JIGSAW-CRL: RECOVERING GLOBAL LATENT CAUSAL ORDER FROM FRAGMENTED MULTI-CLIENT INTERVENTIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

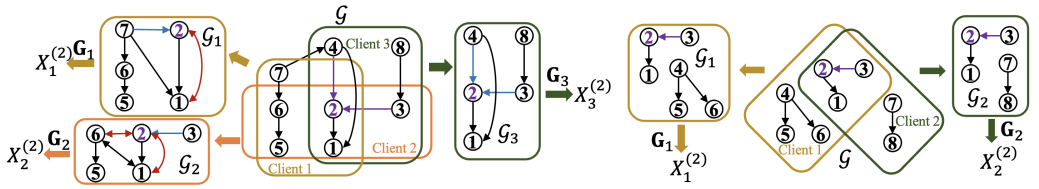
Causal representation learning (CRL) aims to recover latent causal variables and their structural relations from high-dimensional observations. Existing CRL methods typically assume that all environments are defined over the same latent variables, or at least share a common latent representation space. We study a fragmented multi-client setting, where multiple clients interact with the same global latent causal system but each client only accesses and intervenes on a subset of the latent variables. In this regime, marginalizing unused latent variables induces bidirected edges, so a single client no longer admits a node-wise latent causal graph, and the global latent causal order must be recovered by assembling client-specific structural fragments. We propose **Jigsaw-CRL**, a framework for recovering global latent causal order from such fragmented interventions. Under soft interventions, differences between precision matrices across environments exhibit a low-rank structure governed by latent ancestor relations. This enables recovery, for each client, of a block partition, the corresponding block-level ancestral order, and latent subspaces, and then assembly of these fragments into the global node-level latent causal order. We establish identifiability guarantees, develop practical algorithms, and validate the framework on synthetic data. Our codes are available on `https://anonymous.4open.science/r/code-for-Jigsaw-CRL-7B26`

1 INTRODUCTION

Causal representation learning (CRL) aims to recover high-level latent causal variables and their structural relations from high-dimensional observations (Schölkopf et al., 2021; Jin & Syrgkanis, 2024; Buchholz et al., 2023; Varici et al., 2024). Most existing CRL frameworks assume that all observations are defined over the same latent variables, or at least share a common latent representation space (Zhang et al., 2024; Buchholz et al., 2023). Under this shared-space assumption, the natural target is typically a node-level latent causal structure, or a standard equivalence class thereof. We call this single-client setting. In many multi-client systems, however, no single client has access to the full latent causal system. Instead, multiple clients interact with the same global latent causal system, while each client accesses and intervenes on only a subset of the latent variables. Each client therefore provides only a fragmented structural view of the global system, as illustrated in Figure 1a. We call this multi-client setting. This raises a basic question: **can we recover the global latent causal order by assembling fragmented causal information from multiple clients?**

This question is not a straightforward multi-client extension of existing CRL, because fragmented access changes the local causal object itself. From the perspective of a client, the latent variables outside its accessed subset are marginalized out, which can induce directed shortcuts and bidirected dependencies among the remaining latent variables. Thus, the client-specific latent structure is no longer an ordinary DAG, and node-wise latent causal recovery is generally impossible from a single fragmented client. This differs from standard CRL methods (Schölkopf et al., 2021; Squires et al., 2023; Buchholz et al., 2023; Varici et al., 2024; Zhang et al., 2024), which assume a shared latent variable set across environments. It also differs from recent multi-domain or multi-view CRL works (Sturma et al., 2023; Yao et al., 2024): Sturma et al. (2023) assume causally separated shared and client-specific components, as illustrated in Figure 1b, while Yao et al. (2024) allow subset-dependent views but focus on shared latent information rather than ancestral relations or global causal structure.

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.



(a) Multi-client setting (our setting) (b) Multi-domain setting

Figure 1: **Fragmented client views of latent causal systems under different multi-client settings.** All nodes in the figure represent latent variables. For client k, $X_k^{(j)}$ denotes the observed data after intervention j. Purple highlights the intervened node 2 together with the edges affected by the intervention. (a) **Multi-client setting.** Different clients use different subsets of latent variables from the same global causal system. Marginalizing the missing latent variables induces client-specific mixed graphs, where bidirected edges appear in addition to directed relations. In a client-specific induced graph, the intervention on node 2 may manifest as changes in directed edges (blue) and bidirected edges (red), corresponding to $\mathbf{d}_k^{(j)}$ and $\mathbf{u}_k^{(j)}$ in Assumption 2. (b) **Multi-domain setting.** Adopted in Sturma et al. (2023), latent variables are partitioned into a shared global component (e.g., $\{1, 2, 3\}$) and client-specific local components (e.g., $\{4, 5, 6\}$ and $\{7, 8\}$). No edges exist across different local components, nor between global and local components. Consequently, client-specific induced graphs do not generate bidirected edges or extra directed edges after marginalization. This can be viewed as a special case of our more general setting.

This leads to a first question: **when missing latent variables induce bidirected dependencies through marginalization, what causal structure remains identifiable to a fragmented client?**

We show that the natural client-level identifiable object is not a node-wise causal graph, but a block-structured mixed object: a block partition induced by directed ancestry and bidirected dependencies, together with a block-level ancestral order. Nodes within a block are structurally entangled for a single client, whereas ancestral relations between blocks remain identifiable. To recover this object, we propose **Jigsaw-CRL**, which exploits the fact that single-node soft interventions induce low-rank precision perturbations whose row spaces follow block-level ancestor relations. Jigsaw-CRL first extracts from each client a block-level causal fragment, including locally inseparable latent groups, their block-level order, and the associated latent subspaces. It then assembles complementary fragments across clients: a Hasse edge hidden inside a non-singleton block for one client may appear as a singleton-block parent relation for another, enabling recovery of the global node-level latent causal order. Thus, Jigsaw-CRL yields a two-level identifiability picture: *each client reveals a block-level causal fragment, while multiple complementary fragments recover the global node-level causal order*.

Our Contributions. 1) Fragmented multi-client CRL. We introduce a setting where multiple clients partially access and intervene on the same global latent causal system, and show that marginalizing unused latent variables changes each client's local causal object from a DAG to a block-structured mixed graph; **2) Client-level block identifiability.** We study client-level identifiability in the presence of bidirected dependencies induced by latent marginalization, a regime not covered by prior CRL methods. We prove that soft interventions induce low-rank precision perturbations that identify each client's block partition, block-level ancestral order, and latent subspaces up to intrinsic block-level ambiguity in Theorem 1 and Corollary 1; **3) Local-to-global assembly.** We show that complementary client fragments can recover the global latent ancestral relation in Theorem 2: under a cross-client coverage condition, each direct step in the global causal order is resolved by at least one client; **4) Algorithms and experiments.** We develop recovery algorithms based on projected low-rank precision perturbations and validate Jigsaw-CRL on synthetic fragmented multi-client systems.

Organization. Due to space constraints, we defer related work to Appendix A. Section 2 introduces the fragmented multi-client CRL setting and the induced block structure. Sections 3 and 4 establish client-level identifiability and global assembly, respectively. Section 5 reports experimental results.

2 Problem Setup and Client-Specific Block Structure

This section formalizes the fragmented multi-client CRL setting and identifies the causal object that is locally visible to each client. We first define the global latent causal system and the client-specific

observation model. We then show that marginalizing the latent variables unused by a client induces a mixed graph rather than a DAG. This motivates a client-specific block structure, which will be the basic unit of local identifiability in the sequel.

Notation. In the rest of the paper, we use non-bold lowercase letters to denote scalars (e.g., p), bold lowercase letters to denote vectors (e.g., $\mathbf{d}$), non-bold uppercase letters to denote random variables (e.g., Z), and bold uppercase letters to denote matrices (e.g., $\mathbf{A}$). For an index set S, $[\mathbf{A}]_{S,:}$ and $[\mathbf{A}]_{:,S}$ denote the submatrices of $\mathbf{A}$ consisting of the rows and columns indexed by S, respectively, and $[\mathbf{d}]_S$ denotes the subvector of $\mathbf{d}$ consisting of the elements indexed by S. We use $\mathrm{row}(\{\mathbf{A}_1, \ldots, \mathbf{A}_p\})$ to denote the row space jointly spanned by the rows of the matrices $\mathbf{A}_1, \ldots, \mathbf{A}_p$, and use $\mathrm{rank}(\mathbf{A})$ to denote the rank of the matrix $\mathbf{A}$. We use $\langle \mathbf{a}_i | i = 1, \ldots, p \rangle$ to denote the space spanned by $\{\mathbf{a}_i\}_{i=1}^p$. For a positive integer p, let $[p] \triangleq \{1, \ldots, p\}$.

2.1 Problem formulation: fragmented multi-client CRL

Global latent SEM. Let $Z = (Z_{(1)}, \ldots, Z_{(p)})$ denote the global latent variables, generated by the linear structural equation model

$$Z = \mathbf{A}Z + E. \tag{1}$$

Since Z is latent, we may assume without loss of generality that $Z_j \to Z_i$ implies $j > i$. Under this ordering, $\mathbf{A} \in \mathbb{R}^{p \times p}$ is strictly upper triangular. The noise vector E has mean zero and diagonal covariance matrix $\mathbf{\Sigma} \triangleq Cov(E)$. The matrix $\mathbf{A}$ defines a global latent DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = [p]$ and $\mathcal{E} = \{j \to i : \ A_{ij} \neq 0\}$. It induces the global ancestral relation $\prec$ on $\mathcal{V}$: we write $i \prec j$ if there exists a directed path from j to i in $\mathcal{G}$, and write $i \preceq j$ if either $i \prec j$ or $i = j$.

Fragmented clients. We consider K clients. For client $k \in [K]$, the index set $O_k \subset [p]$ specifies the latent variables directly used by client k, and $M_k = [p] \setminus O_k$ denotes the latent variables not directly used by this client. Let $Z_k \triangleq (Z_{(i)})_{i \in O_k} \in \mathbb{R}^{|O_k|}$, $p_k \triangleq |O_k|$. The client-specific observation is

$$X_k = \mathbf{G}_k Z_k, \tag{2}$$

where $\mathbf{G}_k \in \mathbb{R}^{n_k \times p_k}$ is full column rank with $n_k > p_k$. We write $\mathbf{H}_k = \mathbf{G}_k^\dagger$ for its Moore–Penrose pseudoinverse. Thus all clients share the same global latent system, but each client directly uses only a subset of the latent variables.

Soft interventions. For each client k, we observe one observational regime indexed by 0 and multiple soft interventional regimes indexed by $j \in \mathcal{J}_k \subset [p]$. We assume that client k only intervenes on nodes in O_k, and that each intervention targets a single node. Let t_j denote the unknown target node of intervention j. We model the corresponding intervention on the global latent system by

$$\tilde{\mathbf{A}}_k^{(j)} = \mathbf{A} + \mathbf{e}_{t_j} \Delta \mathbf{a}_k^{(j)^T}, \qquad \tilde{\mathbf{\Sigma}}_k^{(j)} = \mathbf{\Sigma} + \Delta\sigma_k^{(j)} \mathbf{e}_{t_j} \mathbf{e}_{t_j}^T,$$

where $\mathbf{e}_{t_j}$ is the vector whose t_j-th entry is 1 and all others are 0, $\Delta \mathbf{a}_k^{(j)} \in \mathbb{R}^p$ is supported on $pa(t_j)$ and represents the intervention-induced perturbation of the parent effects of node t_j, and $\Delta\sigma_k^{(j)} \in \mathbb{R}$ perturbs its exogenous noise variance. We write $X_k^{(j)}$ for the observation of client k under intervention j. Throughout the paper, we work with the second-order information $\mathbf{\Theta}_k^{(j)} \triangleq \mathbb{E}(X_k^{(j)} X_k^{(j)^\top})^\dagger$ for $j \in \mathcal{J}_k \cup \{0\}$, where $X_k^{(0)} \triangleq X_k$ denotes the observational regime.

Recovery goal. Given the collection of precision matrices $\{\mathbf{\Theta}_k^{(j)}\}_{k=1,\ldots,K,\ j \in \mathcal{J}_k \cup \{0\}}$, the fragmented multi-client CRL problem has two recovery levels. At the client level, we first characterize and recover the structural information that remains identifiable from each client's fragmented access to the global latent system. At the global level, we assemble these client-specific fragments across $k = 1, \ldots, K$ to recover the global latent ancestral relation $\prec$.

2.2 Client-specific induced graph and block structure

Since client k directly uses only $Z_k = Z_{O_k}$, the variables in $M_k = [p] \setminus O_k$ are marginalized out from this client's perspective. This marginalization has two effects. First, directed paths that pass through variables in M_k can appear as directed shortcuts among variables in O_k. Second, shared dependence through the marginalized variables can induce correlated effective noises, represented as bidirected edges. After the normalization discussed in Appendix B, we work with $\mathbf{\Sigma} = \mathbf{I}$. Thus the client-level latent object is generally a mixed graph rather than a DAG.

