated relative marginal scale $\alpha_{ej} = s_{ej}/s_{0j}$ and set $\widehat{Q}_e := \mathrm{diag}(\widehat{\alpha}_{e1}, \ldots, \widehat{\alpha}_{ep})$. Let $\widehat{R}_e$ denote the estimated latent correlation matrix obtained from the rank-based bridge inversions followed by the correlation-matrix projection. The mixed-data front end then assembles $\widehat{\Omega}_e = \widehat{Q}_e \widehat{R}_e \widehat{Q}_e$. Appendix F specifies the totalized empirical transforms, bridge inversions, clipping/flooring conventions, and correlation-matrix projection that define these estimators on the whole sample space.

For coefficient estimation, write $M := I - B^\circ$, which has unit diagonal. For a matrix M, let $M_{j\cdot}$ denote its jth row, and define the off-diagonal ℓ_1 norm by $\|M\|_{1,\mathrm{off}} := \sum_{i \ne j} |M_{ij}|$. For fixed criterion weights $w_e > 0$ satisfying $\sum_{e=0}^m w_e = 1$, define

$$\widehat{Q}_{\mathrm{prof}}(M) = \sum_{e=0}^{m} w_e \left[\sum_{j=1}^{p} \log\Big(M_{j\cdot} \widehat{\Omega}_e M_{j\cdot}^\top \Big) - 2 \log |\det M| \right]. \tag{8}$$

For a regularization parameter $\lambda_n > 0$, the coefficient estimator (C1 estimator) is

$$\widehat{M} \in \arg \min_{\substack{\mathrm{diag}(M) = \mathbf{1} \\ \det(M) \ne 0}} \left\{ \widehat{Q}_{\mathrm{prof}}(M) + \lambda_n \|I - M\|_{1,\mathrm{off}} \right\}, \qquad \widehat{B}^\circ := I - \widehat{M}, \tag{9}$$

where $\mathbf{1}$ is the p-vector of ones. Thus C1 estimator is a covariance plug-in criterion with a separate sparsity penalty on the structural coefficients.

Graph extraction is separated from coefficient estimation. For a threshold $\tau_n > 0$, define

$$\widehat{G}_{\tau_n} = \{ j \to i : |\widehat{B}^\circ_{ij}| > \tau_n \}. \tag{10}$$

For the finite-sample statements below, $\|\cdot\|_{\mathrm{op}}$ denotes the spectral/operator norm and $\|\cdot\|_{\max}$ the elementwise maximum norm. The detailed high-probability bounds in Appendix E use a failure probability $\delta \in (0, 1)$.

Theorem 4 (Finite-sample summary)**.** *Under the corresponding conditions detailed in Appendix E, the following hold.*

1. ***Covariance-family estimation (F1).*** *Under the conditions of Lemma E.1, for fixed δ, p, m, $\max_e \|\widehat{\Omega}_e - \Omega_e\|_{\mathrm{op}} = O_p(n_{\min}^{-1/2} + n_0^{-1})$.*
2. ***Coefficient estimation (F2).*** *Under H2 and the finite-sample conditions of Lemma E.2, any fixed measurable global-minimizer selection satisfies $\|\widehat{B}^\circ - B^\circ\|_{\max} = O_p(n_{\min}^{-1/2} + n_0^{-1} + \lambda_n)$.*
3. ***Graph recovery (F3).*** *Under the conditions of Lemma E.3, $\Pr(\widehat{G}_{\tau_n} = G_0) \to 1$.*

Appendix E gives the full F1–F3 statements and supporting derivations, including explicit high-probability radii, localization and curvature conditions, tuning requirements, beta-min conditions, and pointwise versus polynomial-tail envelopes.

6 EXPERIMENTS

6.1 EXPERIMENTAL SETUP

The synthetic design uses $p = 10$ latent variables and five environments with an invariant acyclic coefficient matrix of expected degree 2. Variables V1–V5 are continuous, V6–V7 are binary, and V8–V10 are ordinal. Per-environment sample sizes are $n_e \in \{500, 1500, 5000\}$, and the environment-variance profiles are evaluated under strong and weak separation settings $\delta = 0.50$ and $\delta = 0.12$. Graphs satisfy fixed pre-data conditioning, beta-min, and profile-separation requirements before any recovery evaluation. Appendix G gives the exact coefficient range, variance-profile codebook, measurement maps, thresholds, and graph-generation rules. Figure 2 shows the eight frozen graph topologies used throughout the simulation study.

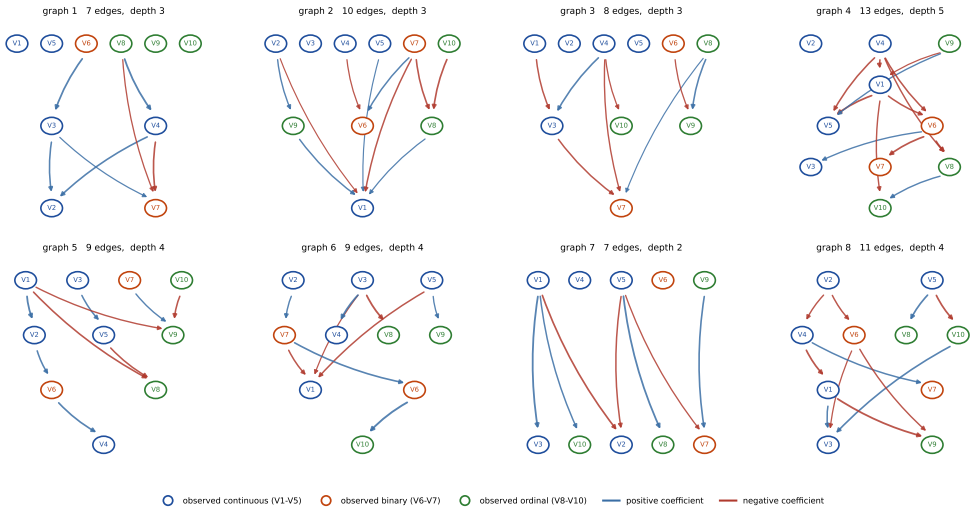


Figure 2: The eight frozen $p = 10$ ground-truth DAGs used in the simulation study, drawn by topological level. Blue, orange, and green nodes denote continuous, binary, and ordinal observed variables, respectively; blue and red edges denote positive and negative coefficients.

6.2 METHODS AND COMPARISONS

The proposed estimator is C1 estimator applied after the estimated mixed-data front end. AJD is an internal same-front backend ablation: it receives the same estimated family $\{\widehat{\Omega}_e\}$ as C1 estimator. Corrected BACKSHIFT is an external multi-environment same-front covariance backend that also receives $\{\widehat{\Omega}_e\}$; it is not a raw-mixed-X pipeline (Rothenhäusler et al., 2015). In contrast, latent-PC and Copula PC are native mixed-data baselines operating on raw mixed $X^{(0)}$ in the reference environment and are compared only with graph-output-compatible metrics. Oracle diagnostics receive either the true canonical Z or the true covariance family $\{\Omega_e\}$ and are labeled as information-advantaged diagnostics. Additional native mixed-data, oracle, and specialization comparisons are reported in Appendix H.

6.3 METRICS

Canonical-coefficient error is reported only for methods that estimate B°. AUPRC and AUROC are used only when a method supplies an authorized continuous edge-ranking score; fixed CPDAG or I-MEC outputs are evaluated with their corresponding skeleton/orientation or equivalence-class metrics rather than coerced into score curves or fully directed DAGs. The coefficient-based methods also use the fixed threshold-sensitivity path specified in Appendix G.

6.4 PRIMARY STRUCTURAL RECOVERY

Figure 3 reports coefficient AUPRC for estimated-front C1 estimator, same-front AJD, and corrected same-front BACKSHIFT. Recovery improves with sample size in both profile regimes, while weak variance-profile separation is substantially harder. In the difficult weak $n = 500$ cell, mean AUPRC is 0.644 for C1 estimator (7/8 successful graphs), versus 0.209 for AJD and 0.171 for BACKSHIFT (8/8 each); at $n = 5000$, the corresponding means are 0.963, 0.555, and 0.654. Under strong heterogeneity, the diagonalization-style backends become competitive as n grows. Three C1 estimator runs have CV tuning failure: G2 at $n = 500$ in both regimes and G8 at $n = 500$ in the strong regime, so the strong and weak C1 estimator means at $n = 500$ use 6/8 and 7/8 graphs, respectively. Graph-level points show variation across G1–G8, not repeated-sampling uncertainty.

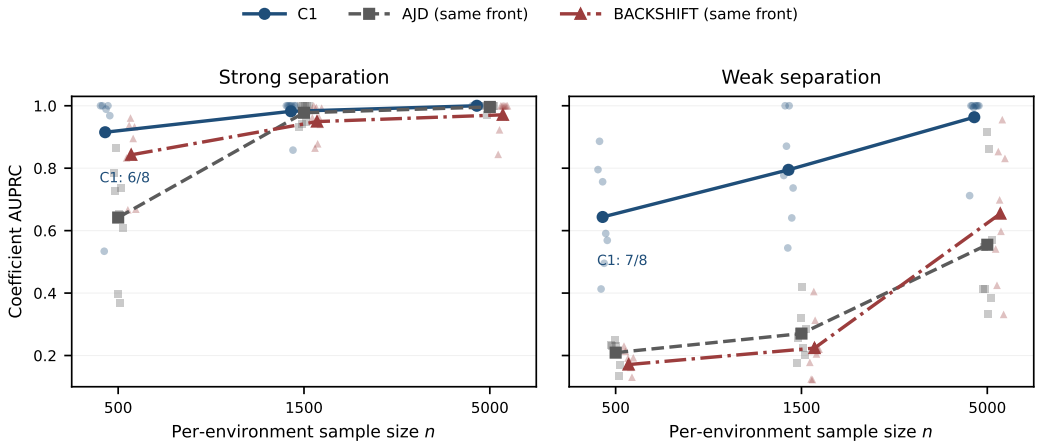


Figure 3: Coefficient AUPRC under strong and weak variance-profile separation. Lines show means across the eight frozen graph topologies and faint points show graph-level results. C1 estimator uses 6/8 graphs for strong $n = 500$ and 7/8 for weak $n = 500$ because failed CV runs are excluded rather than zero-imputed; all remaining cells use 8/8 graphs.

Native mixed-data baselines that use only raw mixed $X^{(0)}$ in the reference environment are reported in Appendix H (Table 4). Because they do not use cross-environment heterogeneity, their skeleton F1 is not a coefficient-AUPRC or full-DAG ranking against the same-front covariance backends.

For the same successful C1 estimator cases in the weak $n = 500$ cell, replacing the estimated covariance family by Oracle-Ω raises mean AUPRC from 0.644 to 0.937 (7/8 in both), indicating that mixed-front-end covariance error contributes materially in this difficult finite-sample regime; Oracle-Ω is an information-advantaged diagnostic, not an external competitor or universal upper bound. For C1 estimator, Oracle-Ω reuses the paired λ selected for the corresponding estimated-front run where applicable; it is not independently retuned (Appendix Table 5).

6.5 BRFSS 2015 APPLICATION

We apply the fitted pipeline to the public 2015 CDC BRFSS data (Centers for Disease Control and Prevention, 2016). After the frozen recoding and geographic exclusions, the 22-variable analysis contains 249,350 observations across Northeast, Midwest, South, and West, treated as naturally occurring environments with Northeast as reference; variable typing follows Zhou et al. (2026). Because BMI is rounded and MentHlth and PhysHlth are tied day counts, the atomless continuous-margin model is only approximate here, and the regional common-B condition remains a modeling assumption rather than an empirical conclusion.

The real-data analysis uses the adaptive ordinal-correlation cutpoint procedure detailed in Appendix I. Cross-validation over the prespecified c_λ grid selects 0.25, its smallest value. C1 estimator and AJD both succeed on the main/holdout fits and all 20 regional bootstrap refits; their held-out covariance losses are -5.814457 and -5.750857, respectively. At the illustrative threshold $\tau = 0.20$, C1

estimator has 17 directed edges, 13 with bootstrap selection frequency at least 0.80; Figure 4 summarizes their coefficient magnitudes and selection stability. These results are not presented as validated medical causal effects.