Proposition 1. *For Z generated by Eq. equation 1, let $Z_k = (Z_{(i)})_{i \in O_k}$. Then*

$$Z_k = \mathbf{A}_k Z_k + E_k, \tag{3}$$

where

$$\mathbf{A}_k = [\mathbf{A}]_{O_k,O_k} + [\mathbf{A}]_{O_k,M_k}(\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-1}[\mathbf{A}]_{M_k,O_k}, \tag{4}$$

$$E_k = E_{O_k} + [\mathbf{A}]_{O_k,M_k}(\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-1}E_{M_k}. \tag{5}$$

Moreover, E_k has mean zero and covariance

$$\mathbf{\Sigma}_k = [\mathbf{\Sigma}]_{O_k,O_k} + [\mathbf{A}]_{O_k,M_k}(\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-1}[\mathbf{\Sigma}]_{M_k,M_k}(\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-T}[\mathbf{A}]^T_{O_k,M_k}. \tag{6}$$

Proposition 1 shows that after marginalizing out M_k, client k no longer inherits a latent DAG on O_k. Instead, it induces a mixed graph $\mathcal{G}_k = (\mathcal{V}_k, \mathcal{E}_k, \mathcal{D}_k)$ with $\mathcal{V}_k = O_k$, directed edges $\mathcal{E}_k = \{j \to i : [\mathbf{A}_k]_{\rho_k(i),\rho_k(j)} \neq 0\}$, and bidirected edges $\mathcal{D}_k = \{j \leftrightarrow i : i \neq j,\ [\mathbf{\Sigma}_k]_{\rho_k(i),\rho_k(j)} \neq 0\}$, where ρ_k denotes the local index map from nodes in O_k to coordinates in Z_k. This mixed graph induces a client-specific ancestral relation $\prec_k$ on $\mathcal{V}_k$: we write $i \prec_k j$ if there exists a directed path from j to i in $\mathcal{G}_k$, and $i \preceq_k j$ if either $i \prec_k j$ or $i = j$. By the proof of Proposition 1 in Appendix E, $\prec_k$ is exactly the restriction of the global ancestral relation $\prec$ to O_k.

For notational consistency, set $\mathbf{A}_k^{(0)} \triangleq \mathbf{A}_k$ and $\mathbf{\Sigma}_k^{(0)} \triangleq \mathbf{\Sigma}_k$. For $j \in \mathcal{J}_k$, marginalizing out M_k under intervention j yields $\mathbf{A}_k^{(j)}$ and $\mathbf{\Sigma}_k^{(j)}$ by replacing $\mathbf{A}$ and $\mathbf{\Sigma}$ in Eqs. equation 4 and equation 6 with $\tilde{\mathbf{A}}_k^{(j)}$ and $\tilde{\mathbf{\Sigma}}_k^{(j)}$, respectively. Then, by Proposition 1 and Lemma A.3, for all $j \in \mathcal{J}_k \cup \{0\}$,

$$\mathbf{\Theta}_k^{(j)} = \mathbf{H}_k^\top (\mathbf{I} - \mathbf{A}_k^{(j)})\mathbf{\Sigma}_k^{(j)^{-1}}(\mathbf{I} - \mathbf{A}_k^{(j)})^\top \mathbf{H}_k. \tag{7}$$

The mixed graph $\mathcal{G}_k$ contains both directed ancestral relations and bidirected dependencies induced by latent marginalization. Consequently, node-wise structure is no longer the right unit of description at the client level. We therefore introduce a client-specific block structure.

Definition 1 (Reachability and mutual reachability)**.** *For two nodes $i, j \in \mathcal{V}_k$, we say that i is* reachable *to j, written $i \rightsquigarrow_k j$, if there exists a sequence of nodes $i = v_0, v_1, \ldots, v_m = j$, such that for every $\ell = 0, \ldots, m-1$, either $v_\ell \prec_k v_{\ell+1}$ or $v_\ell \leftrightarrow v_{\ell+1} \in \mathcal{D}_k$. Two nodes $i, j \in \mathcal{V}_k$ are said to be* mutually reachable *if $i \rightsquigarrow_k j$ and $j \rightsquigarrow_k i$.*

Definition 2 (Block)**.** *For client k, a* block *is a maximal subset $C \subseteq \mathcal{V}_k$ such that every pair of nodes in C is mutually reachable.*

Since mutual reachability is an equivalence relation, all blocks form a partition $\mathcal{C}_k = \{C_{k,1}, \ldots, C_{k,m_k}\}$ of $\mathcal{V}_k$. For example, in Figure 1, the block partition for client k is $\mathcal{C}_k = \big\{\{1,2\}, \{5\}, \{6\}, \{7\}\big\}$.

Intuitively, nodes within the same block are structurally entangled by directed ancestry and bidirected confounding, while a meaningful ancestral order remains well defined across blocks.

Definition 3 (Block-level order)**.** *For two distinct blocks $C_{k,a}, C_{k,b} \in \mathcal{C}_k$, write $C_{k,a} \prec_k^{\mathrm{blk}} C_{k,b}$ if there exist $u \in C_{k,a}$ and $v \in C_{k,b}$ such that $u \prec_k v$. Also write $C_{k,a} \preceq_k^{\mathrm{blk}} C_{k,b}$ if either $C_{k,a} \prec_k^{\mathrm{blk}} C_{k,b}$ or $C_{k,a} = C_{k,b}$. A block $C_{k,b}$ is a* block-level parent *of $C_{k,a}$ if $C_{k,a} \prec_k^{\mathrm{blk}} C_{k,b}$ and there is no block $C_{k,m} \in \mathcal{C}_k$ such that $C_{k,a} \prec_k^{\mathrm{blk}} C_{k,m} \prec_k^{\mathrm{blk}} C_{k,b}$.*

Ancestor set notation. For the global graph, define $anc(i) \triangleq \{j \in \mathcal{V} : i \prec j\}$ and $pa(i) \triangleq \{j \in \mathcal{V} : j \to i \in \mathcal{E}\}$, together with $Anc(i) \triangleq anc(i) \cup \{i\}$ and $Pa(i) \triangleq pa(i) \cup \{i\}$. For client k and node $i \in \mathcal{V}_k$, define $anc_k(i) \triangleq \{j \in \mathcal{V}_k : i \prec_k j\}$, $pa_k(i) \triangleq \{j \in \mathcal{V}_k : j \to i \in \mathcal{E}_k\}$, $Anc_k(i) \triangleq anc_k(i) \cup \{i\}$, and $Pa_k(i) \triangleq pa_k(i) \cup \{i\}$. We further define the spouse set $sp_k(i) \triangleq \{j \in \mathcal{V}_k : j \leftrightarrow i \in \mathcal{D}_k\}$. For any block $C \in \mathcal{C}_k$, define $anc_k^{\mathrm{blk}}(C) \triangleq \{C' \in \mathcal{C}_k : C \prec_k^{\mathrm{blk}} C'\}$, $pa_k^{\mathrm{blk}}(C) \triangleq \{C' \in \mathcal{C}_k : C' \text{ is a block-level parent of } C\}$, $Anc_k^{\mathrm{blk}}(C) \triangleq anc_k^{\mathrm{blk}}(C) \cup \{C\}$, and $Pa_k^{\mathrm{blk}}(C) \triangleq pa_k^{\mathrm{blk}}(C) \cup \{C\}$. For a node $v \in \mathcal{V}_k$, let $C_k(v)$ denote the block containing v.

The next proposition clarifies why block ancestry is the right coarse structural object: across blocks, non-ancestral relations remain absent, while block-level parents still witness genuine node-level parent relations.

Proposition 2. *For each client k, let $C_{k,i}, C_{k,j} \in \mathcal{C}_k$ be two blocks.*

(i) If $C_{k,j} \notin anc_k^{\mathrm{blk}}(C_{k,i})$, then for $\forall u \in C_{k,j}$ and $v \in C_{k,i}$, we have $u \notin anc_k(v)$.

(ii) If $C_{k,j} \in pa_k^{\mathrm{blk}}(C_{k,i})$, then there $\exists u \in C_{k,j}$ and $v \in C_{k,i}$ such that $u \in pa_k(v)$.

For example, in client $k = 1$ of Figure 1a, $\{7\} \in pa_k^{\mathrm{blk}}(\{1,2\})$, $\{7\} \in anc_k^{\mathrm{blk}}(\{5\})$, but $\{7\} \notin pa_k^{\mathrm{blk}}(\{5\})$. Without loss of generality, for each client k, we assume that if $C_{k,i} \prec_k^{\mathrm{blk}} C_{k,j}$, then $\rho_k(v) < \rho_k(u)$ for all $v \in C_{k,i}$ and $u \in C_{k,j}$.

3 Client-level Identifiability from Low-rank Precision Perturbations

Section 2 shows that latent marginalization changes each client's local object from a node-wise DAG to a block-structured mixed graph. This section answers the client-level question raised in Section 1: **when missing latent variables induce bidirected dependencies through marginalization, what causal structure remains identifiable to a fragmented client?** We show that a single client can identify its block partition, block-level ancestral order, and associated latent subspaces up to intrinsic block-level ambiguity. The argument combines a block-adapted coordinate system, low-rank precision perturbations under soft interventions, and an iterative recovery rule based on projected row-space dimensions.

3.1 Client-level assumptions

We first state the assumptions for client-level identifiability.

Assumption 1 (Intervention indexing and coverage)**.** *We assume: (i)* client-level intervention coverage: *for each client k, $\mathcal{J}_k$ is in bijection with $\mathcal{V}_k$, so each node in $\mathcal{V}_k$ is targeted by exactly one intervention label; and (ii)* cross-client label alignment: *interventions from different clients that target the same global latent node share the same label.*

Let $\psi : j \mapsto t_j$ denote the mapping from the global intervention index to its target node, so that ψ is a permutation on $[p]$.

Assumption 2 (Intervention genericity)**.** *For each client k and intervention $j \in \mathcal{J}_k$ with target $t_j \in C_{k,\ell} \subseteq O_k$, define*

$$\mathbf{d}_k^{(j)} \triangleq [\Delta \mathbf{a}_k^{(j)}]_{O_k} + [\mathbf{A}]_{M_k,O_k}^T (\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-T} [\Delta \mathbf{a}_k^{(j)}]_{M_k},$$
$$\mathbf{u}_k^{(j)} \triangleq [\mathbf{A}]_{O_k,M_k} (\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-1} [\mathbf{\Sigma}]_{M_k,M_k} (\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-T} [\Delta \mathbf{a}_k^{(j)}]_{M_k}.$$

We assume that the intervention parameters are chosen such that:

(i) $[\mathbf{d}_k^{(j)}]_{\rho_k(p)} \neq 0$ *for every $p \in pa_k(t_j)$;*

(ii) $[\mathbf{u}_k^{(j)}]_{\rho_k(s)} \neq 0$ *for every $s \in sp_k(t_j)$;*

(iii) for any $\mathcal{B} \subsetneq C_{k,\ell}$ and any $i \in \mathcal{B}$, let $\mathcal{N}_i$ be the support of

$$[\mathbf{u}_k^{(\psi^{-1}(i))}]_{\rho_k(C_{k,\ell})} + [\mathbf{\Sigma}_k]_{\rho_k(C_{k,\ell}),\rho_k(C_{k,\ell})} \big(\mathbf{I} - [\mathbf{A}_k]_{\rho_k(C_{k,\ell}),\rho_k(C_{k,\ell})}^T\big)^{-1} [\mathbf{d}_k^{(\psi^{-1}(i))}]_{\rho_k(C_{k,\ell})}.$$

Then $\left|\bigcup_{i \in \mathcal{B}} \mathcal{N}_i\right| > |\mathcal{B}|$.

Assumption 1 ensures that intervention environments are sufficiently informative and consistently indexed across clients. Assumption 2 is a genericity condition that rules out pathological cancellations in the directed, bidirected, and within-block effects induced by soft interventions. The two assumptions are standard in spirit, and closely related conditions appear frequently in intervention-based latent causal discovery and CRL (Varici et al., 2025; Buchholz et al., 2023; Squires et al., 2023). More detailed discussion and motivation for these assumptions are provided in Appendix C.

3.2 Analytical coordinates: block RQ decomposition

To relate observable precision perturbations to the client-specific block structure, we introduce a block-adapted factorization of $\mathbf{H}_k$. This factorization is not computed by the learner, since $\mathbf{H}_k$ is

unknown. Rather, it provides a coordinate system in which row spaces can be described according to the block ancestral order. Its construction and further discussion are deferred to Appendix G.

Definition 4 (Block RQ decomposition)**.** *For client $k \in [K]$ with block ancestor structure $(\prec_k^{\mathrm{blk}}, \mathcal{C}_k)$, a block RQ decomposition of $\mathbf{H}_k \in \mathbb{R}^{p_k \times n_k}$ is a factorization*

$$\mathbf{H}_k = \mathbf{R}_k \mathbf{Q}_k,$$

where $\mathbf{R}_k \in \mathbb{R}^{p_k \times p_k}$ is block upper triangular with respect to $\mathcal{C}_k$, namely $[\mathbf{R}_k]_{\rho_k(C_{k,i}),\rho_k(C_{k,j})} = 0$ unless $C_{k,i} \preceq_k^{\mathrm{blk}} C_{k,j}$, and $\mathbf{Q}_k \in \mathbb{R}^{p_k \times n_k}$ is row full-rank with unit-norm rows. Moreover, for every block $C \in \mathcal{C}_k$ and every $i \in \rho_k(C)$, the row $[\mathbf{Q}_k]_{i,:}$ is orthogonal to every row of $[\mathbf{Q}_k]_{\bigcup_{C' \in anc_k^{\mathrm{blk}}(C)} \rho_k(C'),:}$.

Although this decomposition is generally not unique, any such factorization is sufficient for our analysis. Intuitively, $\mathbf{R}_k$ records block-ancestor mixing, while $\mathbf{Q}_k$ isolates block-associated row directions.

Proposition 3. *For every block $C \in \mathcal{C}_k$ and every $i \in \rho_k(C)$, we have $[\mathbf{H}_k]_{i,:} \in \mathrm{row}([\mathbf{Q}_k]_{\bigcup_{C' \in Anc_k^{\mathrm{blk}}(C)} \rho_k(C'),:})$.*

Thus, the rows of $\mathbf{H}_k$ are organized according to the block ancestral relation in the $\mathbf{Q}_k$ coordinates, which will allow us to interpret observable precision perturbations in terms of block-ancestor subspaces.

3.3 Low-rank perturbation of the precision matrix

Client-level recovery is driven not by a single precision matrix, but by how the precision matrix changes across intervention environments. We therefore study the perturbation $\mathbf{\Theta}_k^{(j)} - \mathbf{\Theta}_k^{(0)}$. The next two lemmas show that such perturbations are low-rank and that their row spaces are confined to block-ancestor subspaces.

Lemma 1 (Rank-2 perturbation)**.** *For each client k and each intervention j targeting node t_j (which is unknown to the learner), we have*

$$\mathrm{row}(\mathbf{\Theta}_k^{(j)} - \mathbf{\Theta}_k^{(0)}) \subseteq \left\langle \mathbf{H}_k^T (\mathbf{I} - \mathbf{A}_k)^T \mathbf{\Sigma}_k^{-1} \mathbf{e}_{\rho_k(t_j)},\ \mathbf{H}_k^T \Big((\mathbf{I} - \mathbf{A}_k)^T \mathbf{\Sigma}_k^{-1} \mathbf{u}_k^{(j)} + \mathbf{d}_k^{(j)} \Big) \right\rangle, \quad (8)$$

where $\mathbf{u}_k^{(j)}$ and $\mathbf{d}_k^{(j)}$ are defined as in Assumption 2. In particular, $\mathrm{rank}(\mathbf{\Theta}_k^{(j)} - \mathbf{\Theta}_k^{(0)}) \leq 2$.

Thus, a single-node soft intervention affects only a two-dimensional row space, despite producing a $p_k \times p_k$ perturbation matrix. To use this fact for recovery, we must further locate this row space relative to the client-specific block structure.

Lemma 2 (Row-space characterization via $\mathbf{Q}_k$)**.** *For each client k and intervention j, suppose that the intervention target node t_j belongs to block $C_{k,a} \in \mathcal{C}_k$. Then*

$$\mathrm{row}(\mathbf{\Theta}_k^{(j)} - \mathbf{\Theta}_k^{(0)}) \subseteq \left\langle [\mathbf{Q}_k]_{\rho_k(i),:} \;\middle|\; i \in C_{k,b},\ C_{k,b} \in Anc_k^{\mathrm{blk}}(C_{k,a}) \right\rangle.$$

Lemma 2 shows that if the intervention target lies in block $C_{k,a}$, then the resulting perturbation only involves row directions associated with $C_{k,a}$ and its block ancestors. In this sense, soft interventions reveal ancestor-closed portions of the client-specific block structure. For example, in Figure 1a, if intervention j for client $k = 1$ targets node $t_j = 2$, then $\mathrm{row}(\mathbf{\Theta}_k^{(j)} - \mathbf{\Theta}_k^{(0)}) \subseteq \langle [\mathbf{Q}_k]_{i,:} \mid i = \rho_k(1), \rho_k(2), \rho_k(7) \rangle$. Combined with the orthogonality relations in Definition 4, this implies that after projecting out the directions of recovered ancestor blocks, the remaining perturbation isolates the next unrecovered block.

3.4 Client-level recovery algorithms and identifiability result

The preceding lemmas yield the population recovery rule behind Algorithms 1 and 2. For a label set $C \subseteq \mathcal{J}_k$ and a row subspace W, define $\mathcal{P}_k(C; W) \triangleq \Pi_{W^\perp}\Big(\mathrm{row}\{\Theta_k^{(j)} - \Theta_k^{(0)} : j \in C\}\Big)$. Once the row spaces of ancestor blocks are included in W, the intervention labels of the next block are exactly the smallest set C satisfying $\dim \mathcal{P}_k(C; W) = |C|$. Algorithm 2 iteratively applies this rule

Algorithm 1: ID-Block-Ancestor

Input: $\hat{C} \subseteq \mathcal{J}_k$, $\{\Theta_k^{(j)}\}_{j\in\{0\}\cup\hat{C}}$, $\mathcal{I}$, $\{\hat{\mathbf{q}}_{k,i}\}_{i\in\mathcal{I}}$
Output: $\{\hat{\mathbf{q}}_{k,i}\}_{i\in\hat{C}}$, $\mathcal{A}$

$\mathcal{A} \leftarrow \mathcal{I}$
for each block $B \in \mathcal{I}$ **do**
  $W_{-B} \leftarrow \langle \hat{\mathbf{q}}_{k,i} : i \in \mathcal{I}\backslash B\rangle$
  $V_{-B} \leftarrow \mathcal{P}_k(\hat{C}; W_{-B})$
  if $\dim(V_{-B}) = |\hat{C}|$ **then**
    $\mathcal{A} \leftarrow \mathcal{A}\backslash B$
  end if
end for
$W \leftarrow \langle \hat{\mathbf{q}}_{k,i} : i \in \mathcal{A}\rangle$; $V \leftarrow \mathcal{P}_k(\hat{C}; W)$
Construct normalized vectors $\{\hat{\mathbf{q}}_{k,i}\}_{i\in\hat{C}}$ such that $V = \langle \hat{\mathbf{q}}_{k,i} : i \in \hat{C}\rangle$
return $\{\hat{\mathbf{q}}_{k,i}\}_{i\in\hat{C}}, \mathcal{A}$

Algorithm 2: ID-Block-Causal-Order

Input: $\{\Theta_k^{(j)}\}_{j\in\{0\}\cup\mathcal{J}_k}, p_k$
Output: $\hat{\mathcal{C}}_k$, $\hat{\prec}_k^{\mathrm{blk}}$, $\hat{\mathbf{Q}}_k$

$\mathcal{I}_0 \leftarrow \emptyset$, $R_0 \leftarrow \mathcal{J}_k$, $\hat{\mathbf{Q}}_k \leftarrow \mathbf{0}_{p_k\times n_k}$, $t \leftarrow 1$
while $R_{t-1} \neq \emptyset$ **do**
  $r \leftarrow 1$, flag $\leftarrow$ True
  while flag **do**
    $W_t \leftarrow \langle \hat{\mathbf{q}}_{k,i} : i \in \mathcal{I}_{t-1}\rangle$
    $\mathcal{R} \leftarrow \{C \subseteq R_{t-1} : |C| = r\}$
    if there exists $\hat{C} \in \mathcal{R}$ such that $\dim(\mathcal{P}_k(\hat{C}; W_t)) = r$ **then**
      Pick one such $\hat{C}$, flag $\leftarrow$ False
    else
      $r \leftarrow r + 1$
    end if
  end while
  $(\{\hat{\mathbf{q}}_{k,i}\}_{i\in\hat{C}}, \mathcal{A}) \leftarrow$ ID-Block-Ancestor$(\hat{C}, \{\Theta_k^{(j)}\}_{j\in\{0\}\cup\hat{C}}, \mathcal{I}_{t-1}, \{\hat{\mathbf{q}}_{k,i}\}_{i\in\mathcal{I}_{t-1}})$
  Add $\hat{C} \hat{\prec}_k^{\mathrm{blk}} C'$ for any $\tilde{C} \hat{\preceq}_k^{\mathrm{blk}} C'$ with $\tilde{C} \in \mathcal{A}$.
  Update $\hat{\mathbf{Q}}_k \leftarrow [\tilde{\mathbf{Q}}_{k,\hat{C}}; \hat{\mathbf{Q}}_k]$, where $\tilde{\mathbf{Q}}_{k,\hat{C}}$ stacks $\{\hat{\mathbf{q}}_{k,i}\}_{i\in\hat{C}}$.
  $\mathcal{I}_t \leftarrow \{\hat{C}\} \cup \mathcal{I}_{t-1}$
  $R_t \leftarrow R_{t-1}\backslash\hat{C}$
  $t \leftarrow t + 1$
end while
$\hat{\mathcal{C}}_k \leftarrow \mathcal{I}_{t-1}$
return $\hat{\mathcal{C}}_k$, $\hat{\prec}_k^{\mathrm{blk}}$, $\hat{\mathbf{Q}}_k$

Algorithm 3: ID-Global-Causal-Order

Input: $\{\hat{\mathcal{C}}_k, \hat{\prec}_k^{\mathrm{blk}}\}_{k=1}^K$
Output: $\hat{\prec}$

Initialize $\tilde{\prec} \leftarrow \emptyset$
for $k = 1, 2, \ldots, K$ **do**
  Find all pairs of blocks $C_{k,i}, C_{k,j} \in \hat{\mathcal{C}}_k$ such that $|C_{k,i}| = |C_{k,j}| = 1$ and $C_{k,j} \in \hat{pa}_k^{\mathrm{blk}}(C_{k,i})$
  For each such pair $C_{k,i} = \{i\}$ and $C_{k,j} = \{j\}$, add $i \tilde{\prec} j$
end for
$\hat{\prec} \leftarrow$ the transitive closure of $\tilde{\prec}$
return $\hat{\prec}$

to recover blocks, while Algorithm 1 determines which previously recovered blocks are ancestors of the newly found block.

Combining Lemmas 1 and 2 with the orthogonality structure in Definition 4 yields the following client-level identifiability result.

Theorem 1 (Client-level recovery)**.** *Assume Assumptions 1 and 2. Let $(\hat{\mathcal{C}}_k, \hat{\prec}_k^{\mathrm{blk}}, \hat{\mathbf{Q}}_k)$ be the output of Algorithm 2 for client k. Then:*

(i) $(\hat{\mathcal{C}}_k, \hat{\prec}_k^{\mathrm{blk}})$ recovers $(\mathcal{C}_k, \prec_k^{\mathrm{blk}})$ up to relabeling under the bijection $\psi : j \mapsto t_j$.

(ii) For each true block $C \in \mathcal{C}_k$, let $\hat{C} = \psi^{-1}(C)$ denote the corresponding estimated block. Then $\langle \hat{\mathbf{q}}_{k,i} : i \in \hat{C}\rangle = \mathrm{row}([\mathbf{Q}_k]_{\rho_k(C),:})$.

Consequently, $\hat{\mathbf{Q}}_k = \mathbf{S}_k\mathbf{P}_{\sigma,k}\mathbf{Q}_k$, where $\mathbf{S}_k$ is an invertible block-diagonal norm-preserving matrix and $\mathbf{P}_{\sigma,k}$ is a row permutation matrix preserving the block ancestral order.

Theorem 1 shows that a single fragmented client cannot generally recover a node-wise latent DAG, but can exactly recover the client-specific block structure, block-level ancestral relation, and row subspaces of $\mathbf{Q}_k$. The relabeling by $\psi : j \mapsto t_j$ only changes the names of recovered latent targets. The additional ambiguity, captured by $\mathbf{S}_k$ and $\mathbf{P}_{\sigma,k}$, is intrinsic to the non-uniqueness of the block RQ decomposition: it only changes the within-block basis and order-preserving block labels, while preserving the identifiable block-associated row subspaces and causal structure.

Applying the recovered rows of $\hat{\mathbf{Q}}_k$ in Theorem 1 to X_k therefore yields the corresponding client-specific latent variables, up to the intrinsic ambiguity induced by block-level ancestors, as follows.

Corollary 1 (Recovery of Z_k up to block-level ancestral ambiguity)**.** *Assume Assumptions 1 and 2 hold. Let $\hat{\mathbf{Q}}_k$ be the output of Algorithm 2, and define*

$$\hat{Z}_k = \hat{\mathbf{Q}}_k X_k.$$

For each true block $C \in \mathcal{C}_k$, let $\hat{C} = \psi^{-1}(C)$ denote the corresponding estimated block. Then, for every $i \in \hat{C}$, the estimated latent variable $\hat{Z}_{k,(i)}$ lies in the span of the true latent variables indexed by the block ancestors of C, namely

$$\hat{Z}_{k,(i)} \in \mathrm{span}\{Z_{(j)} : j \in C', C' \in Anc_k^{\mathrm{blk}}(C)\}.$$

Thus, up to the relabeling induced by ψ, each estimated latent variable is identified up to block-level ancestor mixing.

Corollary 1 gives the latent-level interpretation of the client-specific recovery result. For a single fragmented client, exact node-wise latent recovery is generally impossible because marginalizing unobserved latent variables induces bidirected dependencies and merges nodes into block-structured objects. Nevertheless, the recovered intervention-induced subspaces identify the latent variables at the finest level permitted by the model, namely up to block-level ancestor mixing.

This completes the client-level part of the identifiability picture: from a single client, one can recover its block structure, the associated block-level ancestral relation, and the corresponding latent subspaces up to block-level ancestral ambiguity. In the next section, we show how such client-specific fragments can be assembled to recover the global latent causal order.

4 Global Ancestral Recovery Across Clients

Section 3 shows that each client recovers a correct block-structured fragment of the global latent system. We now return to the question raised in Section 1: **can we recover the global latent causal order by assembling fragmented causal information from multiple clients?** This is not possible from a single client in general: marginalization may merge several global latent variables into one client-specific block. The key observation is that different clients may induce different blockings. Thus, two latent variables that are entangled in one client's view may be separated and ordered in another client's recovered block structure.