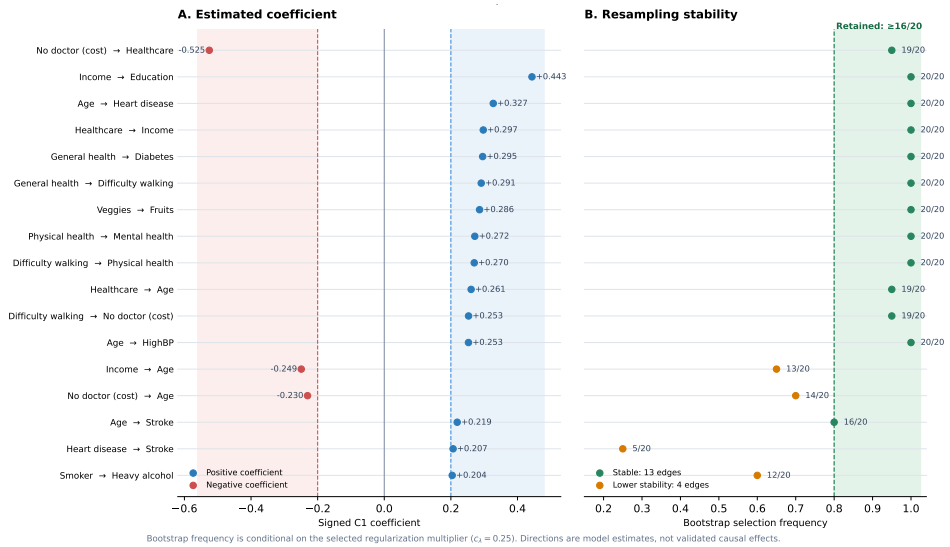


Figure 4: C1 estimator coefficient magnitudes and bootstrap selection frequencies for the 17 BRFSS edges selected at $\tau = 0.20$. Panel A shows signed canonical coefficients and panel B shows selection frequencies over 20 respondent-stratified regional bootstrap fits. Bootstrap frequency measures procedure-conditional edge-selection stability, not a confidence interval for a population causal effect; directions are model estimates rather than medically validated causal effects.

7 Discussion and Conclusion

The main scientific result is an end-to-end identification chain from mixed observations to a canonical latent causal model. Environment-invariant continuous maps and ordinal thresholds recover the relative latent scales needed to align covariance information, and environment-specific structural-noise variance signatures then identify B° and the full latent DAG. The same identified object supports regime-specific interventions on the canonical latent variables and a sharp characterization of their physical measured responses.

The role of measurement invariance is especially important: information that is unavailable inside any one environment becomes identifiable through comparison across environments. This is stronger than recovering a latent correlation matrix separately in each environment, because the causal backend requires the covariance family to live on a common numerical scale. In that sense, the measurement and heterogeneity assumptions work together rather than as two independent modeling devices.

Two ambiguities are intrinsic to the observation model: positive latent-coordinate rescaling leaves raw B unidentified while preserving the canonical target B°, and jumps in a monotone continuous measurement can leave a directly targeted physical value set- rather than point-identified. The theory assumes invariant one-coordinate-per-latent measurement, diagonal structural-noise covariance, pairwise variance-profile separation, and fixed-dimensional regularity for the stated rates. Within this scope, mixed and partly noninvertible measurements still identify a full reference-scale latent SCM, its DAG, and its intervention semantics. Finite-sample experiments, without changing the population claims, show that weak heterogeneity needs larger samples and front-end error can matter (Oracle-Ω); BRFSS uses an adaptive correlation cutpoint rather than the frozen simulation front end. BRFSS has no causal ground truth; common-B invariance is assumed; bootstraps assess selection stability, not confidence in causal effects.

AI USE STATEMENT

In this work, we used generative AI tools, including ChatGPT, to assist with three types of tasks: language editing and polishing of the manuscript; literature retrieval and discovery of potentially relevant prior work; and partial mathematical derivation, proof development, and proof checking.

All AI-assisted work was reviewed and verified by the authors. Mathematical arguments and theoretical claims were independently checked by the authors, and bibliographic information and citation-related claims were checked against the relevant source materials. AI-assisted language was reviewed and revised by the authors before inclusion in the manuscript. Synthetic datasets used in the experiments were generated by the specified data-generating processes and computational code, rather than by generative AI.

Generative AI tools did not make final scientific decisions or replace author verification of mathematical claims, experimental results, or cited sources. We take responsibility for the final content of this work.

ETHICS STATEMENT

The theoretical results do not require collection of new human-subject data. The BRFSS application is a secondary analysis of publicly released CDC survey data (Centers for Disease Control and Prevention, 2016); no new data were collected for this work. Our regional analysis is unweighted and model-based and is not presented as survey-population causal inference.

REPRODUCIBILITY STATEMENT

The main text states the model, identifiable target, population identification results, estimator, and headline rates. The appendix provides the population proofs, finite-sample derivations and conditions, specifies the mixed-data estimator and numerical C1 estimator implementation, and records the V2 synthetic data-generating process, graph-acceptance rules, method roles, and metric semantics. It also records the BRFSS adaptive cutpoint train/freeze rule, CV grid, region construction, threshold path, and bootstrap semantics. Code will be made publicly available upon cleanup.

REFERENCES

Philippe Brouillard, Sébastien Lachapelle, Alexandre Lacoste, Simon Lacoste-Julien, and Alexandre Drouin. Differentiable causal discovery from interventional data. In *Advances in Neural Information Processing Systems*, volume 33, pp. 21865–21877, 2020.

Ruichu Cai, Feng Xie, Clark Glymour, Zhifeng Hao, and Kun Zhang. Triad constraints for learning causal structure of latent variables. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Zhanrui Cai, Dong Xi, Xuan Zhu, and Runze Li. Causal discoveries for high dimensional mixed data. *Statistics in Medicine*, 41(24):4924–4940, 2022. doi: 10.1002/sim.9544.

Centers for Disease Control and Prevention. 2015 BRFSS survey data and documentation. `https://www.cdc.gov/brfss/annual_data/annual_2015.html`, 2016. Behavioral Risk Factor Surveillance System; 2015 survey data released August 2016; accessed 2026-09-24.

Zhengming Chen, Ruichu Cai, Feng Xie, Jie Qiao, Anpeng Wu, Zijian Li, Zhifeng Hao, and Kun Zhang. Learning discrete latent variable structures with tensor rank conditions. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Ruifei Cui, Perry Groot, and Tom Heskes. Copula pc algorithm for causal discovery from mixed data. In *Machine Learning and Knowledge Discovery in Databases*, pp. 377–392. Springer, 2016. doi: 10.1007/978-3-319-46227-1_24.

Jianqing Fan, Han Liu, Yang Ning, and Hui Zou. High dimensional semiparametric latent graphical model for mixed data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(2):405–421, 2017. doi: 10.1111/rssb.12168.

Ambroise Heurtebise, Omar Chehab, Pierre Ablin, Alexandre Gramfort, and Aapo Hyvärinen. Multi-view causal discovery without non-gaussianity: Identifiability and algorithms. In *International Conference on Machine Learning*, 2026.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830.

Biwei Huang, Kun Zhang, Jiji Zhang, Joseph Ramsey, Ruben Sanchez-Romero, Clark Glymour, and Bernhard Schölkopf. Causal discovery from heterogeneous/nonstationary data. *Journal of Machine Learning Research*, 21(89):1–53, 2020.

Amin Jaber, Murat Kocaoglu, Karthikeyan Shanmugam, and Elias Bareinboim. Causal discovery from soft interventions with unknown targets: Characterization and learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 9551–9561, 2020.

Kasra Jalaldoust, Saber Salehkaleybar, and Negar Kiyavash. Multi-domain causal discovery in bijective causal models. In *Proceedings of the Fourth Conference on Causal Learning and Reasoning*, volume 275 of *Proceedings of Machine Learning Research*, pp. 1268–1289, 2025.

Erich Kummerfeld and Joseph Ramsey. Causal clustering for 1-factor measurement models. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1655–1664, 2016. doi: 10.1145/2939672.2939838.

Pascal Massart. The tight constant in the dvoretzky–kiefer–wolfowitz inequality. *The Annals of Probability*, 18(3):1269–1283, 1990. doi: 10.1214/aop/1176990746.

Francesco Montagna. On the identifiability of causal graphs with multiple environments. In *International Conference on Learning Representations*, 2026.

Francesco Montagna and Francesco Locatello. Causal learning with the invariance principle. *arXiv preprint arXiv:2605.13589*, 2026.

Ronan Perry, Julius von Kügelgen, and Bernhard Schölkopf. Causal discovery in heterogeneous environments under the sparse mechanism shift hypothesis. In *Advances in Neural Information Processing Systems*, volume 35, pp. 10904–10917, 2022. doi: 10.52202/068431-0792.

R. L. Plackett. A reduction formula for normal multivariate integrals. *Biometrika*, 41(3–4):351–360, 1954. doi: 10.1093/biomet/41.3-4.351.

Dominik Rothenhäusler, Christina Heinze, Jonas Peters, and Nicolai Meinshausen. Backshift: Learning causal cyclic graphs from unknown shift interventions. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Ricardo Silva, Richard Scheines, Clark Glymour, and Peter Spirtes. Learning the structure of linear latent variable models. *Journal of Machine Learning Research*, 7:191–246, 2006.

Chandler Squires, Yuhao Wang, and Caroline Uhler. Permutation-based causal structure learning with unknown intervention targets. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence*, volume 124 of *Proceedings of Machine Learning Research*, pp. 1039–1048, 2020.

Chandler Squires, Anna Seigal, Salil S. Bhate, and Caroline Uhler. Linear causal disentanglement via interventions. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 32540–32560, 2023.

Seth Sullivant, Kelli Talaska, and Jan Draisma. Trek separation for gaussian graphical models. *The Annals of Statistics*, 38(3):1665–1685, 2010.

Burak Varici, Emre Acartürk, Karthikeyan Shanmugam, and Ali Tajer. General identifiability and achievability for causal representation learning. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 2314–2322, 2024.

Burak Varici, Emre Acartürk, Karthikeyan Shanmugam, Abhishek Kumar, and Ali Tajer. Score-based causal representation learning: Linear and general transformations. *Journal of Machine Learning Research*, 26(112):1–90, 2025.

Julius von Kügelgen, Michel Besserve, Liang Wendong, Luigi Gresele, Armin Kekić, Elias Bareinboim, David M. Blei, and Bernhard Schölkopf. Nonparametric identifiability of causal representations from unknown interventions. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Feng Xie, Biwei Huang, Zhengming Chen, Ruichu Cai, Clark Glymour, Zhi Geng, and Kun Zhang. Generalized independent noise condition for estimating causal structure with latent variables. *Journal of Machine Learning Research*, 25(191):1–61, 2024.

Can Zhou, Nan Li, Shuai Wang, Xiaofei Wang, and Jianhua Guo. Learning semi-parametric tree models from mixed data. *Artificial Intelligence*, 353:104499, 2026. doi: 10.1016/j.artint.2026.104499.

A Additional Notation and Assumptions

We collect the assumptions used by the main results and keep their logical roles separate.

Observational population layer. (S1) B is invariant across environments, has zero diagonal, and defines a DAG under the convention $B_{ij} \neq 0 \Leftrightarrow j \to i$. (S2) Conditional on environment e, the structural noise is centered Gaussian with positive diagonal covariance D_e. (O1) Each observed coordinate measures the corresponding latent coordinate, with no observed coordinate mixing multiple latent variables. (O2) Every continuous measurement map and every ordinal threshold vector is invariant across environments. (O3) A binary coordinate has a nonzero internal threshold; for $K_j \geq 3$, no extra condition is needed because all internal thresholds cannot simultaneously be zero. (H1) Every environment is observed with positive probability. (H2) The normalized noise-variance profiles $r_j = (d_{1j}/d_{0j}, \ldots, d_{mj}/d_{0j})^\top$ are pairwise distinct.

Intervention-semantics layer. IIMM is used only when interpreting physical measured post-intervention quantities: a hard intervention replaces the targeted canonical latent structural equations, while untargeted structural equations and all measurement functions/thresholds remain invariant. It is not needed for observational identification.

Finite-sample layers. F1 adds fixed-$p, m, K_{\max}$ sampling and front-end regularity. F2 adds quantitative curvature, conditioning, Hessian continuity, and localization for C1 estimator. F3 adds only a graph-recovery margin when considering shrinking-edge triangular arrays; a fixed finite model already has a positive minimum nonzero coefficient if it has any edges.

B Mixed-Observation Identification Proofs

This appendix expands Proposition 1 without adding assumptions.

B.1 Continuous coordinates

Let $F_{0j}(x) = P(X_j \leq x \mid E = 0)$ and $h_j(x) = \Phi^{-1}\{F_{0j}(x)\}$. Since $X_j^{(0)} = g_j(Y_j^{(0)})$, g_j is strictly increasing, and $Y_j^{(0)} \sim N(0, s_{0j}^2)$, strict increase alone gives

$$g_j(t) \leq g_j(y) \quad \Longleftrightarrow \quad t \leq y;$$

continuity is not used. Therefore

$$F_{0j}(g_j(y)) = P(Y_j^{(0)} \leq y) = \Phi(y/s_{0j}).$$

Hence $h_j(g_j(y)) = y/s_{0j}$. Invariance of g_j across environments gives

$$h_j(X_j^{(e)}) = Y_j^{(e)}/s_{0j} \quad \text{a.s.}$$

and therefore

$$\alpha_{ej}^2 := \frac{s_{ej}^2}{s_{0j}^2} = \mathrm{Var}\{h_j(X_j^{(e)})\}.$$

For a continuous–continuous pair (j, k),

$$\Omega_{e,jk} = \mathrm{Cov}\{h_j(X_j^{(e)}), h_k(X_k^{(e)})\}.$$

B.2 Ordinal coordinates and mixed pairs

For an internal ordinal threshold k,

$$P(X_j^{(e)} \leq k \mid E = e) = P(Y_j^{(e)} \leq \gamma_{jk}) = \Phi(\gamma_{jk}/s_{ej}),$$

so

$$\delta_{ejk} := \Phi^{-1}\{P(X_j^{(e)} \leq k \mid E = e)\} = \gamma_{jk}/s_{ej}.$$

If $\gamma_{jk} \neq 0$, measurement invariance yields

$$\alpha_{ej} = \frac{s_{ej}}{s_{0j}} = \frac{\delta_{0jk}}{\delta_{ejk}}.$$

A binary coordinate with threshold zero would have $P(X_j = 1 \mid e) = 1/2$ for every latent variance and hence would reveal no marginal scale information. This motivates O3 as a sufficient visibility condition, not as a claim of necessity in all larger model classes.

For a standard bivariate Gaussian CDF $\Phi_2(a, b; \rho)$, Plackett's identity (Plackett, 1954) gives

$$\frac{\partial}{\partial \rho}\Phi_2(a, b; \rho) = \phi_2(a, b; \rho) > 0$$

for finite a, b and $|\rho| < 1$. Thus, for ordinal–ordinal pairs,

$$P(X_j^{(e)} \leq k, X_\ell^{(e)} \leq r \mid e) = \Phi_2(\delta_{ejk}, \delta_{e\ell r}; \rho_{e,j\ell})$$

uniquely determines $\rho_{e,j\ell}$. For continuous–ordinal pairs, let $U_{ej} = h_j(X_j^{(e)})/\alpha_{ej} = Y_j^{(e)}/s_{ej} \sim N(0, 1)$; then

$$P(U_{ej} \leq a, X_\ell^{(e)} \leq r \mid e) = \Phi_2(a, \delta_{e\ell r}; \rho_{e,j\ell})$$

uniquely determines the same correlation parameter. In either case,

$$\Omega_{e,j\ell} = \alpha_{ej}\alpha_{e\ell}\rho_{e,j\ell}.$$

B.3 Proof of Proposition 1

H1 makes every conditional law $P(X \mid E = e)$ observable. The preceding continuous and ordinal arguments identify α_{ej} for every coordinate and environment, so the diagonal entries $\Omega_{e,jj} = \alpha_{ej}^2$ are identified. The continuous–continuous, continuous–ordinal, and ordinal–ordinal arguments identify every off-diagonal entry. Hence each Ω_e and the full family are identified.

Absolute latent scales remain free. A positive diagonal rescaling of the latent coordinates changes Σ_e by a shared positive diagonal congruence and can be absorbed into the measurement mechanisms. Thus the original covariance family is not exactly identified beyond that shared congruence.

C Source Separation, Permutation Removal, and Sharpness

C.1 Common eigendirections

From $\Omega_e = GD_eG^\top$ define

$$H_e = \Omega_0^{-1/2}\Omega_e\Omega_0^{-1/2}, \qquad O = \Omega_0^{-1/2}GD_0^{1/2}.$$

Since $\Omega_0 = GD_0G^\top$, $OO^\top = I$, and

$$H_e = OR_eO^\top, \qquad R_e = D_0^{-1/2}D_eD_0^{-1/2} = \operatorname{diag}(d_{e1}/d_{01}, \ldots, d_{ep}/d_{0p}).$$

For any $a \in \mathbb{R}^m$,

$$H(a) := \sum_{e=1}^{m} a_eH_e = O\operatorname{diag}(a^\top r_1, \ldots, a^\top r_p)O^\top.$$

If the profiles are pairwise distinct, each equality $a^\top r_j = a^\top r_k$ defines a proper hyperplane. A finite union of such hyperplanes cannot cover $\mathbb{R}^m$, so some a gives simple spectrum. The eigenvectors of a real symmetric simple-spectrum matrix are unique up to column signs and permutation; hence O is unique up to signed permutation. Conversely, if $r_j = r_k$ for some pair, every R_e is scalar on their two-dimensional coordinate subspace, so a nontrivial orthogonal rotation preserves all H_e and uniqueness of the common eigendirections from $\{H_e\}$ fails. This converse concerns uniqueness of the common eigendirections used by our simultaneous-diagonalization route. Pairwise profile distinctness is a sufficient condition for the identification theorem above; it is not claimed to be necessary for DAG identifiability at every special parameter point.

Because $\Omega_0^{1/2}O = GD_0^{1/2}$ and D_0 is unknown positive diagonal, the covariance family identifies G exactly up to $GP\Lambda$ with a permutation P and nonsingular diagonal Λ.

C.2 Acyclicity removes the permutation

Let $W = I - B$, so $W_{ii} = 1$. Suppose a permutation matrix Q makes every diagonal entry of QW nonzero. If $Q \neq I$, its permutation contains a cycle $i_1 \mapsto i_2 \mapsto \cdots \mapsto i_q \mapsto i_1$ with $q \geq 2$. Nonzero diagonal entries of QW imply $W_{i_{\ell+1}, i_\ell} \neq 0$ around this cycle; off the diagonal this means $B_{i_{\ell+1}, i_\ell} \neq 0$, which creates the directed cycle $i_1 \to i_2 \to \cdots \to i_q \to i_1$. Contradiction. Therefore $Q = I$.

Take any representative $V = GP\Lambda$. Then

$$V^{-1} = \Lambda^{-1} P^\top (I - B) S_0.$$

The true row permutation P makes PV^{-1} have a nonzero diagonal. If another Q did so, diagonal left/right scalings show that $(QP^\top)W$ would also have a nonzero diagonal, forcing $QP^\top = I$. Thus the correct source permutation is unique. After this row permutation, row-normalizing to unit diagonal yields

$$S_0^{-1}(I - B)S_0 = I - B^\circ,$$

and hence $B^\circ = S_0^{-1} B S_0$.

C.3 Positive-rescaling sharpness

For any positive diagonal H, define $\widetilde{Y} = HY$, $\widetilde{B} = HBH^{-1}$, $\widetilde{D}_e = HD_eH$, continuous maps $\widetilde{g}_j(y) = g_j(y/h_j)$, and ordinal thresholds $\widetilde{\gamma}_{jk} = h_j \gamma_{jk}$. The transformed model satisfies the same assumptions and induces exactly the same $P(X, E)$. Moreover, the reference scale transforms as

$$\widetilde{S}_0 = HS_0,$$

so

$$\widetilde{B}^\circ = (HS_0)^{-1}(HBH^{-1})(HS_0) = B^\circ.$$

The variance profiles are unchanged because $\widetilde{d}_{ej}/\widetilde{d}_{0j} = d_{ej}/d_{0j}$. Therefore B cannot be identified more finely than positive diagonal similarity. Every matrix in that class has the same directed support and coefficient signs, and the invariant representative is B°.

D Intervention Proofs

D.1 Canonical latent hard interventions

Let $M_0 = I - B^\circ$. Since B° and Ω_e are identified,

$$D_e^\circ = M_0 \Omega_e M_0^\top = S_0^{-1} D_e S_0^{-1}$$

is identified as well. Under $do(Z_\mathcal{S} = a_\mathcal{S})$, acyclicity makes $I - B^\circ_{\mathcal{R}\mathcal{R}}$ invertible and gives Equation (6). Therefore the mean and covariance are

$$\mu_\mathcal{R}^{e,do} = (I - B^\circ_{\mathcal{R}\mathcal{R}})^{-1} B^\circ_{\mathcal{R}\mathcal{S}} a_\mathcal{S},$$

$$\Sigma_{\mathcal{R}\mathcal{R}}^{e,do} = (I - B^\circ_{\mathcal{R}\mathcal{R}})^{-1} D^\circ_{e,\mathcal{R}\mathcal{R}} (I - B^\circ_{\mathcal{R}\mathcal{R}})^{-\top},$$

which proves identification of the full canonical latent post-intervention law in any observed regime.

D.2 Sharp pointwise physical identified set

Fix one continuous coordinate and omit the subscript. Write $m(z) = g(s_0 z)$ and $F(x) = P(X \leq x)$ under the reference regime. Let $u = \Phi(z)$ and

$$L(z) = \inf\{x : F(x) \geq u\}, \qquad U(z) = \inf\{x : F(x) > u\}.$$

For a strictly increasing m, the one-sided limits

$$a = m(z-) = \sup_{t<z} m(t), \qquad b = m(z+) = \inf_{t>z} m(t)$$

exist and satisfy $a \leq m(z) \leq b$. Since the reference latent coordinate is standard Gaussian, direct monotonicity arguments give $L(z) = a$ and $U(z) = b$.

For any $c \in [a, b]$, define $m^{(c)}$ to equal m away from z and take value c at z. The modified map remains strictly increasing. Every observed Gaussian regime assigns probability zero to the single latent point z, so this point modification leaves the entire observational law $P(X, E)$ unchanged. Hence every $c \in [L(z), U(z)]$ is observationally compatible, while monotonicity rules out values outside this interval. Therefore

$$\mathrm{IdSet}\{m(z) \mid P(X, E)\} = [L(z), U(z)].$$

For the full-law statement we first prove that untargeted continuous coordinates introduce no additional ambiguity. For any continuous coordinate j, define the observational ambiguity set

$$J_j := \{z : L_j(z) < U_j(z)\}.$$

The pointwise result above applies to every observationally compatible strictly increasing canonical measurement map $\widetilde{m}_j$: for every z,

$$\widetilde{m}_j(z) \in [L_j(z), U_j(z)].$$

Hence all compatible maps agree at every $z \notin J_j$. For any one compatible representative m_j, the one-sided-limit identities give

$$J_j = \{z : m_j(z-) < m_j(z+)\},$$

the jump-discontinuity set of a monotone function. This set is at most countable. Indeed, for each $z \in J_j$ choose a rational number in the nonempty interval $(m_j(z-), m_j(z+))$; these open jump intervals are pairwise disjoint for distinct z, so this choice injects J_j into $\mathbb{Q}$.

Now let $\mathcal{R} = [p] \setminus \mathcal{S}$ and

$$K_{\mathcal{R}} := (I - B^{\circ}_{\mathcal{R}\mathcal{R}})^{-1}.$$

Acyclicity makes $K_{\mathcal{R}}$ invertible, and $D^{\circ}_{e,\mathcal{R}\mathcal{R}} \succ 0$ is diagonal. Therefore

$$\Sigma^{e,do}_{\mathcal{R}\mathcal{R}} = K_{\mathcal{R}} D^{\circ}_{e,\mathcal{R}\mathcal{R}} K_{\mathcal{R}}^{\top} \succ 0.$$

Thus every untargeted coordinate Z_j, $j \in \mathcal{R}$, has a nondegenerate Gaussian post-intervention marginal. In particular, for each $j \in \mathcal{R} \cap \mathcal{C}$,

$$P_{e,do}\{Z_j \in J_j\} = 0.$$

Because there are only finitely many coordinates,

$$P_{e,do}\{\exists j \in \mathcal{R} \cap \mathcal{C} : \ Z_j \in J_j\} = 0.$$

No independence among the untargeted coordinates is required. Consequently, any two observationally compatible collections of continuous measurement maps agree simultaneously almost surely on all untargeted continuous coordinates under the identified post-intervention latent law. The canonical ordinal thresholds together with the fixed endpoint convention likewise determine every untargeted ordinal response, so the joint law of all untargeted measured coordinates is unique; denote it by $\nu_{e,\mathcal{S},a}$.