We first formalize the cross-client visibility needed for global assembly.

Assumption 3 (Cross-client coverage for global assembly)**.** *If $j \to i$ is a Hasse edge of the global latent DAG $\mathcal{G}$, namely $j \prec i$ and there exists no node m such that $j \prec m \prec i$, there exists a client $k \in [K]$ such that (i) $i, j \in O_k$; (ii) $\{i\}, \{j\} \in \mathcal{C}_k$; (iii) $\{j\} \in pa_k^{\mathrm{blk}}(\{i\})$.*

Under Assumption 3, every global Hasse edge is witnessed by at least one client as a parent relation between singleton blocks. Thus Algorithm 3 assembles the global order by collecting all recovered singleton-block parent relations $C_{k,j} \in \hat{pa}_k^{\mathrm{blk}}(C_{k,i})$ with $C_{k,i} = \{i\}$ and $C_{k,j} = \{j\}$, and then taking their transitive closure. Theorem 2 shows that this procedure recovers the global latent ancestral relation, up to the intervention-target relabeling $\psi : j \mapsto t_j$.

Theorem 2 (Correctness of global ancestral recovery)**.** *Assume Assumptions 1–3 hold. Let $\{(\hat{\mathcal{C}}_k, \hat{\prec}_k^{\mathrm{blk}})\}_{k=1}^K$ be the outputs of Algorithm 2 and the input to Algorithm 3. Let $\hat{\prec}$ be the output of Algorithm 3. Then $(\mathcal{V}, \hat{\prec})$ recovers $(\mathcal{V}, \prec)$ up to relabeling under the bijection $\psi : j \mapsto t_j$.*

Remark 1. *Theorems 1 and 2, together with Corollary 1, yield a two-level identifiability picture: a single client recovers only a block-structured fragment and the corresponding latent variables up to block-level ancestor mixing, whereas multiple such fragments can be assembled to recover the global node-level ancestral relation. This local-to-global gap is intrinsic to the fragmented multi-client setting.*

5 Finite-sample Method and Experiments

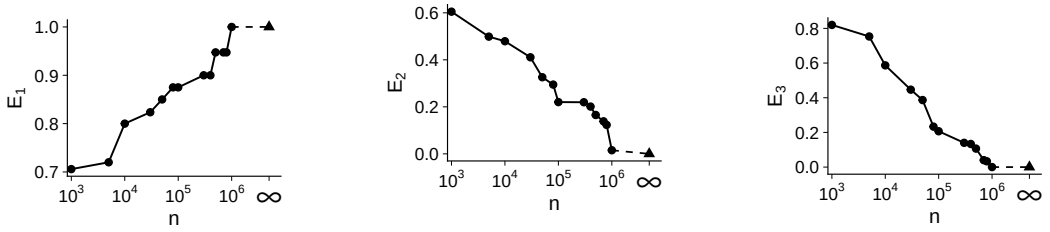
The preceding sections establish population-level identifiability from exact precision matrices. We now evaluate a finite-sample implementation of the proposed recovery procedure on synthetic fragmented multi-client systems, and include the population case $n = \infty$ to verify the theoretical recovery guarantees.

Finite-sample implementation. In finite samples, each population precision matrix $\mathbf{\Theta}_k^{(j)}$ is replaced by an estimator $\hat{\mathbf{\Theta}}_k^{(j)}$. Since exact row-space dimensions are unstable under estimation noise, we use a thresholded singular-value criterion: for a matrix M, its effective row-space dimension is taken as

the smallest r such that $\frac{\sum_{i=1}^{r}\sigma_i(M)}{\sum_i \sigma_i(M)} > 1-\tau$, where $\sigma_i(M)$ is the i-th largest singular value and τ is a tuning parameter. This gives finite-sample versions of Algorithms 1 and 2; details are deferred to Appendix N.

Setup. We consider three different global causal network structures $\mathbf{A}$ together with different client-specific missing patterns $\{O_k\}_{k=1}^K$, representing different network scales and sparsity levels. We consider two noise distributions, Gaussian and Laplace. For each graph setting and noise setting, the full specifications are deferred to Appendix N. We report results for sample sizes $n \in \{10^3, \dots, 10^6\}$ together with the population case $n = \infty$.

Evaluation metrics. We use three metrics. For ease of exposition, we first align the row ordering of $\hat{\mathbf{Q}}_k$ with that of $\mathbf{Q}_k$ according to the permutation induced by $\mathbf{P}_{\sigma,k}$; this does not affect the evaluation itself. The first metric, E_1, is the F1-score between the true global partial order $\prec$ and the estimate $\hat{\prec}$. The second metric, E_2, measures the average block-level row-subspace error between the true and estimated rows of $\mathbf{Q}_k$. The third metric, E_3, measures support leakage in the estimated latent variables $\hat{Z}_k = \hat{\mathbf{Q}}_k X_k$, namely the fraction of nonzero coefficients appearing outside the block-level ancestral support allowed by Corollary 1. All three metrics take values in $[0, 1]$, where larger E_1 indicates better performance, while smaller E_2 and E_3 indicate more accurate recovery. Complete formulas are given in Appendix N.



(a) Global ancestral F1.

(b) Subspace error of $\{\hat{\mathbf{Q}}_k\}_{k=1}^K$.

(c) Support error of $\{\hat{Z}_k\}_{k=1}^K$.

Figure 2: Finite-sample performance under one graph setting A and Gaussian noise. As n increases, global ancestral recovery improves while subspace and support errors decrease. The rightmost marker at ∞ corresponds to the population-level precision-matrix input.

Results. For the six scenarios induced by three graph settings and two noise settings, Figure 2 and Figures A.1-A.5 in Appendix N plot E_1, E_2, E_3 against the sample size n. Across all scenarios, performance improves as n increases: the estimated global ancestor relation becomes more accurate, the client-specific subspace error decreases, and the support leakage becomes smaller. In the population case $n = \infty$, where the estimated precision matrices are replaced by the corresponding population precision matrices, we obtain $E_1 = 1$ and $E_2 = E_3 = 0$. This exactly matches Theorems 1 and 2 together with Corollary 1, and confirms that the finite-sample implementation converges to the population recovery guaranteed by the theory. The same qualitative behavior is observed across all three graph settings and under both Gaussian and Laplace noise, indicating that the method is robust to variations in graph scale, sparsity, and noise distribution. Additional sensitivity and scalability analyses varying p, K, and the client coverage probability q, including random missingness patterns that may violate Assumption 3, are provided in Appendix N.6. A semiconductor-manufacturing case study is reported in Appendix N.7.

6 Conclusion

We study causal representation learning in a fragmented multi-client setting, where clients share a global latent causal system but each accesses and intervenes on only a subset of latent variables. We show that latent marginalization changes the client-level identifiable object from a node-wise latent DAG to a block structure, a block-level ancestral relation, and latent subspaces up to block-level ancestor mixing. We further show that these client-specific fragments can be assembled across clients to recover the global node-level latent causal order. Future directions include finite-sample guarantees, extensions to nonlinear models and more general observation mappings, and stronger identifiability under hard interventions, potentially including exact recovery of $\mathbf{A}$ and $\mathbf{H}$.

AI USE STATEMENT

In this work, we used generative AI tools to assist with writing refinement and language polishing, as well as literature retrieval and discovery for identifying related research. We have not used generative AI tools for generating research ideas, developing the theoretical methodology, designing algorithms, conducting experiments, analyzing experimental results, generating code, or producing mathematical proofs. Other uses of generative AI tools beyond the above-mentioned purposes are not applicable to this work. We have carefully reviewed all AI-assisted content and verified its correctness, consistency, and originality through manual inspection. In particular, all theoretical results, proofs, algorithmic developments, and experimental analyses were independently developed and validated by the authors. We take full responsibility for the final content of this work, including all text, claims, and artifacts produced with the aid of generative AI tools.

ETHICS STATEMENT

This work does not involve human subjects, personal data, or real-world deployment of potentially harmful systems. The experiments are conducted on synthetic datasets generated under controlled simulation settings. The proposed method is developed for fundamental research in causal representation learning, and we are not aware of any ethical concerns arising from this work.

REPRODUCIBILITY STATEMENT

We have made efforts to ensure the reproducibility of this work. The proposed algorithms and experimental implementation are provided in the appendix and abstract through an anonymous code repository. The main paper describes the problem formulation, assumptions, model structure, and algorithmic framework, while complete proofs of all theoretical results are provided in the appendix. Detailed experimental settings, including simulation configurations, evaluation metrics, and additional results, are also included in the appendix. These materials together provide the necessary information to reproduce the theoretical analysis and experimental results reported in this paper.

REFERENCES

Kartik Ahuja, Divyat Mahajan, Yixin Wang, and Yoshua Bengio. Interventional causal representation learning. In *International conference on machine learning*, pp. 372–407. PMLR, 2023.

Simon Bing, Urmi Ninad, Jonas Wahl, and Jakob Runge. Identifying linearly-mixed causal representations from multi-node interventions. In *Causal Learning and Reasoning*, pp. 843–867. PMLR, 2024.

Simon Buchholz, Goutham Rajendran, Elan Rosenfeld, Bryon Aragam, Bernhard Schölkopf, and Pradeep Ravikumar. Learning linear causal representations from interventions under general nonlinear mixing. *Advances in Neural Information Processing Systems*, 36:45419–45462, 2023.

David Danks, Clark Glymour, and Robert Tillman. Integrating locally learned causal structures with overlapping variables. In *Advances in Neural Information Processing Systems*, volume 21, 2008.

Anish Dhir and Ciarán M Lee. Integrating overlapping datasets using bivariate causal discovery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 3781–3790, 2020.

Mathias Drton, Rina Foygel, and Seth Sullivant. Global identifiability of linear structural equation models. *The Annals of Statistics*, 39(2):865–886, 2011. doi: 10.1214/10-AOS859.

Rina Foygel, Jan Draisma, and Mathias Drton. Half-trek criterion for generic identifiability of linear structural equation models. *The Annals of Statistics*, 40(3):1682–1713, 2012. doi: 10.1214/12-AOS1012.

Thomas Nall Eden Greville. Note on the generalized inverse of a matrix product. *Siam Review*, 8(4): 518–521, 1966.

Jikai Jin and Vasilis Syrgkanis. Learning linear causal representations from general environments: Identifiability and intrinsic ambiguity. *Advances in Neural Information Processing Systems*, 37: 63466–63509, 2024.

Yuhang Liu, Zhen Zhang, Dong Gong, Mingming Gong, Biwei Huang, Anton van den Hengel, Kun Zhang, and Javen Qinfeng Shi. Identifiable latent polynomial causal models through the lens of change. In *The Twelfth International Conference on Learning Representations*, 2024.

Yuhang Liu, Zhen Zhang, Dong Gong, Mingming Gong, Biwei Huang, Anton van den Hengel, Kun Zhang, and Javen Qinfeng Shi. Identifying weight-variant latent causal models. *Journal of Machine Learning Research*, 27(4):1–49, 2026.

Ignavier Ng, Shaoan Xie, Xinshuai Dong, Peter Spirtes, and Kun Zhang. Causal representation learning from general environments under nonparametric mixing. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025.

Thomas S. Richardson and Peter Spirtes. Ancestral graph markov models. *The Annals of Statistics*, 30(4):962–1030, 2002. doi: 10.1214/aos/1031689015.

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, 2021.

Chandler Squires, Anna Seigal, Salil S Bhate, and Caroline Uhler. Linear causal disentanglement via interventions. In *International conference on machine learning*, pp. 32540–32560. PMLR, 2023.

Nils Sturma, Chandler Squires, Mathias Drton, and Caroline Uhler. Unpaired multi-domain causal representation learning. *Advances in Neural Information Processing Systems*, 36:34465–34492, 2023.

Jin Tian and Judea Pearl. A general identification condition for causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pp. 567–573. AAAI Press, 2002.

Sofia Triantafillou and Ioannis Tsamardinos. Constraint-based causal discovery from multiple interventions over overlapping variable sets. *Journal of Machine Learning Research*, 16(66): 2147–2205, 2015.

Sofia Triantafillou, Ioannis Tsamardinos, and Ioannis G. Tollis. Learning causal structure from overlapping variable sets. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 860–867. PMLR, 2010.

Burak Varici, Emre Acartürk, Karthikeyan Shanmugam, and Ali Tajer. General identifiability and achievability for causal representation learning. In *International Conference on Artificial Intelligence and Statistics*, pp. 2314–2322. PMLR, 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 Blei, and Bernhard Schölkopf. Nonparametric identifiability of causal representations from unknown interventions. *Advances in Neural Information Processing Systems*, 36:48603–48638, 2023.

Dingling Yao, Danru Xu, Sébastien Lachapelle, Sara Magliacane, Perouz Taslakian, Georg Martius, Julius von Kügelgen, and Francesco Locatello. Multi-view causal representation learning with partial observability. In *12th International Conference on Learning Representations*, 2024.

Jiaqi Zhang, Kristjan Greenewald, Chandler Squires, Akash Srivastava, Karthikeyan Shanmugam, and Caroline Uhler. Identifiability guarantees for causal disentanglement from soft interventions. *Advances in Neural Information Processing Systems*, 36:50254–50292, 2023.