It remains to identify the directly targeted measured values. Let

$$T = \mathcal{S} \cap \mathcal{C}, \qquad Q = \mathcal{S} \cap \mathcal{O}.$$

For $j \in Q$, let $\kappa_j(a_j)$ be the deterministic category defined in the main text, and write $\kappa_Q(a_Q)$ for the resulting vector. For every

$$c_T \in \prod_{j \in T} [L_j(a_j), U_j(a_j)],$$

modify each targeted continuous map only at a_j so that its value there is c_j, leaving all other latent points and all untargeted measurement maps unchanged. The pointwise argument above shows that each modified map remains strictly increasing. Since the finite union of these targeted latent points has probability zero in every observational Gaussian regime, the modified model induces the same $P(X, E)$; under IIMM and $do(Z_{\mathcal{S}} = a_{\mathcal{S}})$, its targeted continuous responses are exactly c_T, its targeted ordinal responses are $\kappa_Q(a_Q)$, and its untargeted measured law is $\nu_{e,\mathcal{S},a}$. Hence every law in

$$\left\{ \delta_{c_T} \otimes \delta_{\kappa_Q(a_Q)} \otimes \nu_{e,\mathcal{S},a} : c_T \in \prod_{j \in T} [L_j(a_j), U_j(a_j)] \right\}$$

is attainable, up to returning coordinates to their original order.

Conversely, every observationally compatible model has each directly targeted continuous response in $[L_j(a_j), U_j(a_j)]$ by the pointwise theorem, has the same deterministic targeted ordinal vector $\kappa_Q(a_Q)$ by the fixed threshold convention, and has the same untargeted measured joint law $\nu_{e,\mathcal{S},a}$ by the countable-ambiguity argument above. Therefore the displayed family is the exact physical post-intervention-law identified set. It is a singleton if and only if every directly targeted continuous interval collapses. This proves the full mixed-X uniqueness criterion stated in Theorem 3.

E FINITE-SAMPLE THEORY

E.1 F1: FRONT-END COVARIANCE FAMILY

For each environment observe n_e i.i.d. samples and let $n_{\min} = \min_e n_e$. The theory fixes $p, m, K_{\max}$. Scale/correlation cutpoints and continuous anchor levels are prespecified. The cumulative probabilities used by the ordinal front end stay in $[\eta_p, 1-\eta_p]$; scale-anchor probits stay at least $d_{\mathrm{sc}} > 0$ away from zero; continuous anchor levels stay away from $0, 1, 1/2$; relative scales satisfy $0 < \alpha_{\min} \le \alpha_{ej} \le \alpha_{\max} < \infty$; pairwise correlations satisfy $|\rho_{e,jk}| \le \rho_* < 1$; and correlation cutoffs lie in a fixed compact region. There is no balanced-sample, sparsity, beta-min, or ridge requirement in F1.

Let $p_C = |\mathcal{C}|$, $p_O = |\mathcal{O}|$ and

$$N_F = (m+1)(p_C + 2p_O), \qquad N_\tau = (m+1)\binom{p}{2}.$$

For $\delta \in (0,1)$ define

$$t_F(\delta) = \sqrt{\frac{1}{2n_{\min}} \log \frac{4N_F}{\delta}}, \qquad t_\tau(\delta) = \sqrt{\frac{8}{n_{\min}} \log \frac{4N_\tau}{\delta}},$$

$$r_{\Omega,n}(\delta) = t_F(\delta) + t_\tau(\delta) + \frac{1}{n_0}.$$

Lemma E.1 (F1: mixed front-end covariance-family error)**.** *Under the population model and the preceding regularity class, there exists a finite class constant C_Ω such that, for every $\delta \in (0,1)$, whenever $n_{\min} \ge n_\star(\delta)$ for a finite threshold determined by the fixed regularity margins and δ,*

$$\mathbb{P}\Big\{ \max_e \|\widehat{\Omega}_e - \Omega_e\|_{\max} \le C_\Omega r_{\Omega,n}(\delta) \Big\} \ge 1 - \delta.$$

Consequently,

$$\mathbb{P}\Big\{ \max_e \|\widehat{\Omega}_e - \Omega_e\|_{\mathrm{op}} \le \Delta_{\Omega,n}(\delta) \Big\} \ge 1 - \delta,$$

where

$$\Delta_{\Omega,n}(\delta) = pC_\Omega \left[\sqrt{\frac{1}{2n_{\min}} \log \frac{4N_F}{\delta}} + \sqrt{\frac{8}{n_{\min}} \log \frac{4N_\tau}{\delta}} + \frac{1}{n_0} \right].$$

For fixed δ, p, m, $\max_e \|\widehat{\Omega}_e - \Omega_e\|_{\mathrm{op}} = O_p(n_{\min}^{-1/2} + n_0^{-1})$.

Proof ingredients. DKW control with Massart's sharp constant (Massart, 1990) and Hoeffding bounds (Hoeffding, 1963) control the empirical CDF/cumulative-probability terms at scale $t_F(\delta)$, while the bounded-kernel U-statistic inequality of Hoeffding (1963) gives $t_\tau(\delta)$ for Kendall statistics. The empirical reference quantile contributes the $1/n_0$ discretization term. On the fixed interior margins, the probit map, relative-scale ratios, and binary–continuous/binary–binary bridge inverses have finite Lipschitz constants. Large-sample admissibility makes clipping, denominator flooring, correlation capping, and bridge attainable-range projections inactive. If $\widetilde{R}_e$ is the pairwise correlation matrix and $\widehat{R}_e$ is a max-norm nearest correlation matrix, population feasibility gives

$$\|\widehat{R}_e - R_e\|_{\max} \le 2\|\widetilde{R}_e - R_e\|_{\max}.$$

Expanding $\widehat{Q}_e \widehat{R}_e \widehat{Q}_e - Q_e R_e Q_e$ then yields the stated bound. Finally $\|A\|_{\mathrm{op}} \le p\|A\|_{\max}$ gives the operator-norm radius.

E.2 F2: C1 ESTIMATOR COEFFICIENT ERROR

Throughout F2, the criterion weights satisfy $w_e > 0$ and $\sum_{e=0}^{m} w_e = 1$. Let

$$\mathcal{M} = \{M \in \mathbb{R}^{p\times p} : \operatorname{diag}(M) = \mathbf{1},\ \det M \neq 0\}, \qquad M_0 = I - B^\circ,$$

and let $\theta = \operatorname{off}(M) \in \mathbb{R}^d$, $\theta_0 = \operatorname{off}(M_0)$, where $d = p(p-1)$. Define

$$D_e^\circ = M_0\Omega_e M_0^\top = S_0^{-1}D_e S_0^{-1} = \operatorname{diag}(d_{e1}^\circ, \dots, d_{ep}^\circ), \qquad d_{ei}^\circ = d_{ei}/s_{0i}^2,$$

$$d_*^\circ = \min_{e,i} d_{ei}^\circ, \qquad R_0 = \max_i \|m_{0i}\|_2, \qquad V_0 = \max_i \|(M_0^{-T})_{i:}\|_\infty,$$

$$C_{\text{score}} = \frac{4}{d_*^\circ}(R_0 + R_0^2 V_0), \qquad \mu_\Omega = \min_e \lambda_{\min}(\Omega_e) > 0.$$

The population smooth profiled criterion is

$$q(M) = \sum_e w_e \left[\sum_i \log(m_i^\top \Omega_e m_i) - 2\log|\det M|\right]$$

with gradient

$$\nabla q(M) = 2\sum_e w_e \left[\mathcal{D}_e(M)^{-1}M\Omega_e - M^{-T}\right], \quad \mathcal{D}_e(M) = \operatorname{Diag}\{\operatorname{diag}(M\Omega_e M^\top)\}.$$

At M_0, $\nabla q(M_0) = 0$. If $E_e = \widehat{\Omega}_e - \Omega_e$, $\delta_{ei} = m_{0i}^\top E_e m_{0i}$, and $v_i^\top = (M_0^{-T})_{i:}$, the exact rowwise score perturbation identity is

$$\frac{m_{0i}^\top(\Omega_e + E_e)}{d_{ei}^\circ + \delta_{ei}} - v_i^\top = \frac{m_{0i}^\top E_e - \delta_{ei}v_i^\top}{d_{ei}^\circ + \delta_{ei}}.$$

Thus, if $R_0^2\Delta_\Omega \le d_*^\circ/2$,

$$\|\nabla_{\text{off}}\widehat{q}(M_0)\|_\infty \le C_{\text{score}}\Delta_\Omega.$$

Population Hadamard gap and global uniqueness. Remove the M-independent covariance determinants from q by defining

$$f_\Omega(M) := q(M) - \sum_e w_e \log\det(\Omega_e).$$

Using $\det(M\Omega_e M^\top) = \det(M)^2\det(\Omega_e)$ gives the exact representation

$$f_\Omega(M) = \sum_e w_e \log\left\{\frac{\prod_{i=1}^p (M\Omega_e M^\top)_{ii}}{\det(M\Omega_e M^\top)}\right\}. \tag{11}$$

For every $M \in \mathcal{M}$, $M\Omega_e M^\top$ is positive definite. Hadamard's determinant inequality therefore makes every summand nonnegative. Equality in Hadamard's inequality for a positive-definite matrix holds exactly when its coordinate vectors are mutually orthogonal, equivalently when the matrix is diagonal. Because every w_e is positive,

$$f_\Omega(M) = 0 \iff M\Omega_e M^\top \text{ is diagonal for every } e. \tag{12}$$

Under H2, the frozen simultaneous-diagonalization result in Appendix C implies that any such joint diagonalizer has rows equal to those of G^{-1} up to a common row permutation and nonsingular row scaling. The unit-diagonal constraint removes the row scaling once a row permutation is fixed. The acyclicity permutation-removal argument in Appendix C then leaves only the true row assignment, whose unit-diagonal normalization is $M_0 = I - B^\circ$. Hence

$$M \in \mathcal{M}, \qquad f_\Omega(M) = 0 \iff M = M_0. \tag{13}$$

This uses H2 only for uniqueness of the simultaneous-diagonalization directions and does not introduce faithfulness.

Sample Hadamard gap and the global ℓ_1 radius. On a covariance-good event write $\max_e \|E_e\|_{\mathrm{op}} \le \Delta_\Omega$ and assume $\Delta_\Omega \le \mu_\Omega/2$, so every $\widehat{\Omega}_e$ is positive definite. Define

$$\widehat{f}(M) := \widehat{q}(M) - \sum_e w_e \log\det(\widehat{\Omega}_e).$$

The same determinant identity and Hadamard equality condition give $\widehat{f}(M) \ge 0$ on $\mathcal{M}$. Since the C1 estimator penalty is $\lambda_n\|\theta\|_1$, let $\lambda_n > 0$ and let $\widehat{M}$ be any global minimizer. Comparing it with M_0 yields

$$\widehat{f}(\widehat{M}) + \lambda_n\|\widehat{\theta}\|_1 \le \widehat{f}(M_0) + \lambda_n\|\theta_0\|_1, \tag{14}$$

and therefore

$$\|\widehat{\theta}\|_1 \le \|\theta_0\|_1 + \frac{\widehat{f}(M_0)}{\lambda_n}. \tag{15}$$

The requirement $\lambda_n > 0$ is essential for this global radius argument; the finite-sample global guarantee here does not cover $\lambda_n = 0$. Moreover, $\mathrm{diag}(M) = \mathbf{1}$ keeps every row nonzero, so $\widehat{f}(M) \to +\infty$ along bounded sequences approaching the singular boundary, while $\lambda_n\|\theta\|_1 \to +\infty$ as $\|\theta\|_1 \to \infty$. Hence the extended penalized criterion attains a global minimum on the unit-diagonal parameter space.

The sample Hadamard gap at M_0 is second order in the covariance perturbation. Indeed, put $R_M = \|M_0\|_{\mathrm{op}}$ and

$$F_e = (D_e^\circ)^{-1/2} M_0 E_e M_0^\top (D_e^\circ)^{-1/2}.$$

Then $\|F_e\|_{\mathrm{op}} \le R_M^2 \Delta_\Omega / d_*^\circ$. If this is at most $1/2$, the Hadamard representation (11) at M_0 becomes

$$h_e := \sum_i \log(1 + (F_e)_{ii}) - \log\det(I + F_e).$$

The value at $F_e = 0$ is zero and the first variation cancels because $\sum_i (F_e)_{ii} = \mathrm{tr}(F_e)$. More explicitly, for $|x| \le 1/2$, $|\log(1+x) - x| \le x^2$, and applying the same scalar bound to the eigenvalues of the symmetric matrix F_e gives

$$0 \le h_e \le \sum_i (F_e)_{ii}^2 + \sum_i \lambda_i(F_e)^2 \le 2\|F_e\|_F^2 \le 2p\|F_e\|_{\mathrm{op}}^2.$$

Consequently,

$$0 \le \widehat{f}(M_0) \le C_H \Delta_\Omega^2, \qquad C_H := \frac{2pR_M^4}{(d_*^\circ)^2}, \tag{16}$$

where C_H depends only on fixed p, the population noise-variance lower bound, and the conditioning of M_0. Combining displays (15)–(16),

$$\|\widehat{\theta}\|_1 \le \|\theta_0\|_1 + C_H \frac{\Delta_\Omega^2}{\lambda_n}. \tag{17}$$

Thus the existing condition $\Delta_{\Omega,n}(\delta)^2/\lambda_n = O(1)$ is used first to confine *all* global penalized minimizers to one fixed bounded ℓ_1 region. Concretely, if $\Delta_\Omega^2/\lambda_n \le L$ for a fixed finite L, we may take

$$K := \|\theta_0\|_1 + C_H L + 1. \tag{18}$$

Bounded-region separation and uniform criterion perturbation. Fix a local radius $\rho > 0$ small enough that the unit-diagonal ρ-ball around M_0 remains in $\mathcal{M}$ and lies inside the curvature neighborhood below, and define

$$g_{\rho,K} := \inf\left\{ f_\Omega(M) : \mathrm{diag}(M) = \mathbf{1},\ \|\theta\|_1 \le K,\ \|\theta - \theta_0\|_2 \ge \rho \right\}, \tag{19}$$

where $f_\Omega(M)$ is extended to $+\infty$ at singular M. The closed set in (19) is compact because the off-diagonal entries are bounded. The extension is lower semicontinuous: as a unit-diagonal matrix approaches singularity, every diagonal quadratic form $m_i^\top \Omega_e m_i$ stays at least μ_Ω, whereas the determinant in (11) tends to zero. By the unique-zero property in (13), the compact set excludes the only zero M_0; hence

$$g_{\rho,K} > 0. \tag{20}$$

The sample and population gaps are uniformly close on the invertible part of this bounded region. If $\mathrm{diag}(M) = \mathbf{1}$ and $\|\theta\|_1 \le K$, each row satisfies $\|m_i\|_2 \le 1 + K$ and $m_i^\top \Omega_e m_i \ge \mu_\Omega$. Whenever

$$\Delta_\Omega \le \frac{\mu_\Omega}{2(1+K)^2},$$

we have

$$\left|\log(m_i^\top \widehat{\Omega}_e m_i) - \log(m_i^\top \Omega_e m_i)\right| \le \frac{2(1+K)^2}{\mu_\Omega}\Delta_\Omega.$$

Also $\|\Omega_e^{-1/2} E_e \Omega_e^{-1/2}\|_{\mathrm{op}} \le 1/2$, so eigenvalue-wise expansion of $\log\det$ yields

$$|\log\det\widehat{\Omega}_e - \log\det\Omega_e| \le \frac{2p}{\mu_\Omega}\Delta_\Omega.$$

The $-2\log|\det M|$ terms cancel between sample and population gaps, and therefore

$$\sup_{M\in\mathcal{M}:\ \|\theta\|_1\le K} |\widehat{f}(M) - f_\Omega(M)| \le C_K\Delta_\Omega, \qquad C_K := \frac{2p\{(1+K)^2+1\}}{\mu_\Omega}. \tag{21}$$

Thus C_K is a fixed population constant once K is fixed; it depends only on fixed p, K, and the smallest population covariance eigenvalue.

Global localization lemma. Suppose the preceding small-perturbation bounds hold, $\Delta_\Omega^2/\lambda_n \le L$, and

$$C_K\Delta_\Omega + C_H\Delta_\Omega^2 + \lambda_n\|\theta_0\|_1 < g_{\rho,K}. \tag{22}$$

Then every global penalized C1 estimator minimizer lies in $\|\theta - \theta_0\|_2 < \rho$. To see this, Display (17) puts every global minimizer inside $\|\theta\|_1 \le K$. Any point in that region but outside the local ball satisfies, by (20) and (21),

$$\widehat{f}(M) + \lambda_n\|\theta\|_1 \ge g_{\rho,K} - C_K\Delta_\Omega,$$

whereas at the truth (16) gives

$$\widehat{f}(M_0) + \lambda_n\|\theta_0\|_1 \le C_H\Delta_\Omega^2 + \lambda_n\|\theta_0\|_1.$$

The strict bound (22) makes the first display strictly larger than the second, contradicting global optimality. This comparison is deterministic on the covariance-good event and applies to every member of the global argmin set. In particular, if $\Delta_{\Omega,n}(\delta) \to 0$, $\lambda_n \to 0$, and $\Delta_{\Omega,n}(\delta)^2/\lambda_n = O(1)$, the fixed K above and any fixed admissible local ρ satisfy the localization inequality for all sufficiently large samples.

Theta-coordinate curvature and local error transfer. For the curvature calculation write $M = (I + K')M_0$. The second-order off-diagonal contribution for each pair $i < j$ has block

$$C_{ij} = \begin{pmatrix} \sum_e w_e d_{ej}^\circ/d_{ei}^\circ & 1 \\ 1 & \sum_e w_e d_{ei}^\circ/d_{ej}^\circ \end{pmatrix}.$$

H2 makes the ratio $d_{ej}^\circ/d_{ei}^\circ$ nonconstant across environments for every $i \ne j$. Strict Cauchy–Schwarz therefore gives $(\sum_e w_e t_e)(\sum_e w_e/t_e) > 1$ for $t_e = d_{ej}^\circ/d_{ei}^\circ$, so every C_{ij} is positive definite. Diagonal K' directions are exactly the row-scaling null directions of q. On the unit-diagonal tangent space, however, the linear map from a zero-diagonal perturbation $U = M - M_0$ to $\mathrm{off}(K') = \mathrm{off}(UM_0^{-1})$ is injective: if $\mathrm{off}(K') = 0$, then K' is diagonal, while $\mathrm{diag}(U) = \mathrm{diag}(K'M_0) = \mathrm{diag}(K') = 0$, forcing $K' = 0$. Since both spaces have dimension $d = p(p-1)$, this map is invertible. Hence the theta-coordinate Hessian is positive definite. Define, unambiguously in the coordinates used by the error bound,

$$\kappa_0 := \frac{1}{2}\lambda_{\min}\{\nabla^2_{\theta\theta} q(\theta_0)\} > 0. \tag{23}$$

Fixed conditioning of M_0 is absorbed into this constant through the preceding linear coordinate map.

By Hessian continuity choose the same ρ small enough that $\lambda_{\min}\{\nabla^2_{\theta\theta} q(\theta)\} \ge 3\kappa_0/2$ throughout $\|\theta - \theta_0\|_2 \le \rho$. The quantitative F2 backend condition requires the sample-Hessian perturbation on this ball to be at most $\kappa_0/2$ in operator norm; consequently

$$\nabla^2_{\theta\theta}\widehat{q}(\theta) \succeq \kappa_0 I \quad \text{throughout the local ball.}$$

After the global localization lemma, let $u = \widehat{\theta} - \theta_0$ for any global minimizer. Taylor's theorem along the segment from θ_0 to $\widehat{\theta}$ gives

$$\widehat{q}(\theta_0 + u) \geq \widehat{q}(\theta_0) + \nabla_\theta \widehat{q}(\theta_0)^\top u + \frac{\kappa_0}{2}\|u\|_2^2.$$

Combining this with penalized optimality, the score bound, and $\|u\|_1 \leq \sqrt{d}\|u\|_2$ gives

$$\frac{\kappa_0}{2}\|u\|_2^2 \leq \sqrt{d}\big(C_{\mathrm{score}}\Delta_\Omega + \lambda_n\big)\|u\|_2.$$

After canceling $\|u\|_2$ when nonzero,

$$\|u\|_2 \leq \frac{2\sqrt{d}}{\kappa_0}\big(C_{\mathrm{score}}\Delta_\Omega + \lambda_n\big). \tag{24}$$

Because $\|\widehat{B}^\circ - B^\circ\|_{\max} \leq \|u\|_2$, the factor 2 in the theorem below is the exact consequence of the definition (23) and the Taylor factor $1/2$; no coordinate-dependent curvature constant is being reused.

Lemma E.2 (F2: C1 estimator coefficient error)**.** *Under F1, normalized criterion weights $w_e > 0$ with $\sum_{e=0}^m w_e = 1$, H2, and the local Hessian-continuity/sample-Hessian conditions above, fix an admissible local radius ρ. Let $\lambda_n > 0$. On the F1 covariance-good event, if $\Delta_{\Omega,n}(\delta)$ is small enough for the displayed score, Hadamard-gap, positive-definiteness, and uniform-perturbation bounds, if $\Delta_{\Omega,n}(\delta)^2/\lambda_n \leq L$ for a fixed finite L, and if (22) holds with K from (18), then every global C1 estimator minimizer lies in the local basin and*

$$\|\widehat{B}^\circ - B^\circ\|_{\max} \leq \frac{2\sqrt{p(p-1)}}{\kappa_0}\left[C_{\mathrm{score}}\Delta_{\Omega,n}(\delta) + \lambda_n\right].$$

Therefore the same inequality holds with probability at least $1-\delta$ for any fixed measurable global-minimizer selection. For fixed p, m and any tuning sequence satisfying the stated localization conditions,

$$\|\widehat{B}^\circ - B^\circ\|_{\max} = O_p(n_{\min}^{-1/2} + n_0^{-1} + \lambda_n).$$

The radius constants depend on unknown population curvature/conditioning/localization quantities and are not data-computable confidence radii. If $\delta_n = n_{\min}^{-a}$, the convenient deterministic envelope contains $\sqrt{\log n_{\min}/n_{\min}}$; this is a vanishing-failure envelope, not the intrinsic pointwise rate.

E.3 F3: SEPARATE THRESHOLD GRAPH RECOVERY

Let $\|\widehat{B}^\circ - B^\circ\|_{\max} = O_p(a_n)$ with $a_n \to 0$ and choose $\tau_n = c_n a_n$ where $c_n \to \infty$ and $c_n a_n \to 0$. Then $\|\widehat{B}^\circ - B^\circ\|_{\max} = o_p(\tau_n)$.

Lemma E.3 (F3: thresholded support/DAG consistency)**.** *For a fixed finite-dimensional model, if the true edge set is nonempty let $\beta^\circ_{\min} = \min_{B^\circ_{ij}\neq 0}|B^\circ_{ij}| > 0$. Then*

$$\mathbb{P}(\widehat{G}_{\tau_n} = G_0) \to 1.$$

For a triangular array B°_n, the same conclusion holds if

$$\frac{\beta^\circ_{\min,n} - \tau_n}{a_n} \to \infty, \qquad \beta^\circ_{\min,n} > \tau_n \text{ eventually}.$$

For a true zero, a false positive requires coefficient error exceeding τ_n; for a fixed true edge, a false negative requires error exceeding $\beta^\circ_{\min} - \tau_n$. A union bound over finitely many entries proves the fixed-model result. The triangular-array statement uses the displayed margin. This is asymptotic exact support/DAG consistency, not a finite-sample exact-DAG theorem, and beta-min is not an estimator input.

F Estimator and Algorithm Details

F.1 Totalized relative-scale estimation

For each environment define

$$\varepsilon_e = \frac{1}{2(n_e+1)}, \qquad C_e(u) = \min\{1-\varepsilon_e, \max(\varepsilon_e, u)\},$$

$$b_e = \frac{1}{n_e+1}, \qquad s_e(x) = \max\{|x|, b_e\}.$$

All empirical probabilities passed to Φ^{-1} are clipped by C_e and all nonreference scale-ratio denominators use the magnitude floor. Set $\widehat{\alpha}_{0j} = 1$.

For $j \in \mathcal{O}$, with a prespecified scale cutpoint c_j^{sc},

$$\widehat{p}_{ej}^{\mathrm{sc}} = n_e^{-1} \sum_i \mathbf{1}\{X_{ij}^{(e)} \le c_j^{\mathrm{sc}}\}, \quad \widehat{\delta}_{ej}^{\mathrm{sc}} = \Phi^{-1}(C_e(\widehat{p}_{ej}^{\mathrm{sc}})),$$

$$\widehat{\alpha}_{ej} = \frac{s_0(\widehat{\delta}_{0j}^{\mathrm{sc}})}{s_e(\widehat{\delta}_{ej}^{\mathrm{sc}})}, \quad e \ge 1.$$

For $j \in \mathcal{C}$, fix an anchor level $u_j \in (0,1)$, let $\widehat{x}_j$ be the reference empirical u_j-quantile, and set

$$\widehat{z}_{ej} = \Phi^{-1}\Big(C_e\{\widehat{F}_{ej}(\widehat{x}_j)\}\Big), \qquad \widehat{\alpha}_{ej} = \frac{s_0(\widehat{z}_{0j})}{s_e(\widehat{z}_{ej})}.$$

DKW uniform control (Massart, 1990) allows evaluation at the random reference anchor without sample splitting.