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

Kun Zhang, Shaoan Xie, Ignavier Ng, and Yujia Zheng. Causal representation learning from multiple distributions: A general setting. *arXiv preprint arXiv:2402.05052*, 2024.

Appendix Contents

A Related Work **12**

B Normalization of the Noise Covariance **13**

C Discussion of Assumptions **13**

D Discussion of Computational Complexity **15**

E Proof of Proposition 1 **16**

F Proof of Proposition 2 **18**

G Block RQ Decomposition **19**

H Proof of Lemma 1 **21**

I Proof of Lemma 2 **22**

J Proof of Theorem 1 **23**

K Proof of Corollary 1 **32**

L Proof of Theorem 2 **33**

M Auxiliary Lemmas **34**

N Additional finite-sample details **38**

A Related Work

Intervention-based CRL in a shared latent space. A central line of CRL studies identifiability from interventions or multiple environments under a shared latent space (Schölkopf et al., 2021; Squires et al., 2023; Buchholz et al., 2023; Varici et al., 2024; 2025; Zhang et al., 2024; Liu et al., 2026; Zhang et al., 2023; Liu et al., 2024; Ahuja et al., 2023; Ng et al., 2025; Bing et al., 2024; von Kügelgen et al., 2023). These works differ in intervention regimes, mixing assumptions, and identifiability targets, but they typically assume that all environments are defined over the same latent variables. Consequently, the target remains a node-level latent causal structure, or an ambiguity class thereof, in a single latent space. In contrast, we consider fragmented clients that access and intervene on only subsets of a common global latent causal system. Marginalizing the unused latent variables induces client-specific bidirected dependencies, making the local identifiable object block-structured rather than a node-wise latent DAG.

CRL under multi-view or multiple domains. A related line studies CRL with multiple views or multiple domains. Yao et al. (2024) consider multi-view CRL where each view depends on a subset of latent variables and identify shared latent information across views, while Sturma et al. (2023) recover a shared causal representation from unpaired multi-domain marginal distributions. These works are related to ours through partial observability or multiple clients, but they do not address how latent marginalization changes each client's local causal object. In our setting, the unused latent variables induce bidirected dependencies, so the client-level target becomes a block-structured fragment that must be assembled to recover the global latent causal order.

Latent confounding and mixed graphs. Our client-specific induced graphs are related to causal models with latent confounders, where ancestral graphs and acyclic directed mixed graphs (ADMGs) use bidirected edges to encode dependencies induced by marginalization or unobserved common causes (Richardson & Spirtes, 2002; Zhang, 2008; Tian & Pearl, 2002). In linear structural equation models, mixed graphs have also been used to study global or generic parameter identifiability (Drton et al., 2011; Foygel et al., 2012). These works provide important graphical languages for latent confounding, but they mainly concern observed-variable causal structure or parameter identifiability. In contrast, our mixed graphs arise over client-specific latent variables accessed only through high-dimensional observations, and our target is block-level latent identifiability from interventional precision perturbations rather than observed-variable ADMG discovery.

Causal discovery from overlapping variable sets. Another related line studies causal discovery from multiple datasets with overlapping variable sets, including methods that integrate locally learned structures or constraints across observational and interventional datasets (Danks et al., 2008; Triantafillou et al., 2010; Triantafillou & Tsamardinos, 2015; Dhir & Lee, 2020). These works also assemble information from partial views of a system, but they typically operate on observed variables and target a graph or equivalence class over the union of measured variables. In contrast, our clients observe high-dimensional mixtures of latent variables, and missing latent variables induce bidirected dependencies and block-level ambiguity. Our task is therefore not to merge overlapping observed-variable graphs, but to recover a global latent causal order from block-structured latent fragments.

B Normalization of the Noise Covariance

In the main text, we work under the normalized parametrization $\mathbf{\Sigma} = \mathbf{I}$. In this appendix, we explain why this assumption is made without loss of generality.

Suppose there exists an alternative parameterization

$$\left\{\mathbf{A}', \mathbf{\Sigma}', \{\mathbf{H}'_k\}_{k\in[K]}, \{\tilde{\mathbf{A}}^{(j)'}, \tilde{\mathbf{\Sigma}}^{(j)'}\}_{j\in[p]}\right\}$$

with $\mathbf{\Sigma}' \neq \mathbf{I}$. Let

$$\mathbf{D} = (\mathbf{\Sigma}')^{-1/2},$$

and define the transformed parameters

$$\mathbf{A} = \mathbf{D}\mathbf{A}'\mathbf{D}^{-1}, \qquad \mathbf{\Sigma} = \mathbf{I}, \qquad \mathbf{H}_k = [\mathbf{D}]_{O_k,O_k}\mathbf{H}'_k,$$

together with

$$\tilde{\mathbf{A}}^{(j)} = \mathbf{D}\tilde{\mathbf{A}}^{(j)'}\mathbf{D}^{-1}, \qquad \tilde{\mathbf{\Sigma}}^{(j)} = \mathbf{D}\tilde{\mathbf{\Sigma}}^{(j)'}\mathbf{D}.$$

Under this transformation, the latent variables are rescaled but the induced second-order information of the observed variables remains unchanged. In particular, the collection of observed distributions

$$\{X_k^{(j)}\}_{k\in[K],\, j\in\mathcal{J}_k\cup\{0\}}$$

generated by the transformed parameterization is identical to that generated by the original one. Therefore, the two parameterizations are observationally equivalent.

Hence, whenever a model with general noise covariance $\mathbf{\Sigma}'$ is observationally feasible, there exists an equivalent reparameterization with identity noise covariance. It is therefore sufficient to analyze the normalized case $\mathbf{\Sigma} = \mathbf{I}$ throughout the paper.

C Discussion of Assumptions

Discussion of Assumption 1. Assumption 1 combines a standard within-client coverage requirement with a cross-client label alignment condition. The coverage requirement states that, for each client k, the intervention set $\mathcal{J}_k$ is in bijection with the client-specific node set $\mathcal{V}_k$, so that every node is targeted by exactly one intervention environment. Such coverage assumptions are standard in intervention-based latent causal discovery and CRL, where sufficiently rich interventions are generally needed to identify causal orders or ancestral relations in the worst case (Varici et al., 2025; Buchholz et al., 2023; Squires et al., 2023). This requirement is particularly natural in our

fragmented setting, since marginalizing the latent variables unavailable to a client induces bidirected dependencies, and even with full client-level intervention coverage, a single client can generally recover only a block-structured causal object.

The cross-client label alignment condition does not strengthen the client-level recovery result, but it is essential for assembling the recovered fragments into a common global causal order. Without this condition, the ambiguity is structural rather than merely an additional global permutation. To see this, consider $K = 2$ and $p = 3$, with $O_1 = \{1, 2\}$ and $O_2 = \{2, 3\}$. Suppose that each client recovers an unlabeled two-node relation of the form $a_1 \to a_2$. These same client-level fragments are compatible with both the global chain

$$1 \to 2 \to 3$$

and the global fork

$$1 \leftarrow 2 \to 3.$$

Indeed, the two models differ only in how the intervention labels of the first client are matched to global nodes 1 and 2, while the second client's local relation can remain unchanged. Since the client-specific mixing matrices $\mathbf{G}_1$ and $\mathbf{G}_2$ are arbitrary and unknown, the observed mixtures provide no additional common anchor for resolving this correspondence. Moreover, the chain and the fork are not related by any single global permutation: in the chain, all three nodes are totally ordered, whereas in the fork, the two children are incomparable. Thus, allowing an additional global permutation ambiguity does not restore identifiability. Cross-client label alignment is therefore intrinsically required for global assembly, independently of the particular recovery algorithm. Related CRL formulations likewise impose correspondence conditions across environments, such as the coupled-environment condition in Varici et al. (2025). Under this condition, the global map $\psi : j \mapsto t_j$ consistently associates each intervention label with its target node across all clients.

Discussion of Assumption 2. Assumption 2 is a standard genericity condition in intervention-based causal discovery and related CRL work, used to rule out pathological intervention choices that create exceptional cancellations in the structural signal (Varici et al., 2025; Squires et al., 2023). The first two parts correspond to non-degeneracy of the directed and bidirected components, respectively. In view of Lemma A.4, the supports of $\mathbf{d}_k^{(j)}$ and $\mathbf{u}_k^{(j)}$ are structurally constrained by the underlying graph, and these two conditions require that the entries which can be nonzero are, generically, indeed nonzero. Intuitively, this means that the intervention sufficiently excites all directed and bidirected directions that are structurally reachable, rather than being canceled by special parameter choices; such cancellations can typically be regarded as measure-zero pathologies. The third part imposes a non-degeneracy condition within each block. Combining the first two parts, for any $i \in C_{k,\ell}$, the possible elements of $\mathcal{N}_i$ lie in $\rho_k\big((sp_k(Anc_k(i)) \cup Anc_k(i)) \cap C_{k,\ell}\big)$. Hence, for any proper subset $\mathcal{B} \subsetneq C_{k,\ell}$, the union $\bigcup_{i\in\mathcal{B}} \mathcal{N}_i$ typically extends beyond $\mathcal{B}$ itself. Indeed, by the definition of the block structure, any two nodes in the same block are connected through a combination of ancestral relations and bidirected connectivity, so the set of potentially nonzero coordinates necessarily reaches nodes in $C_{k,\ell} \setminus \mathcal{B}$ as well. Therefore, as long as one avoids further pathological cancellations on these potentially active coordinates, one expects $\big|\bigcup_{i\in\mathcal{B}} \mathcal{N}_i\big| > |\mathcal{B}|$. In this sense, the third part likewise excludes degenerate interventions that collapse the within-block information. Overall, Assumption 2 only requires that interventions generically reveal the structural signal allowed by the graph, rather than imposing any strong additional structural restriction.

Discussion of Assumption 3. Assumption 3 is a cross-client coverage condition for exact global recovery. Some such condition is unavoidable: if $\bigcup_{k=1}^{K} O_k \subsetneq [p]$, then some global latent nodes are never accessed and their ancestral relations are generally unidentifiable, whereas if $O_k = [p]$ for every client, the problem reduces to the non-fragmented single-client setting. Assumption 3 adopts a relatively weak intermediate requirement. It neither requires any client to access the full latent system nor requires every global ancestral relation to be directly witnessed. Instead, it only requires each global Hasse edge to appear in some client as a parent relation between singleton blocks. Since the complete ancestral order is the transitive closure of its Hasse edges, this is sufficient, and in general necessary, for exact global assembly.

This assumption is not needed for client-level recovery: Theorem 1 and Corollary 1 remain valid when it fails. Moreover, the proof of Theorem 2 still guarantees $\hat{\prec} \subseteq \prec$, so Algorithm 3 may omit relations associated with unwitnessed Hasse edges but does not introduce incorrect causal relations.

The condition is also mild under random fragmented access. Suppose that every client independently includes each latent node in O_k with probability q, let $\mathcal{H}$ be the set of global Hasse edges, and define $a_i = |\operatorname{Anc}(i)|$. For a Hasse edge $j \to i$, observing all nodes in $\operatorname{Anc}(i)$ is sufficient for one client to witness the edge and occurs with probability q^{a_i}. Therefore, if $\max_i a_i \le d$, a union bound gives

$$\Pr(\text{Assumption 3 holds} \mid \mathcal{G}) \ge 1 - |\mathcal{H}| e^{-Kq^d}.$$

Hence, $K \ge q^{-d} \log(|\mathcal{H}|/\delta)$ is sufficient for the condition to hold with probability at least $1-\delta$. Since $|\mathcal{H}| \le p(p-1)/2$, this yields $K = O(\log p)$ for fixed q, d, and δ. Although conservative, this bound shows that Assumption 3 is primarily a coverage requirement induced by fragmented observations rather than a restrictive causal assumption.

D Discussion of Computational Complexity

We analyze the computational complexity of the proposed method and compare it with the closely related single-client procedure of Squires et al. (2023). Let $p_k = |O_k|$ be the client-level latent dimension, let n_k be the observation dimension of client k, and define

$$b_k \triangleq \max_{C \in \mathcal{C}_k} |C|$$

as the maximum client-specific block size.

As a reference point, the algorithm of Squires et al. (2023) has worst-case complexity $O(n_k^3 p_k^2)$ under our notation. Its main cost comes from repeatedly projecting and decomposing $n_k \times n_k$ precision perturbation matrices while recovering p_k individual latent variables. Our client-level procedure uses the same type of projection and singular-value decomposition operations. The additional computation arises because latent marginalization may merge several intervention targets into an unknown block, so Algorithm 2 must search over candidate subsets rather than individual intervention labels.

For a candidate subset of size r, Algorithm 2 stacks r precision perturbation matrices of size $n_k \times n_k$ and performs the corresponding projection and singular-value decomposition. The cost of one candidate check is therefore $O(rn_k^3)$. Since the search considers candidate subsets up to size b_k and performs at most p_k block-recovery iterations, the subset-search cost is bounded by

$$O\left(p_k n_k^3 \sum_{r=1}^{b_k} r \binom{p_k}{r}\right).$$

In addition, Algorithm 1 compares each newly recovered block with previously recovered blocks. Its total cost is bounded by $O(b_k p_k^2 n_k^3)$. Combining the two parts gives the client-level worst-case complexity

$$O\left(p_k n_k^3 \sum_{r=1}^{b_k} r \binom{p_k}{r} + b_k p_k^2 n_k^3\right).$$

This expression also clarifies the additional cost relative to existing single-client CRL methods. When $b_k = 1$, all client-specific blocks are singletons and no combinatorial block ambiguity is present. In this case, the above complexity reduces to

$$O(n_k^3 p_k^2),$$

which matches the order of the single-client algorithm of Squires et al. (2023). Thus, our method does not introduce a higher-order computational cost when marginalization does not create non-singleton blocks. The additional subset search appears only when it is needed to resolve the block structure induced by fragmented observations.