F.2 Rank bridges and covariance assembly

For ordinal coordinates fix correlation cutpoints c_j^{cor} and binarize as $Z_{ij}^{(e)} = \mathbf{1}\{X_{ij}^{(e)} > c_j^{\mathrm{cor}}\}$. We use the latent-Gaussian Kendall bridge functions of Fan et al. (2017); ordinal coordinates are reduced to binary indicators at the prespecified correlation cutpoints. Pairwise Kendall statistics use the continuous value for $j \in \mathcal{C}$ and this binary reduction for $j \in \mathcal{O}$, with ties contributing zero. Continuous–continuous correlations use

$$\widetilde{\rho} = \sin\Big(\frac{\pi}{2}\widehat{\tau}\Big).$$

For binary–continuous and binary–binary pairs, respectively,

$$F_{BC}(\rho;\Delta) = 4\Phi_2\left(\Delta, 0; \frac{\rho}{\sqrt{2}}\right) - 2\Phi(\Delta),$$

$$F_{BB}(\rho;\Delta_1,\Delta_2) = 2\{\Phi_2(\Delta_1,\Delta_2;\rho) - \Phi(\Delta_1)\Phi(\Delta_2)\}.$$

Their derivatives with respect to ρ are strictly positive on $(-1,1)$. The empirical Kendall value is first projected to the attainable bridge range over $\rho \in [-\rho_{\mathrm{cap,e}}, \rho_{\mathrm{cap,e}}]$ with $\rho_{\mathrm{cap,e}} = 1 - 1/(n_e+1)$ and then inverted. This defines the pairwise estimator on the whole sample space.

Assemble the pairwise estimates into a symmetric unit-diagonal matrix $\widetilde{R}_e$ and define

$$\widehat{R}_e \in \arg\min_{R \succeq 0,\, \mathrm{diag}(R)=\mathbf{1}} \|R - \widetilde{R}_e\|_{\max}, \qquad \widehat{\Omega}_e = \widehat{Q}_e \widehat{R}_e \widehat{Q}_e.$$

There is no environment-specific restandardization and no ridge in the theoretical definition.

F.3 C1 ESTIMATOR OPTIMIZATION AND GRAPH EXTRACTION

The mathematical estimator is exactly Equations (8)–(9); graph extraction is Equation (10). For whole-sample-space completeness, fixed measurable selections are used for set-valued PSD projections and global argmins. If any $\widehat{\Omega}_e$ is non-positive-definite, the theorem-level definition returns $\widehat{B}^\circ = 0$ and an empty graph; this technical fallback is inactive on the F1 good event once the covariance error is smaller than half the minimum population eigenvalue and is not practical advice.

The numerical implementation optimizes the unchanged C1 estimator objective in Equations (8)–(9). After the maximum scheduled iterations, it applies a deterministic polishing step and evaluates the same final stationarity criterion; the polish changes neither the statistical objective nor the graph-extraction rule. The finite-sample theorem concerns a global minimizer of the mathematical criterion; the numerical stationarity check is not a certificate of global optimality. Experimental and cross-validation choices of λ are practical tuning rules; the theorem does not assert that a CV-selected λ automatically satisfies its asymptotic localization conditions.

F.4 AJD COMPARATOR RETAINED FOR CONTROLLED ABLATION

The population simultaneous-diagonalization argument remains part of identification. As an internal finite-sample comparator, the same estimated covariance front end may instead whiten $\widehat{H}_e = \widehat{\Omega}_0^{-1/2}\widehat{\Omega}_e\widehat{\Omega}_0^{-1/2}$ and minimize an approximate joint-diagonalization off-diagonal loss. The retained comparator perturbation analysis yields, under its quantitative separation conditions,

$$\min_{P,S} \|\widehat{O} - OPS\|_F \le \frac{4\sqrt{p}}{\gamma_{\mathrm{JD}}} \max_e \|\widehat{H}_e - H_e\|_{\mathrm{op}},$$

followed by its own permutation/row-normalization backend bound. This estimator is reported only as a same-front-end comparator; it is not the primary statistical chain.

G SIMULATION DESIGN DETAILS

G.1 GRAPH AND COEFFICIENT GENERATION

The primary synthetic design uses $p = 10$, five environments $e = 0, 1, 2, 3, 4$, expected degree 2, and eight independently generated graphs satisfying fixed generation constraints. Candidate acyclic graphs receive coefficient magnitudes sampled from $U[0.35, 0.80]$ with independent random signs. Graphs are accepted before any recovery run only if all of the following hold:

- edge count is between 7 and 13;
- canonical beta-min is at least 0.25;
- $\max_{i,j} |B^\circ_{ij}| \le 1.5$;
- $\kappa_2(I - B^\circ) \le 8$;
- all five population Ω_e are positive definite and $\max_e \kappa_2(\Omega_e) \le 30$;
- all ten variance profiles are distinct and meet the minimum profile-separation requirement.

Graphs are redrawn deterministically until the pre-data constraints pass; recovery performance never enters graph selection. The released V2 records preserve realized seed provenance for the simulation runs, including immutable dataset/run identifiers and recorded data- or method-seed fields where available. Because the paper-facing V2 package does not contain the complete project-level `seed_for` implementation needed to derive every graph, data, and method seed from first principles, we do not claim exact reconstruction of all randomness from a disclosed derivation rule.

G.2 ENVIRONMENT HETEROGENEITY

The five-environment profile family uses a fixed 10-code, four-bit codebook and

$$r_{ej} = \exp(\delta c_{ej}), \qquad e = 1, \ldots, 4,$$

with strong $\delta = 0.50$ and weak $\delta = 0.12$. The strong and weak cells are paired on the same underlying graph and reference-environment sample. The corresponding minimum profile-distance scales are $2\sinh(0.50) = 1.0422$ and $2\sinh(0.12) = 0.2406$.

Codebook table. For $e = 1, \ldots, 4$, bit 1 means $c_{ej} = +1$ and bit 0 means $c_{ej} = -1$. The node order is:

node	code	c_{1j}	c_{2j}	c_{3j}	c_{4j}
1	`1000`	$+1$	-1	-1	-1
2	`0100`	-1	$+1$	-1	-1
3	`0010`	-1	-1	$+1$	-1
4	`0001`	-1	-1	-1	$+1$
5	`1100`	$+1$	$+1$	-1	-1
6	`1010`	$+1$	-1	$+1$	-1
7	`1001`	$+1$	-1	-1	$+1$
8	`0110`	-1	$+1$	$+1$	-1
9	`0101`	-1	$+1$	-1	$+1$
10	`0011`	-1	-1	$+1$	$+1$

Thus $r_{ej} = \exp(\delta c_{ej})$, with $\delta_{\text{strong}} = 0.50$ and $\delta_{\text{weak}} = 0.12$.

G.3 Measurements and sample sizes

Nodes 1–5 are continuous with maps

$$z, \qquad z + 0.15z^3, \qquad \sinh(0.5z), \qquad \operatorname{asinh}(z), \qquad z + 0.1\tanh(z).$$

Nodes 6–7 are binary with single thresholds $[-0.50]$ and $[0.60]$. Nodes 8–10 are ordinal with thresholds

$$[-0.80, 0.40],\ [-0.40, 0.90],\ [-1.00, -0.20, 0.80].$$

For the thresholded coordinates, the frozen primary simulation uses $c_j^{\text{sc}} = 1$ and $c_j^{\text{cor}} = 1$. Per-environment sample sizes are $n_e \in \{500, 1500, 5000\}$. Strong and weak separation cells use the same structural graph and reference sample under the paired design, so changes can be attributed to profile separation rather than graph selection. For Copula PC specifically, the paired cells use profile-specific deterministic Gibbs-sampler seeds, so its stochastic fit can differ slightly even when the shared $X^{(0)}$ input is identical.

G.4 Baselines, oracles, and metrics

latent-PC and Copula PC are native mixed-data baselines on raw mixed $X^{(0)}$ in the reference environment, returning PC/CPDAG-style graph information. AJD is the internal same-front backend ablation: it receives the identical estimated covariance family $\{\widehat{\Omega}_e\}$ as C1 estimator. Corrected BACKSHIFT is an external multi-environment same-front covariance backend receiving that same family; it is not a raw-mixed-X implementation. Oracle diagnostics include methods given the true canonical Z and methods given the true covariance family $\{\Omega_e\}$. They are information-advantaged diagnostics rather than equal-information mixed-X competitors. Oracle-Z UT-IGSP and CD-NOD are retained only in the dedicated structural-diagnostic table, and MSS is retained only in its official Oracle-Z + true-MEC role.

Implementation provenance. The reported latent-PC run uses the authors' `xidongdxi/latentPC` implementation pinned at commit `34e8036e`, with the typed mixed reference sample, its latent-Gaussian correlation recovery, and `pcalg::gaussCItest` at $\alpha = 0.05$ using the `stable.fast` skeleton option. Copula PC uses the authors' `cuiruifei/CausalMissingValues` implementation pinned at `251cba17`, GESS variant, with 1000 Gibbs samples, burn-in 500, $S_0 = I_p/n$, conservative PC, and $\alpha = 0.05$. BACKSHIFT uses `christinaheinze/backShift` pinned at `8cc72909850ac8844f58ff07d97c79a853faa8a9`; its `ffdiag` and `permuteAndScale` routines are unchanged, while the covariance-level wrapper replaces only the

Table 1: Primary Panel-A coefficient recovery. Means are across fixed graph topologies, not repeated-sampling replicates. The denominator column gives successful graphs/attempted graphs; the two $n = 500$ C1 cells with failures therefore use 6/8 (strong) and 7/8 (weak).

Profile	n	Method	AUPRC	AUROC	$\|\widehat{B}^\circ - B^\circ\|_{\max}$	$\|\widehat{B}^\circ - B^\circ\|_F$	Denom.
Strong	500	C1	0.915	0.968	0.227	0.497	6/8
		AJD (same front)	0.642	0.911	0.668	1.654	8/8
		BACKSHIFT (same front)	0.843	0.973	0.481	1.271	8/8
Strong	1500	C1	0.982	0.992	0.184	0.281	8/8
		AJD (same front)	0.978	0.997	0.253	0.646	8/8
		BACKSHIFT (same front)	0.949	0.995	0.313	0.761	8/8
Strong	5000	C1	1.000	1.000	0.051	0.094	8/8
		AJD (same front)	0.996	1.000	0.151	0.346	8/8
		BACKSHIFT (same front)	0.971	0.997	0.273	0.545	8/8
Weak	500	C1	0.644	0.869	0.537	1.031	7/8
		AJD (same front)	0.209	0.674	1.358	3.469	8/8
		BACKSHIFT (same front)	0.171	0.537	1.393	3.958	8/8
Weak	1500	C1	0.795	0.926	0.447	0.781	8/8
		AJD (same front)	0.270	0.753	0.899	2.451	8/8
		BACKSHIFT (same front)	0.224	0.715	1.312	3.303	8/8
Weak	5000	C1	0.963	0.990	0.173	0.287	8/8
		AJD (same front)	0.555	0.911	0.653	1.618	8/8
		BACKSHIFT (same front)	0.654	0.905	0.613	1.633	8/8

raw-data `computeDelta()` step by its covariance identity and transposes the returned coefficient matrix to the paper's $B[\text{child}, \text{parent}]$ convention. Official MSS uses `rflperry/sparse_shift` pinned at `8620c5ccf7b7b28b2d2946d3f84c1b16b8fcfd39`, with `MinChange`, $\alpha = 0.05$, `scale_alpha=True`, and the authors' linear `invariant_residuals`/Whitney–Levene setting on the true observational CPDAG. Oracle-Z UT-IGSP uses `causaldag.unknown_target_igsp` with Gaussian CI/invariance tests, both levels 10^{-5}, depth 4, five runs, and no known intervention targets. Oracle-Z CD-NOD uses `causal-learn cdnod` with Fisher-z on pooled canonical Z plus the environment index. These are frozen implementation settings; none is tuned against truth here.

Canonical-coefficient error is used only for coefficient-estimating methods. AUPRC and AUROC are used only when a method supplies an authorized continuous edge-ranking score. Fixed CPDAG and I-MEC outputs are evaluated with the corresponding pair-state skeleton/orientation or equivalence-class metrics and are not coerced into fully directed DAGs. Coefficient-based methods additionally use the fixed threshold-sensitivity path; graph thresholds are fixed independently of evaluation performance.

H Additional Simulation Results

H.1 Primary coefficient recovery

The eight frozen graph topologies are shown in Figure 2. They are graph-to-graph design instances, not repeated-sampling replicates.

H.2 Native mixed-data baseline comparison

Table 4: Native mixed-data reference-environment skeleton recovery. The strong/weak labels are paired simulation labels; latent-PC and Copula PC use only raw mixed $X^{(0)}$ and do not use cross-environment heterogeneity. Values are mean skeleton F1 over 8/8 graph topologies in every cell.

	$n = 500$		$n = 1500$		$n = 5000$	
Method	Strong	Weak	Strong	Weak	Strong	Weak
latent-PC	0.875	0.875	0.899	0.899	0.909	0.909
Copula PC	0.924	0.925	0.934	0.955	0.947	0.948

Table 2: Per-graph coefficient AUPRC for the strong profile-separation regime. C1 failures are reported explicitly as `CV tuning failure`; they are neither zero-imputed nor silently omitted.

n	Graph	C1	AJD (same front)	BACKSHIFT (same front)
500	G1	1.000	0.783	0.833
500	G2	CV tuning failure	0.726	0.667
500	G3	1.000	0.864	0.938
500	G4	0.534	0.397	0.961
500	G5	0.989	0.653	0.848
500	G6	1.000	0.369	0.895
500	G7	0.968	0.735	0.933
500	G8	CV tuning failure	0.608	0.669
1500	G1	1.000	0.933	0.962
1500	G2	1.000	1.000	1.000
1500	G3	1.000	0.942	0.864
1500	G4	1.000	1.000	0.995
1500	G5	1.000	1.000	0.878
1500	G6	0.858	0.947	0.963
1500	G7	1.000	1.000	0.940
1500	G8	1.000	1.000	0.992
5000	G1	1.000	1.000	0.844
5000	G2	1.000	0.970	0.923
5000	G3	1.000	1.000	1.000
5000	G4	1.000	1.000	1.000
5000	G5	1.000	1.000	1.000
5000	G6	1.000	1.000	1.000
5000	G7	1.000	1.000	1.000
5000	G8	1.000	1.000	1.000

Table 3: Per-graph coefficient AUPRC for the weak profile-separation regime. C1 failures are reported explicitly as `CV tuning failure`; they are neither zero-imputed nor silently omitted.

n	Graph	C1	AJD (same front)	BACKSHIFT (same front)
500	G1	0.795	0.235	0.230
500	G2	CV tuning failure	0.230	0.214
500	G3	0.886	0.209	0.166
500	G4	0.413	0.251	0.192
500	G5	0.757	0.232	0.081
500	G6	0.495	0.213	0.162
500	G7	0.591	0.134	0.131
500	G8	0.569	0.170	0.193
1500	G1	0.777	0.177	0.178
1500	G2	1.000	0.257	0.124
1500	G3	0.870	0.274	0.123
1500	G4	0.545	0.321	0.405
1500	G5	1.000	0.421	0.313
1500	G6	0.789	0.225	0.205
1500	G7	0.641	0.201	0.221
1500	G8	0.737	0.285	0.220
5000	G1	0.712	0.412	0.541
5000	G2	1.000	0.546	0.425
5000	G3	1.000	0.415	0.853
5000	G4	0.995	0.917	0.699
5000	G5	1.000	0.334	0.598
5000	G6	1.000	0.861	0.956
5000	G7	1.000	0.383	0.332
5000	G8	1.000	0.570	0.831

Table 4 gives the complementary native mixed-data comparison. latent-PC and Copula PC remain competitive for reference-environment skeleton recovery, but they use only raw mixed $X^{(0)}$ and do not use cross-environment heterogeneity. Their skeleton F1 therefore is not a coefficient-AUPRC or full-DAG ranking against the same-front covariance backends; the strong/weak labels are paired simulation labels only. Although paired strong/weak cells share the same reference sample, Copula PC uses profile-specific deterministic method seeds for its Gibbs fit, so its stochastic outputs and skeleton F1 can differ slightly across paired cells.

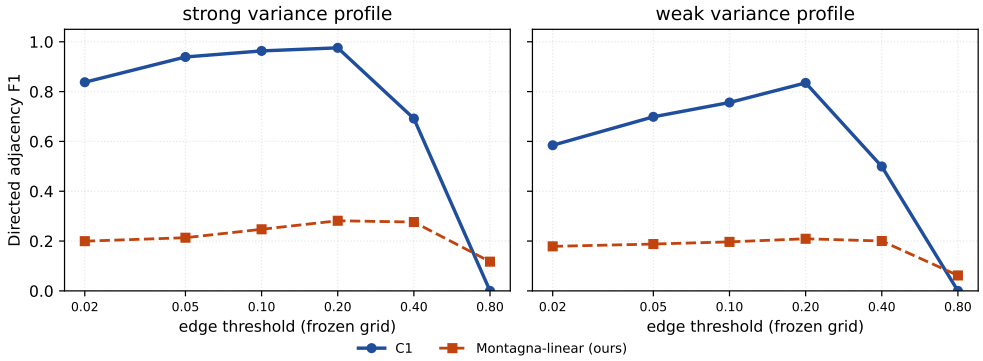


Figure 5: Threshold sensitivity at $n = 1500$. Directed-adjacency F1 is shown across the prespecified threshold grid under strong and weak variance-profile separation. This is a sensitivity analysis, not a truth-selected threshold procedure.

Table 5: Front-end diagnostic: estimated versus Oracle-Ω C1. Oracle-Ω uses the true covariance family and is information-advantaged, not an external competitor or a universal upper bound. It reuses the paired λ selected for the corresponding estimated-front C1 where applicable and is not independently retuned.

Profile	n	Estimated	Denom.	Oracle-Ω	Denom.
Strong	500	0.915	6/8	0.934	6/8
Strong	1500	0.982	8/8	1.000	8/8
Strong	5000	1.000	8/8	1.000	8/8
Weak	500	0.644	7/8	0.937	7/8
Weak	1500	0.795	8/8	0.930	8/8
Weak	5000	0.963	8/8	0.972	8/8

H.3 Threshold, front-end, scale, and specification diagnostics

Figure 5 reports the frozen threshold-sensitivity path; the graph threshold is not chosen by looking at truth.

H.4 Information-advantaged and specialization diagnostics

Table 6 reports Oracle-Z PC, UT-IGSP (Squires et al., 2020), and CD-NOD (Huang et al., 2020). Each receives the true canonical latent variables, so these rows are diagnostic rather than equal-information baselines.

The retained Montagna-related diagnostic is our own linear-Gaussian covariance specialization motivated by Montagna (2026), not an implementation of the authors' official estimator.

The official MSS diagnostic (Perry et al., 2022) is supplied the true canonical variables and the true observational MEC. Consequently its skeleton is fixed by construction and is not used as evidence of a fair recovery advantage.

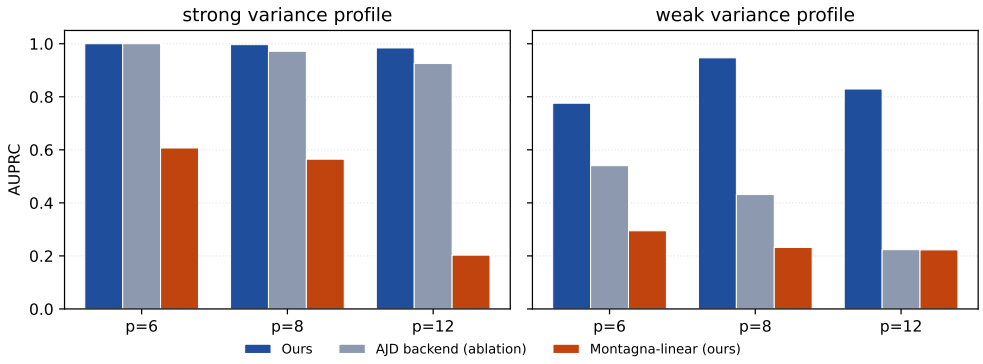


Figure 6: Dimension/scale panel at $n = 1500$. Coefficient AUPRC is reported for $p \in \{6, 8, 12\}$ under strong and weak variance-profile separation. This experiment uses a separate scale-panel design and is not directly comparable to the primary $p = 10$ panel.

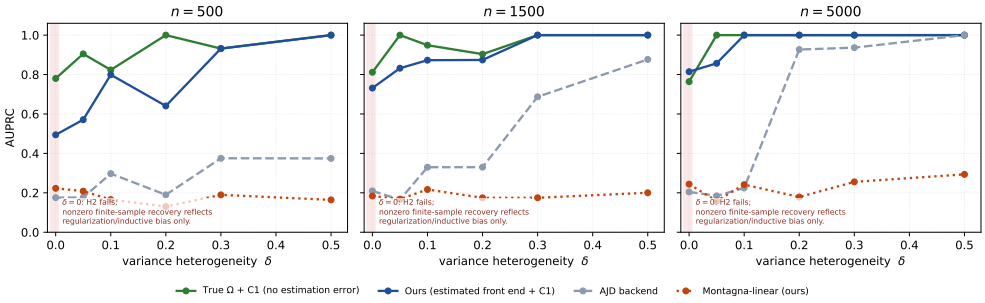


Figure 7: Recovery as a function of between-environment variance-profile separation. Results are shown for $n = 500, 1500, 5000$. At $\delta = 0$, H2 fails and the full-DAG identification theorem does not apply; nontrivial recovery at that particular parameter point should not be interpreted as establishing identification without H2.

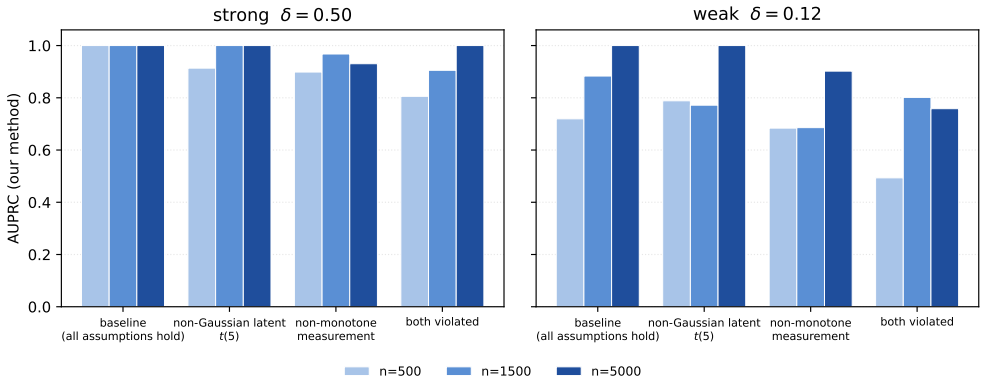


Figure 8: Misspecification robustness under strong and weak variance-profile separation. The panels compare the primary model with non-Gaussian latent variables, non-monotone measurements, and simultaneous violation of both assumptions across $n = 500, 1500, 5000$. These experiments assess empirical robustness outside the exact model class and do not enlarge the scope of the population theorems.

Table 6: Information-advantaged Oracle-Z structural diagnostics. All three methods receive the true canonical latent variables; the displayed common metric is mean undirected skeleton F1 over 8/8 graph topologies. These are not equal-information mixed-data baselines.

Profile	n	Oracle-Z PC	UT-IGSP	CD-NOD
Strong	500	0.957	0.930	0.990
Strong	1500	0.963	0.990	0.970
Strong	5000	0.964	0.990	0.972
Weak	500	0.957	0.935	0.974
Weak	1500	0.963	0.954	0.981
Weak	5000	0.964	0.974	0.989

Table 7: Montagna-linear specialization (ours). The shared-front row is our linear-Gaussian covariance specialization motivated by Montagna (2026), not the authors' official estimator; Oracle-Ω is an information-advantaged diagnostic. Values are coefficient AUPRC.

Profile	n	Same front	Oracle-Ω
Strong	500	0.199	1.000
Strong	1500	0.245	1.000
Strong	5000	0.353	1.000
Weak	500	0.184	1.000
Weak	1500	0.154	1.000
Weak	5000	0.182	1.000

Table 8: MSS (official; Oracle-Z + true observational MEC). Because the true observational MEC is supplied as the search space, skeleton F1 is constructionally 1 and is not reported as a recovery advantage. We instead show mean CPDAG SHD and compelled-orientation F1 over 8/8 graphs.

Profile	n	CPDAG SHD	Compelled orient. F1
Strong	500	3.000	0.780
Strong	1500	2.125	0.831
Strong	5000	2.625	0.804
Weak	500	2.375	0.818
Weak	1500	2.375	0.816
Weak	5000	2.750	0.800

H.5 Runtime context

Table 9 reports the frozen runtime diagnostic available for the retained methods.

Table 9: Recorded V2 runtime diagnostic for retained arms available in the frozen runtime summary. Inputs and implementations differ, so these values provide computational context rather than a normalized complexity benchmark. C1, Copula PC, and UT-IGSP were not present in the frozen T8 runtime summary and are not backfilled post hoc.