More generally, if the maximum block size is bounded by a constant $b_k \le b$, then

$$\sum_{r=1}^{b} r \binom{p_k}{r} = O(p_k^b),$$

and the client-level complexity becomes

$$O(n_k^3 p_k^{b+1}).$$

Therefore, the method remains polynomial in p_k for bounded block size. Relative to the $O(n_k^3 p_k^2)$ single-client baseline, the additional worst-case factor is $O(p_k^{b-1})$, reflecting the search required to identify blocks of size up to b. The cost can become exponential only in the unrestricted worst case where b_k grows proportionally with p_k.

For the complete multi-client procedure, define

$$p_{\max} \triangleq \max_{k\in[K]} p_k, \qquad n_{\max} \triangleq \max_{k\in[K]} n_k, \qquad b \triangleq \max_{k\in[K]} b_k.$$

The total client-level recovery cost is bounded by

$$O\left(Kn_{\max}^3\left[p_{\max}\sum_{r=1}^{b} r\binom{p_{\max}}{r} + bp_{\max}^2\right]\right),$$

which reduces to $O(Kn_{\max}^3 p_{\max}^{b+1})$ for bounded b. Algorithm 3 only collects the recovered singleton-block relations and computes their transitive closure, adding at most $O(\sum_{k=1}^{K} p_k^2 + p^3)$ operations. Hence, the dependence on the number of clients is at most linear, and the client-level dimension and maximum block size are the primary determinants of computational cost. Moreover, the client-level computations are independent across clients and can be parallelized.

This analysis is conservative because the number of unrecovered labels decreases after every iteration, and the subset search terminates as soon as a valid candidate is found. The runtime results in Appendix N.6 are consistent with the analysis: runtime grows more strongly with the latent dimension than with the number of clients, while its dependence on the coverage probability is non-monotone because increasing coverage enlarges p_k and n_k but generally reduces the induced block sizes.

E Proof of Proposition 1

Proposition A.1 (Proposition 1 in the main text)**.** *For Z generated by Eq. equation 1, let $Z_k = (Z_{(i)})_{i\in O_k}$. Then*

$$Z_k = \mathbf{A}_k Z_k + E_k, \tag{9}$$

where

$$\mathbf{A}_k = [\mathbf{A}]_{O_k,O_k} + [\mathbf{A}]_{O_k,M_k}(\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-1}[\mathbf{A}]_{M_k,O_k}, \tag{10}$$

$$E_k = E_{O_k} + [\mathbf{A}]_{O_k,M_k}(\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-1}E_{M_k}. \tag{11}$$

Moreover, E_k has mean zero and covariance

$$\mathbf{\Sigma}_k = [\mathbf{\Sigma}]_{O_k,O_k} + [\mathbf{A}]_{O_k,M_k}(\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-1}[\mathbf{\Sigma}]_{M_k,M_k}(\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-T}[\mathbf{A}]_{O_k,M_k}^T. \tag{12}$$

Furthermore $\prec_k$, which is induced by $\mathbf{A}_k$, is exactly the restriction of the $\prec$, which is induced by $\mathbf{A}$, to $\mathcal{V}_k = O_k$.

Proof. For client k, partition the global latent vector and exogenous noise as

$$Z = \begin{pmatrix} Z_{O_k} \\ Z_{M_k} \end{pmatrix}, \qquad E = \begin{pmatrix} E_{O_k} \\ E_{M_k} \end{pmatrix}.$$

Then the global latent SEM in Eq. (1) can be written in block form as

$$\begin{pmatrix} Z_{O_k} \\ Z_{M_k} \end{pmatrix} = \begin{pmatrix} [\mathbf{A}]_{O_k,O_k} & [\mathbf{A}]_{O_k,M_k} \\ [\mathbf{A}]_{M_k,O_k} & [\mathbf{A}]_{M_k,M_k} \end{pmatrix}\begin{pmatrix} Z_{O_k} \\ Z_{M_k} \end{pmatrix} + \begin{pmatrix} E_{O_k} \\ E_{M_k} \end{pmatrix}.$$

Equivalently,

$$Z_{O_k} = [\mathbf{A}]_{O_k,O_k} Z_{O_k} + [\mathbf{A}]_{O_k,M_k} Z_{M_k} + E_{O_k}, \tag{13}$$

$$Z_{M_k} = [\mathbf{A}]_{M_k,O_k} Z_{O_k} + [\mathbf{A}]_{M_k,M_k} Z_{M_k} + E_{M_k}. \tag{14}$$

Since $\mathbf{A}$ is strictly upper triangular under a topological order, the principal submatrix $[\mathbf{A}]_{M_k,M_k}$ is also strictly upper triangular. Hence $\mathbf{I} - [\mathbf{A}]_{M_k,M_k}$ is invertible, and Eq. (14) gives

$$Z_{M_k} = (\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-1}[\mathbf{A}]_{M_k,O_k} Z_{O_k} + (\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-1}E_{M_k}. \tag{15}$$

Substituting Eq. (15) into Eq. (13), we obtain

$$Z_{O_k} = [\mathbf{A}]_{O_k,O_k} Z_{O_k} + [\mathbf{A}]_{O_k,M_k}(\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-1}[\mathbf{A}]_{M_k,O_k} Z_{O_k} + [\mathbf{A}]_{O_k,M_k}(\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-1} E_{M_k} + E_{O_k}.$$

Therefore,

$$Z_k = \mathbf{A}_k Z_k + E_k, \qquad Z_k = Z_{O_k},$$

where

$$\mathbf{A}_k = [\mathbf{A}]_{O_k,O_k} + [\mathbf{A}]_{O_k,M_k}(\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-1}[\mathbf{A}]_{M_k,O_k},$$

and

$$E_k = E_{O_k} + [\mathbf{A}]_{O_k,M_k}(\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-1} E_{M_k}.$$

This proves the claimed induced structural equation.

It remains to compute the covariance of E_k. Let

$$\mathbf{B}_k \triangleq [\mathbf{A}]_{O_k,M_k}(\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-1}.$$

Then

$$E_k = E_{O_k} + \mathbf{B}_k E_{M_k}.$$

Since the global exogenous noise covariance is $\mathbf{\Sigma}$ and the coordinates of E are mutually uncorrelated, we have

$$Cov(E_{O_k}, E_{M_k}) = \mathbf{0}.$$

Hence

$$\begin{aligned}\mathbf{\Sigma}_k \triangleq Cov(E_k) &= Cov(E_{O_k} + \mathbf{B}_k E_{M_k}) \\ &= [\mathbf{\Sigma}]_{O_k,O_k} + \mathbf{B}_k [\mathbf{\Sigma}]_{M_k,M_k} \mathbf{B}_k^T.\end{aligned}$$

Substituting the definition of $\mathbf{B}_k$ yields

$$\mathbf{\Sigma}_k = [\mathbf{\Sigma}]_{O_k,O_k} + [\mathbf{A}]_{O_k,M_k}(\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-1}[\mathbf{\Sigma}]_{M_k,M_k}(\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-T}[\mathbf{A}]_{O_k,M_k}^T,$$

which is exactly Eq. (12).

Finally, we justify the statement on the ancestral relation $\prec_k$. By construction, the directed edge set of the induced mixed graph $\mathcal{G}_k$ is determined by the nonzero entries of $\mathbf{A}_k$. From the formula above, a directed edge in $\mathcal{G}_k$ arises either from a directed edge already present in $[\mathbf{A}]_{O_k,O_k}$, or from a directed path in the global DAG whose endpoints lie in O_k and whose internal nodes lie in M_k. Indeed, since $[\mathbf{A}]_{M_k,M_k}$ is strictly upper triangular, it is nilpotent, and thus

$$(\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-1} = \mathbf{I} + [\mathbf{A}]_{M_k,M_k} + [\mathbf{A}]_{M_k,M_k}^2 + \cdots,$$

where the sum is finite. Therefore,

$$[\mathbf{A}]_{O_k,M_k}(\mathbf{I} - [\mathbf{A}]_{M_k,M_k})^{-1}[\mathbf{A}]_{M_k,O_k}$$

collects exactly the contributions of all directed paths whose endpoints lie in O_k and whose internal nodes lie in M_k. It follows that every directed path in $\mathcal{G}_k$ corresponds to a directed path in the global latent DAG. Hence, for any $i, j \in O_k$,

$$i \prec_k j \quad \Longrightarrow \quad i \prec j.$$

Conversely, suppose $i, j \in O_k$ and $i \prec j$ in the global latent DAG. Then there exists a directed path from j to i in the global graph. Removing from this path all intermediate nodes that lie in O_k decomposes it into consecutive segments whose endpoints lie in O_k and whose internal nodes lie in M_k. Each such segment induces a directed edge in $\mathcal{G}_k$ through the effective coefficient matrix $\mathbf{A}_k$, possibly together with edges already present in $[\mathbf{A}]_{O_k,O_k}$. Chaining these induced directed edges yields a directed path from j to i in $\mathcal{G}_k$, and therefore

$$i \prec j, \quad i, j \in O_k \quad \Longrightarrow \quad i \prec_k j.$$

Combining the two directions, we conclude that $\prec_k$ is exactly the restriction of the global ancestral relation $\prec$ to $\mathcal{V}_k = O_k$. □

F PROOF OF PROPOSITION 2

Proposition A.2 (Proposition 2 in the main text)**.** *For each client k, let $C_{k,i}, C_{k,j} \in \mathcal{C}_k$ be two blocks.*

(i) If $C_{k,j} \notin anc_k^{\mathrm{blk}}(C_{k,i})$, then for $\forall u \in C_{k,j}$ and $v \in C_{k,i}$, we have $u \notin anc_k(v)$.

(ii) If $C_{k,j} \in pa_k^{\mathrm{blk}}(C_{k,i})$, then there $\exists u \in C_{k,j}$ and $v \in C_{k,i}$ such that $u \in pa_k(v)$.

Proof. We prove the two claims separately.

For the first claim, suppose $C_{k,j} \notin Anc_k^{\mathrm{blk}}(C_{k,i})$. We show that there do not exist $u \in C_{k,j}$ and $v \in C_{k,i}$ such that $u \in Anc_k(v)$. Indeed, if such a pair existed, then by the definition of the block-level ancestral relation, the existence of $u \prec_k v$ would imply

$$C_{k,j} \prec_k^{\mathrm{blk}} C_{k,i},$$

or equivalently $C_{k,j} \in Anc_k^{\mathrm{blk}}(C_{k,i})$, which contradicts the assumption. Therefore, for any $u \in C_{k,j}$ and $v \in C_{k,i}$, we must have $u \notin Anc_k(v)$.

We next prove the second claim. Suppose $C_{k,j} \in Pa_k^{\mathrm{blk}}(C_{k,i})$. By definition of the block parent relation, there exist $u \in C_{k,j}$ and $v \in C_{k,i}$ such that $u \in Anc_k(v)$, that is, there exists a directed path from u to v in the directed part of $\mathcal{G}_k$. Among all such directed paths from a node in $C_{k,j}$ to a node in $C_{k,i}$, choose one of shortest length, and write it as

$$u = w_0 \to w_1 \to \cdots \to w_m = v.$$

Consider the sequence of blocks containing these nodes:

$$C_k(w_0),\, C_k(w_1),\, \ldots,\, C_k(w_m),$$

where $C_k(w_\ell)$ denotes the block in $\mathcal{C}_k$ containing w_ℓ. Since $w_0 = u \in C_{k,j}$ and $w_m = v \in C_{k,i}$, the path starts in $C_{k,j}$ and ends in $C_{k,i}$. We claim that no intermediate node w_ℓ can belong to a block different from both $C_{k,j}$ and $C_{k,i}$. Otherwise, for some $1 \le \ell \le m-1$, there exists a block $C_{k,m'} \in \mathcal{C}_k$ such that

$$w_\ell \in C_{k,m'}, \qquad C_{k,m'} \neq C_{k,j}, \qquad C_{k,m'} \neq C_{k,i}.$$

Then the directed path above implies

$$C_{k,j} \prec_k^{\mathrm{blk}} C_{k,m'} \prec_k^{\mathrm{blk}} C_{k,i},$$

which contradicts the assumption that $C_{k,j} \in Pa_k^{\mathrm{blk}}(C_{k,i})$, since a block parent cannot have another block strictly between it and $C_{k,i}$ in the block-level partial order.