Method	Mean seconds	Max seconds	Runs
latent-PC	7.525	11.810	48
AJD (same front)	6.223	9.529	48
BACKSHIFT (same front)	5.151	9.321	48
Montagna-linear specialization (ours)	4.010	4.682	48
Oracle-Z PC	6.895	12.920	48
Oracle-Z CD-NOD	0.240	0.318	48
Oracle-Ω AJD	5.638	7.546	48
Oracle-Ω BACKSHIFT	4.758	8.427	48
Oracle-Ω specialization	4.064	4.834	48
MSS (official; Oracle-Z + true MEC)	793.012	2776.000	48

I BRFSS 2015 Application Details

I.1 Data provenance, mixed typing, and regional environments

The official CDC BRFSS 2015 raw file contains 441,456 records and 330 variables (Centers for Disease Control and Prevention, 2016). After the frozen raw-missing filter and author recodes, the downstream 22-variable Diabetes Health Indicators representation also analyzed by Zhou et al. (2026) contains 253,680 records. Excluding reporting areas outside the four Census regions removes 1,020 records with area code 66 and 3,310 with area code 72 (Guam and Puerto Rico in this pipeline), yielding the four-region analysis size 249,350. Following the paper-level typing of Zhou et al. (2026), the analysis treats BMI, MentHlth, and PhysHlth as continuous, Diabetes_012, GenHlth, Age, Education, and Income as ordinal, and 14 remaining variables as binary. In the released representation, however, BMI is rounded by the author recipe $\mathrm{round}(_BMI5/100)$ and MentHlth/PhysHlth are health-day counts with ties. Therefore the exact atomless continuous-margin observation model does not literally hold for these three coordinates; the BRFSS analysis is an empirical robustness/application illustration rather than validation of the exact population assumptions. The application likewise does not validate the model's zero latent-mean restriction across regions; observed regional marginal shifts alone do not distinguish latent mean changes from variance-driven changes under the measurement model. We treat Census regions as naturally occurring environments; no empirical claim is made that they prove the common-B assumption.

I.2 Adaptive front end and positive definiteness

The latest adaptive-regions4 analysis uses a real-data-specific multilevel-ordinal correlation bridge. On the relevant reference/training data, for each multilevel ordinal variable it selects the category boundary whose cumulative probability is closest to 0.5 and then freezes that c_{cor} value for validation/test. The scale bridge remains fixed at $c_{\mathrm{sc}} = 1$. This is an adaptive real-data sensitivity and not the unmodified frozen primary simulation front end. All relevant covariance families in the latest adaptive-regions4 run are strictly positive definite. No 10^{-4} identity shrinkage is used. The earlier fixed-cutpoint-plus-shrinkage pipeline is archived provenance only and is not mixed into the results reported here.

I.3 Cross-validation, held-out fit, and bootstrap stability

For this subsection only, we index the four Census regions by $e = 1, \ldots, 4$. The prespecified real-data CV grid is $c_\lambda \in \{0.25, 0.50, 1, 2, 4\}$. The selected value is $c_\lambda = 0.25$, which is the smallest value

Table 10: BRFSS four-region analysis counts after excluding non-Census-region reporting areas.

Region	Observations
Northeast	46,199
Midwest	71,864
South	71,452
West	59,835
Total	249,350

Table 11: Adaptive multilevel-ordinal correlation cutpoints c_{cor} in the latest BRFSS analysis. Each cutpoint is selected on the relevant reference training data by choosing the category boundary whose cumulative mass is closest to 0.5, then frozen for validation/test. The scale cutpoint remains $c_{\mathrm{sc}} = 1$.

Variable	c_{cor}
Diabetes_012	1
GenHlth	2
Age	8
Education	5
Income	6

in the grid; we therefore report it explicitly as a boundary selection and do not imply an interior optimum. C1 estimator and AJD both return successful main and holdout fits. The reported held-out covariance loss is exactly the unpenalized profiled criterion

$$Q_{\mathrm{hold}}(M) = \sum_{e=1}^{4} \frac{1}{4} \left[\sum_{i=1}^{22} \log\Big(m_i^{\top} \widehat{\Omega}_e^{\mathrm{hold}} m_i \Big) - 2 \log |\det M| \right],$$

with $M = I - \widehat{B}^{\circ}$ and the same frozen adaptive held-out covariance family $\{\widehat{\Omega}_e^{\mathrm{hold}}\}$ used for both methods; no graph threshold or ℓ_1 penalty enters this score. The held-out loss is the unpenalized profiled covariance criterion underlying C1 estimator and is therefore reported as a descriptive held-out fit diagnostic rather than a method-neutral model-selection score. It equals -5.814457 for C1 estimator and -5.750857 for AJD, so this comparison is conditional on a shared covariance front end.

Twenty respondent-stratified regional bootstrap refits succeed for C1 estimator (20/20) and AJD (20/20). These resamples characterize edge-selection stability conditional on the analysis procedure; they are not complex-survey population confidence intervals. Across the prespecified C1 estimator coefficient-threshold grid, the directed edge counts are 216, 133, 57, 17, 2, and 0 at $\tau = 0.02, 0.05, 0.10, 0.20, 0.40, 0.80$, respectively. The $\tau = 0.20$ graph is only an illustrative member of this prespecified grid and is not an optimized threshold. It contains 17 directed edges, of which 13 have bootstrap selection frequency at least 0.80; the stable 13-edge subgraph is acyclic.

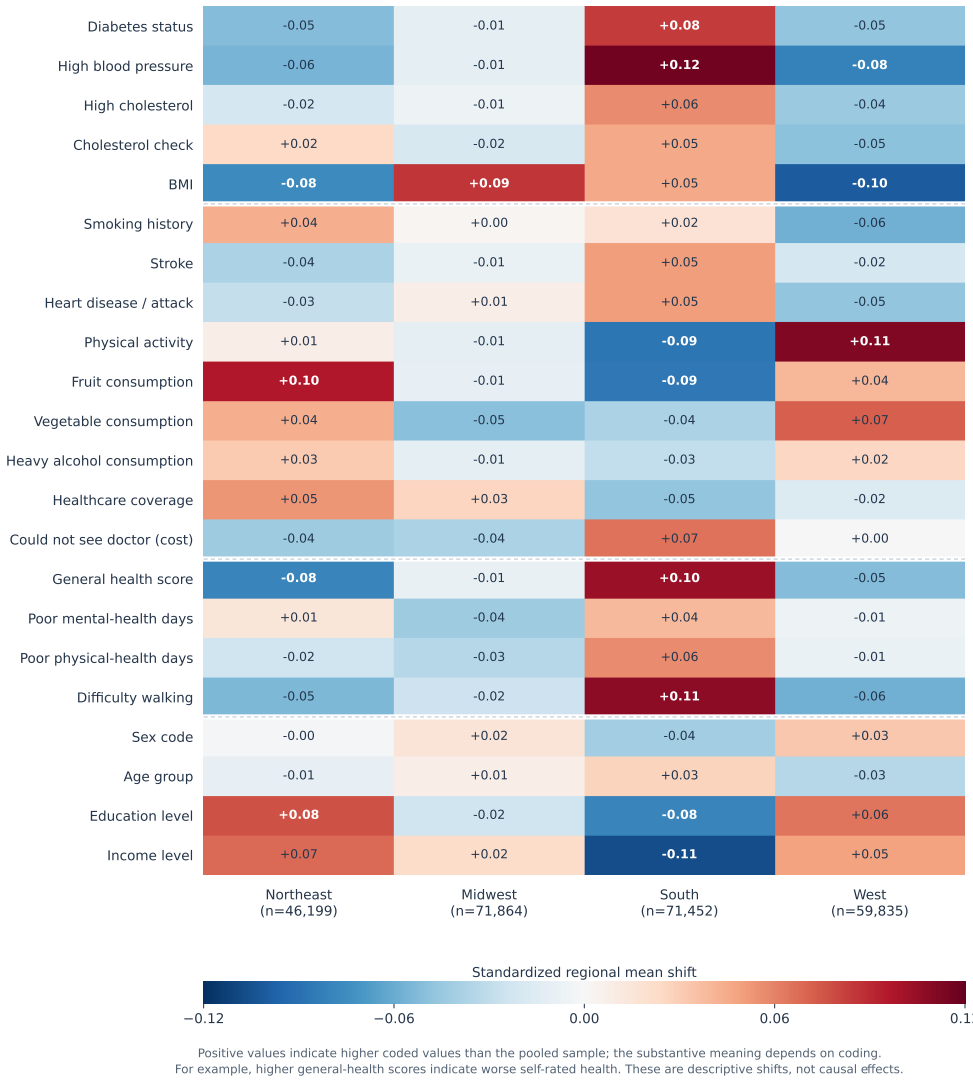


Figure 9: Regional distribution shifts in the BRFSS analysis sample. Each cell reports (regional mean − pooled mean)/pooled SD for one variable and Census region. The regions are naturally occurring environments rather than randomized interventions, and the displayed shifts are descriptive rather than causal effects.

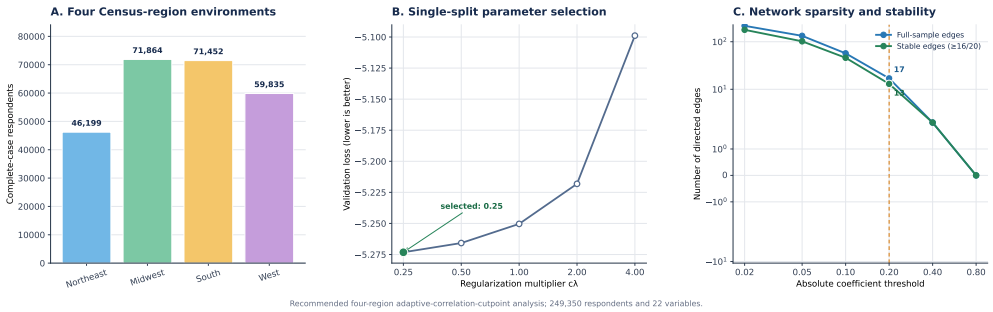


Figure 10: BRFSS sample, tuning, and sparsity diagnostics. (A) Complete-case sample sizes in the four Census-region environments. (B) Held-out covariance loss over the prespecified c_λ grid, with $c_\lambda = 0.25$ selected. (C) Numbers of full-sample and bootstrap-stable directed edges across coefficient thresholds.

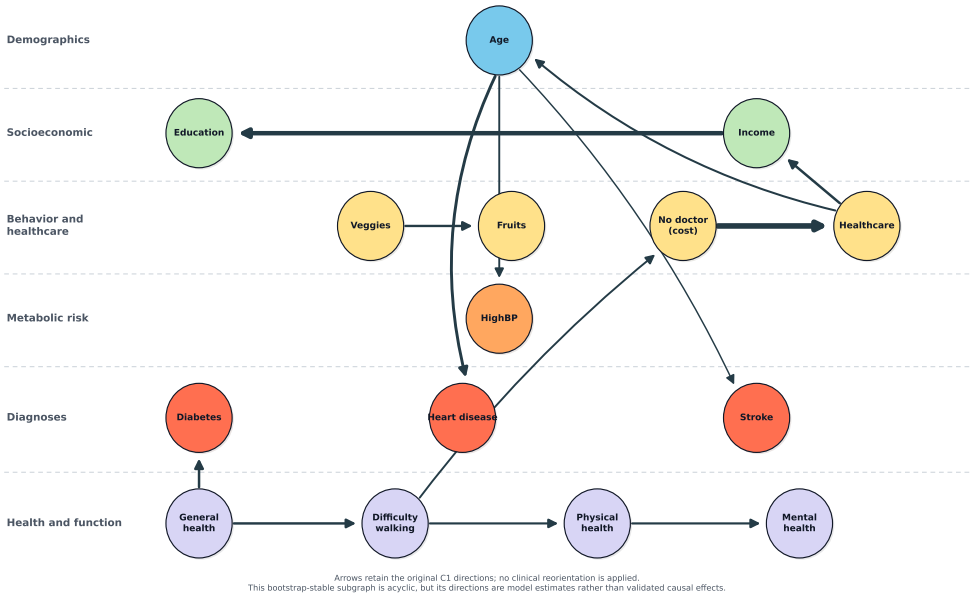


Figure 11: Bootstrap-stable 13-edge C1 estimator subgraph at $\tau = 0.20$. The graph retains the original model-estimated C1 estimator directions for edges with bootstrap selection frequency at least 0.80; the resulting subgraph is acyclic. Directions are shown as procedure-conditional model estimates, not as clinically validated causal effects.