Hence the path cannot pass through any third block. It follows that along the chosen directed path, once the path leaves $C_{k,j}$, it must enter $C_{k,i}$ immediately. Therefore, there exists an index $r \in \{1, \ldots, m\}$ such that

$$w_{r-1} \in C_{k,j}, \qquad w_r \in C_{k,i},$$

and

$$w_{r-1} \to w_r$$

is a directed edge in $\mathcal{G}_k$. Consequently,

$$w_{r-1} \in Pa_k(w_r).$$

Setting

$$u^\star \triangleq w_{r-1} \in C_{k,j}, \qquad v^\star \triangleq w_r \in C_{k,i},$$

we obtain a pair of nodes such that $u^\star \in Pa_k(v^\star)$, as required. □

G BLOCK RQ DECOMPOSITION

We now give a constructive procedure for the block-RQ decomposition introduced in Section 3.2. Given $\mathbf{H}_k$, a block topological order, and the corresponding block ancestral relation, the algorithm builds $\mathbf{Q}_k$ block by block in reverse topological order. At each step, the current block row of $\mathbf{H}_k$ is projected onto the orthogonal complement of the row space already explained by its ancestor blocks, and the resulting innovation component is used to define the new block row of $\mathbf{Q}_k$. The coefficient matrix $\mathbf{R}_k$ is then obtained by expressing each block row of $\mathbf{H}_k$ in the row span of the corresponding ancestor blocks together with its own newly added block.

Algorithm 4 Block-RQ decomposition

Input: $\mathbf{H}_k \in \mathbb{R}^{p_k \times n_k}$, block topological order $C_r, C_{r-1}, \ldots, C_1$, block ancestor relation $anc_k^{\mathrm{blk}}(C_\ell)$ for each block C_ℓ
Output: $\mathbf{R}_k \in \mathbb{R}^{p_k \times p_k}$, $\mathbf{Q}_k \in \mathbb{R}^{p_k \times n_k}$

1: Initialize $\mathbf{R}_k = \mathbf{0}_{p_k \times p_k}$, $\mathbf{Q}_k = \mathbf{0}_{p_k \times n_k}$
2: **for** $\ell = r, r-1, \ldots, 1$ **do**
3: Let $\widetilde{\mathbf{H}}_\ell = [\mathbf{H}_k]_{C_\ell,:}$
4: Let $\mathbf{W}_\ell = [\mathbf{Q}_k]_{anc_k^{\mathrm{blk}}(C_\ell),:}$
5: **if** $\mathbf{W}_\ell$ is empty **then**
6: $\widetilde{\mathbf{H}}_\ell^{\perp} = \widetilde{\mathbf{H}}_\ell$
7: **else**
8: $\widetilde{\mathbf{H}}_\ell^{\perp} = \Pi_{\mathbf{W}_\ell^{\perp}}(\widetilde{\mathbf{H}}_\ell)$
9: **end if**
10: Normalize each row of $\widetilde{\mathbf{H}}_\ell^{\perp}$ to unit norm, and denote the resulting matrix by $[\mathbf{Q}_k]_{C_\ell,:}$
11: Let $J_\ell = C_\ell \cup Anc(C_\ell)$
12: Set

$$[\mathbf{R}_k]_{C_\ell, J_\ell} = \widetilde{\mathbf{H}}_\ell \big([\mathbf{Q}_k]_{J_\ell,:}\big)^{\dagger}$$

13: Set $\mathbf{R}_k[C_\ell, J_\ell^c] = 0$
14: **end for**
15: **return** $\mathbf{R}_k, \mathbf{Q}_k$

The next proposition shows that the above procedure indeed produces a valid block-RQ decomposition in the sense of Section 3.2.

Proposition A.3. *The output $(\mathbf{R}_k, \mathbf{Q}_k)$ of Algorithm 4 satisfies the block-RQ requirements in Section 3.2. In particular,*

1. $\mathbf{H}_k = \mathbf{R}_k \mathbf{Q}_k$*;*
2. *for each block C_ℓ, the row block $[\mathbf{R}_k]_{C_\ell,:}$ is supported only on $J_\ell = Anc_k^{\mathrm{blk}}(C_\ell)$, namely $[\mathbf{R}_k]_{C_\ell, J_\ell^c} = 0$;*
3. *for each block C_ℓ, the row block $[\mathbf{Q}_k]_{C_\ell,:}$ is orthogonal to the row space generated by its strict block ancestors, i.e.*

$$\mathrm{row}\big([\mathbf{Q}_k]_{C_\ell,:}\big) \perp \mathrm{row}\big([\mathbf{Q}_k]_{anc_k^{\mathrm{blk}}(C_\ell),:}\big),$$

and

$$\mathrm{row}\big([\mathbf{Q}_k]_{C_\ell,:}\big) = \mathrm{row}\left(\Pi_{\mathrm{row}([\mathbf{Q}_k]_{anc_k^{\mathrm{blk}}(C_\ell),:})^{\perp}}\big([\mathbf{H}_k]_{C_\ell,:}\big)\right).$$

Proof. Fix a block C_ℓ. By construction,

$$\widetilde{\mathbf{H}}_\ell^{\perp} = \Pi_{\mathrm{row}([\mathbf{Q}_k]_{anc_k^{\mathrm{blk}}(C_\ell),:})^{\perp}}\big([\mathbf{H}_k]_{C_\ell,:}\big),$$

with the convention that $\widetilde{\mathbf{H}}_\ell^{\perp} = [\mathbf{H}_k]_{C_\ell,:}$ if $anc_k^{\mathrm{blk}}(C_\ell) = \emptyset$. Since Step 10 only normalizes the rows of $\widetilde{\mathbf{H}}_\ell^{\perp}$, it does not change its row space. Therefore

$$\mathrm{row}\big([\mathbf{Q}_k]_{C_\ell,:}\big) = \mathrm{row}\big(\widetilde{\mathbf{H}}_\ell^{\perp}\big) = \mathrm{row}\left(\Pi_{\mathrm{row}([\mathbf{Q}_k]_{anc_k^{\mathrm{blk}}(C_\ell),:})^{\perp}}\big([\mathbf{H}_k]_{C_\ell,:}\big)\right).$$

Because $\widetilde{\mathbf{H}}_\ell^\perp$ is defined as the orthogonal projection onto

$$\mathrm{row}([\mathbf{Q}_k]_{anc_k^{\mathrm{blk}}(C_\ell),:})^\perp,$$

its row space is orthogonal to $\mathrm{row}([\mathbf{Q}_k]_{anc_k^{\mathrm{blk}}(C_\ell),:})$. Hence

$$\mathrm{row}\big([\mathbf{Q}_k]_{C_\ell,:}\big) \perp \mathrm{row}\big([\mathbf{Q}_k]_{anc_k^{\mathrm{blk}}(C_\ell),:}\big),$$

which proves the third claim.

Next, by the orthogonal decomposition of $[\mathbf{H}_k]_{C_\ell,:}$ with respect to

$$\mathrm{row}\big([\mathbf{Q}_k]_{anc_k^{\mathrm{blk}}(C_\ell),:}\big),$$

we have

$$[\mathbf{H}_k]_{C_\ell,:} = \Pi_{\mathrm{row}([\mathbf{Q}_k]_{anc_k^{\mathrm{blk}}(C_\ell),:})}\big([\mathbf{H}_k]_{C_\ell,:}\big) + \widetilde{\mathbf{H}}_\ell^\perp.$$

Therefore,

$$\mathrm{row}\big([\mathbf{H}_k]_{C_\ell,:}\big) \subseteq \mathrm{row}\big([\mathbf{Q}_k]_{J_\ell,:}\big), \qquad J_\ell = Anc_k^{\mathrm{blk}}(C_\ell).$$

It follows that

$$[\mathbf{H}_k]_{C_\ell,:} = [\mathbf{R}_k]_{C_\ell,J_\ell}[\mathbf{Q}_k]_{J_\ell,:},$$

where

$$[\mathbf{R}_k]_{C_\ell,J_\ell} = [\mathbf{H}_k]_{C_\ell,:}\big([\mathbf{Q}_k]_{J_\ell,:}\big)^\dagger.$$

Together with the assignment $[\mathbf{R}_k]_{C_\ell,J_\ell^c} = 0$, this gives

$$[\mathbf{H}_k]_{C_\ell,:} = [\mathbf{R}_k]_{C_\ell,:}\mathbf{Q}_k.$$

Stacking over all blocks yields

$$\mathbf{H}_k = \mathbf{R}_k\mathbf{Q}_k,$$

which proves the first claim.

Finally, the second claim follows directly from the construction of $\mathbf{R}_k$, since Step 13 explicitly sets

$$[\mathbf{R}_k]_{C_\ell,J_\ell^c} = 0$$

for every block C_ℓ. Thus each block row of $\mathbf{R}_k$ is supported only on the current block and its block ancestors. □

This decomposition is generally not unique, since the algorithm only determines each block through its row space, while different normalized bases of the same subspace lead to different matrix representatives $\mathbf{Q}_k$ and hence different $\mathbf{R}_k$. However, for Theorem 1, the relevant object is precisely the family of block row spaces $\{\mathrm{row}([\mathbf{Q}_k]_{C_\ell,:})\}_\ell$, rather than a particular matrix representative. These block row spaces are uniquely determined by $\mathbf{H}_k$, the prescribed block order, and the block ancestral relation, since each step extracts the orthogonal innovation subspace relative to already determined ancestor blocks. In this subspace sense, the decomposition is unique, which is sufficient for the identifiability argument in Theorem 1.

Proposition A.4 (Proposition 3 in the main text)**.** *For every block $C \in \mathcal{C}_k$ and every $i \in \rho_k(C)$, we have* $[\mathbf{H}_k]_{i,:} \in \mathrm{row}([\mathbf{Q}_k]_{\bigcup_{C' \in Anc_k^{\mathrm{blk}}(C)} \rho_k(C'),:})$.

Proof. Fix a block $C \in \mathcal{C}_k$ and an index $i \in \rho_k(C)$. By Definition 2, the block-RQ decomposition satisfies

$$\mathbf{H}_k = \mathbf{R}_k\mathbf{Q}_k.$$

Hence the ith row of $\mathbf{H}_k$ can be written as

$$[\mathbf{H}_k]_{i,:} = \sum_{m=1}^{p_k} [\mathbf{R}_k]_{i,m}[\mathbf{Q}_k]_{m,:}.$$

That is, $[\mathbf{H}_k]_{i,:}$ is a linear combination of the rows of $\mathbf{Q}_k$, with coefficients given by the ith row of $\mathbf{R}_k$.

Since $i \in \rho_k(C)$, the ith row belongs to the block row indexed by C. By the block support property in Definition 2, this block row of $\mathbf{R}_k$ can be nonzero only on the column blocks corresponding to $Anc_k^{\mathrm{blk}}(C)$. Equivalently, for any block $C' \notin Anc_k^{\mathrm{blk}}(C)$,

$$[\mathbf{R}_k]_{i,\rho_k(C')} = 0.$$

Therefore, the above sum reduces to

$$[\mathbf{H}_k]_{i,:} = \sum_{C' \in Anc_k^{\mathrm{blk}}(C)} \sum_{m \in \rho_k(C')} [\mathbf{R}_k]_{i,m} [\mathbf{Q}_k]_{m,:}.$$

This shows that $[\mathbf{H}_k]_{i,:}$ lies in the row space generated by the rows of $\mathbf{Q}_k$ indexed by the ancestor blocks of C, namely

$$[\mathbf{H}_k]_{i,:} \in \mathrm{row}\Big([\mathbf{Q}_k]_{\cup_{C' \in Anc_k^{\mathrm{blk}}(C)} \rho_k(C'),:} \Big).$$

The claim follows. □

H Proof of Lemma 1

Lemma A.1 (Lemma 1 in the main text)**.** *For each client k and each intervention j targeting node t_j (which is unknown to the learner), we have*

$$\mathrm{row}(\mathbf{\Theta}_k^{(j)} - \mathbf{\Theta}_k^{(0)}) \subseteq \langle \mathbf{H}_k^T (\mathbf{I} - \mathbf{A}_k)^T \mathbf{\Sigma}_k^{-1} \mathbf{e}_{\rho_k(t_j)}, \mathbf{H}_k^T \Big((\mathbf{I} - \mathbf{A}_k)^T \mathbf{\Sigma}_k^{-1} \mathbf{u}_k^{(j)} + \mathbf{d}_k^{(j)} \Big) \rangle. \tag{16}$$

where $\mathbf{u}_k^{(j)}, \mathbf{d}_k^{(j)}$ have the same definition as in Assumption 2. Therefore, it immediately follows that $\mathrm{rank}(\mathbf{\Theta}_k^{(j)} - \mathbf{\Theta}_k^{(0)}) \leq 2$.

Proof. Fix a client k and an intervention $j \in \mathcal{J}_k$, and write

$$\mathbf{B}_k \coloneqq \mathbf{I} - \mathbf{A}_k, \qquad \mathbf{\Omega}_k \coloneqq \mathbf{\Sigma}_k^{-1}, \qquad e \coloneqq \mathbf{e}_{\rho_k(t_j)}, \qquad d \coloneqq \mathbf{d}_k^{(j)}, \qquad u \coloneqq \mathbf{u}_k^{(j)}.$$

By the precision factorization established earlier, we have

$$\mathbf{\Theta}_k^{(j)} - \mathbf{\Theta}_k^{(0)} = \mathbf{H}_k^T \Big((\mathbf{I} - \mathbf{A}_k^{(j)})^T (\mathbf{\Sigma}_k^{(j)})^{-1} (\mathbf{I} - \mathbf{A}_k^{(j)}) - \mathbf{B}_k^T \mathbf{\Omega}_k \mathbf{B}_k \Big) \mathbf{H}_k.$$

Hence it suffices to study

$$\mathbf{\Delta}_k^{(j)} \coloneqq (\mathbf{I} - \mathbf{A}_k^{(j)})^T (\mathbf{\Sigma}_k^{(j)})^{-1} (\mathbf{I} - \mathbf{A}_k^{(j)}) - \mathbf{B}_k^T \mathbf{\Omega}_k \mathbf{B}_k.$$

By Lemma A.4, the intervention changes only the row corresponding to the target node, so

$$\mathbf{A}_k^{(j)} = \mathbf{A}_k + e d^T, \qquad \text{hence} \qquad \mathbf{I} - \mathbf{A}_k^{(j)} = \mathbf{B}_k - e d^T.$$

Moreover, there exists a scalar $\eta_k^{(j)}$ such that

$$\mathbf{\Sigma}_k^{(j)} = \mathbf{\Sigma}_k + e u^T + u e^T + \eta_k^{(j)} e e^T.$$

For brevity, write $\eta \coloneqq \eta_k^{(j)}$, and define

$$\mathbf{V} \coloneqq [e, u], \qquad \mathbf{W} \coloneqq [u + \eta e, e].$$

Then

$$\mathbf{\Sigma}_k^{(j)} = \mathbf{\Sigma}_k + \mathbf{V} \mathbf{W}^T.$$

Applying the Woodbury identity gives

$$(\mathbf{\Sigma}_k^{(j)})^{-1} = \mathbf{\Omega}_k - \mathbf{\Omega}_k \mathbf{V} \big(\mathbf{I} + \mathbf{W}^T \mathbf{\Omega}_k \mathbf{V} \big)^{-1} \mathbf{W}^T \mathbf{\Omega}_k.$$

Substituting this into $\mathbf{\Delta}_k^{(j)}$ yields

$$\begin{aligned} \mathbf{\Delta}_k^{(j)} = {} & (\mathbf{B}_k - e d^T)^T \mathbf{\Omega}_k (\mathbf{B}_k - e d^T) - \mathbf{B}_k^T \mathbf{\Omega}_k \mathbf{B}_k \\ & - (\mathbf{B}_k - e d^T)^T \mathbf{\Omega}_k \mathbf{V} \big(\mathbf{I} + \mathbf{W}^T \mathbf{\Omega}_k \mathbf{V} \big)^{-1} \mathbf{W}^T \mathbf{\Omega}_k (\mathbf{B}_k - e d^T). \end{aligned} \tag{17}$$

We now introduce

$$q \coloneqq \mathbf{B}_k^T \mathbf{\Omega}_k e, \qquad s \coloneqq \mathbf{B}_k^T \mathbf{\Omega}_k u.$$

Then the first line of equation 17 becomes

$$(\mathbf{B}_k - ed^T)^T \mathbf{\Omega}_k (\mathbf{B}_k - ed^T) - \mathbf{B}_k^T \mathbf{\Omega}_k \mathbf{B}_k = -qd^T - dq^T + (e^T \mathbf{\Omega}_k e)\, dd^T.$$

For the second line, direct multiplication gives

$$(\mathbf{B}_k - ed^T)^T \mathbf{\Omega}_k \mathbf{V} = \big[q - (e^T \mathbf{\Omega}_k e)d,\ s - (e^T \mathbf{\Omega}_k u)d\big],$$

and

$$\mathbf{W}^T \mathbf{\Omega}_k (\mathbf{B}_k - ed^T) = \begin{bmatrix} \big(s + \eta q - (e^T \mathbf{\Omega}_k u + \eta\, e^T \mathbf{\Omega}_k e)d\big)^T \\ \big(q - (e^T \mathbf{\Omega}_k e)d\big)^T \end{bmatrix}.$$

Let

$$\tau \coloneqq e^T \mathbf{\Omega}_k e, \qquad \kappa \coloneqq e^T \mathbf{\Omega}_k u, \qquad \nu \coloneqq u^T \mathbf{\Omega}_k u.$$

Then

$$\mathbf{I} + \mathbf{W}^T \mathbf{\Omega}_k \mathbf{V} = \begin{pmatrix} 1 + \kappa + \eta\tau & \nu + \eta\kappa \\ \tau & 1 + \kappa \end{pmatrix},$$

which is invertible since $\mathbf{\Sigma}_k^{(j)}$ is invertible. A straightforward simplification of equation 17 then yields

$$\mathbf{\Delta}_k^{(j)} = c_1\, qq^T + c_2\, q(s+d)^T + c_2\, (s+d)q^T + c_3\, (s+d)(s+d)^T$$

for some scalars c_1, c_2, c_3 depending on (k, j). For example, one may take

$$c_1 = \frac{\nu - \eta}{(1+\kappa)^2 + \tau(\eta - \nu)}, \qquad c_2 = -\frac{1+\kappa}{(1+\kappa)^2 + \tau(\eta - \nu)}, \qquad c_3 = \frac{\tau}{(1+\kappa)^2 + \tau(\eta - \nu)}.$$

In particular,

$$\mathrm{row}(\mathbf{\Delta}_k^{(j)}) \subseteq \langle q,\ s + d \rangle.$$

Recalling the definitions of q and s, this becomes

$$\mathrm{row}(\mathbf{\Delta}_k^{(j)}) \subseteq \left\langle \mathbf{B}_k^T \mathbf{\Omega}_k e,\ \mathbf{B}_k^T \mathbf{\Omega}_k u + d \right\rangle = \left\langle (\mathbf{I} - \mathbf{A}_k)^T \mathbf{\Sigma}_k^{-1} \mathbf{e}_{\rho_k(t_j)},\ (\mathbf{I} - \mathbf{A}_k)^T \mathbf{\Sigma}_k^{-1} \mathbf{u}_k^{(j)} + \mathbf{d}_k^{(j)} \right\rangle.$$

Finally, since

$$\mathbf{\Theta}_k^{(j)} - \mathbf{\Theta}_k^{(0)} = \mathbf{H}_k^T \mathbf{\Delta}_k^{(j)} \mathbf{H}_k,$$

we conclude that

$$\mathrm{row}(\mathbf{\Theta}_k^{(j)} - \mathbf{\Theta}_k^{(0)}) \subseteq \left\langle \mathbf{H}_k^T (\mathbf{I} - \mathbf{A}_k)^T \mathbf{\Sigma}_k^{-1} \mathbf{e}_{\rho_k(t_j)},\ \mathbf{H}_k^T \Big((\mathbf{I} - \mathbf{A}_k)^T \mathbf{\Sigma}_k^{-1} \mathbf{u}_k^{(j)} + \mathbf{d}_k^{(j)} \Big) \right\rangle.$$

Therefore

$$\mathrm{rank}(\mathbf{\Theta}_k^{(j)} - \mathbf{\Theta}_k^{(0)}) \le 2.$$

This proves the lemma. □

I Proof of Lemma 2

Lemma A.2 (Lemma 2 in the main text)**.** *For each client k and intervention j, suppose that the intervention target node t_j belongs to block $C_{k,a} \in \mathcal{C}_k$. Then*

$$\mathrm{row}(\mathbf{\Theta}_k^{(j)} - \mathbf{\Theta}_k^{(0)}) \subseteq \left\langle [\mathbf{Q}_k]_{\rho_k(i),:} \;\middle|\; i \in C_{k,b},\ C_{k,b} \in Anc_k^{\mathrm{blk}}(C_{k,a}) \right\rangle.$$

Proof. Let $t_j \in C_{k,a}$ and write $r = \rho_k(t_j)$. By Lemma 1,

$$\mathrm{row}(\mathbf{\Theta}_k^{(j)} - \mathbf{\Theta}_k^{(0)}) \subseteq \left\langle \mathbf{H}_k^T (\mathbf{I} - \mathbf{A}_k)^T \mathbf{\Sigma}_k^{-1} \mathbf{e}_r,\ \mathbf{H}_k^T \Big((\mathbf{I} - \mathbf{A}_k)^T \mathbf{\Sigma}_k^{-1} \mathbf{u}_k^{(j)} + \mathbf{d}_k^{(j)} \Big) \right\rangle.$$

Hence it suffices to show that both generators on the right-hand side belong to

$$\left\langle [\mathbf{Q}_k]_{\rho_k(i),:} \;\middle|\; i \in C_{k,b},\ C_{k,b} \in Anc_k^{\mathrm{blk}}(C_{k,a}) \right\rangle.$$

For the first generator, define

$$\boldsymbol{\alpha} \coloneqq (\mathbf{I} - \mathbf{A}_k)^T \boldsymbol{\Sigma}_k^{-1} \mathbf{e}_r.$$

Then

$$\mathbf{H}_k^T (\mathbf{I} - \mathbf{A}_k)^T \boldsymbol{\Sigma}_k^{-1} \mathbf{e}_r = \mathbf{H}_k^T \boldsymbol{\alpha} = \sum_{i \in V_k} \alpha_i [\mathbf{H}_k]_{i,:}^T.$$

Since $r = \rho_k(t_j)$ and $t_j \in C_{k,a}$, any nonzero coefficient α_i can only arise from indices i belonging to blocks in $Anc_k^{\mathrm{blk}}(C_{k,a})$. Therefore the above vector is a linear combination of rows of $\mathbf{H}_k$ indexed by nodes in blocks belonging to $Anc_k^{\mathrm{blk}}(C_{k,a})$.

For the second generator, define

$$\boldsymbol{\beta} \coloneqq (\mathbf{I} - \mathbf{A}_k)^T \boldsymbol{\Sigma}_k^{-1} \mathbf{u}_k^{(j)} + \mathbf{d}_k^{(j)}.$$

Then

$$\mathbf{H}_k^T \left((\mathbf{I} - \mathbf{A}_k)^T \boldsymbol{\Sigma}_k^{-1} \mathbf{u}_k^{(j)} + \mathbf{d}_k^{(j)} \right) = \mathbf{H}_k^T \boldsymbol{\beta} = \sum_{i \in V_k} \beta_i [\mathbf{H}_k]_{i,:}^T.$$

By Lemma A.4,

$$\mathrm{supp}(\mathbf{d}_k^{(j)}) \subseteq \rho_k\big(pa_k(t_j)\big), \qquad \mathrm{supp}(\mathbf{u}_k^{(j)}) \subseteq \rho_k\big(sp_k(t_j)\big).$$

Hence the nonzero coordinates of $\boldsymbol{\beta}$ can only come from set $\rho_k(Pa_k(dis_k(t_j) \cup \{t_j\}))$, where $dis_k(i) \triangleq \{j \in \mathcal{V}_k \mid i \text{ and } j \text{ are connected by a bidirected path in } \mathcal{D}_k\}$. By the definition of the block structure and the block-level ancestor relation, these nodes belong to blocks contained in $Anc_k^{\mathrm{blk}}(C_{k,a})$. It follows that the second generator is also a linear combination of rows of $\mathbf{H}_k$ indexed by nodes in blocks belonging to $Anc_k^{\mathrm{blk}}(C_{k,a})$.

Now apply Proposition 3. For any node i lying in a block $C \in \mathcal{C}_k$, Proposition 3 implies that

$$[\mathbf{H}_k]_{i,:} \in \left\langle [\mathbf{Q}_k]_{\rho_k(\ell),:} \;\middle|\; \ell \in C_{k,b},\ C_{k,b} \in Anc_k^{\mathrm{blk}}(C) \right\rangle.$$

In particular, whenever $C \in Anc_k^{\mathrm{blk}}(C_{k,a})$, we also have

$$[\mathbf{H}_k]_{i,:} \in \left\langle [\mathbf{Q}_k]_{\rho_k(\ell),:} \;\middle|\; \ell \in C_{k,b},\ C_{k,b} \in Anc_k^{\mathrm{blk}}(C_{k,a}) \right\rangle.$$

Therefore both generators in Lemma 1 lie in

$$\left\langle [\mathbf{Q}_k]_{\rho_k(i),:} \;\middle|\; i \in C_{k,b},\ C_{k,b} \in Anc_k^{\mathrm{blk}}(C_{k,a}) \right\rangle.$$

Taking their span proves that

$$\mathrm{row}(\boldsymbol{\Theta}_k^{(j)} - \boldsymbol{\Theta}_k^{(0)}) \subseteq \left\langle [\mathbf{Q}_k]_{\rho_k(i),:} \;\middle|\; i \in C_{k,b},\ C_{k,b} \in Anc_k^{\mathrm{blk}}(C_{k,a}) \right\rangle.$$

This completes the proof. □

J Proof of Theorem 1

Theorem A.1 (Theorem 1 in the main text)**.** *Assume Assumptions 1 and 2. Let $(\hat{\mathcal{C}}_k, \hat{\prec}_k^{\mathrm{blk}}, \hat{\mathbf{Q}}_k)$ be the output of Algorithm 2 for client k. Then:*

(i) $(\hat{\mathcal{C}}_k, \hat{\prec}_k^{\mathrm{blk}})$ recovers $(\mathcal{C}_k, \prec_k^{\mathrm{blk}})$ up to relabeling under the bijection $\psi : j \mapsto t_j$.

(ii) For each true block $C \in \mathcal{C}_k$, let $\hat{C} = \psi^{-1}(C)$ denote the corresponding estimated block. Then $\langle \hat{\mathbf{q}}_{k,i} : i \in \hat{C} \rangle = \mathrm{row}([\mathbf{Q}_k]_{\rho_k(C),:})$.

Consequently, $\hat{\mathbf{Q}}_k = \mathbf{S}_k \mathbf{P}_{\sigma,k} \mathbf{Q}_k$, where $\mathbf{S}_k$ is an invertible block-diagonal norm-preserving matrix and $\mathbf{P}_{\sigma,k}$ is a row permutation matrix preserving the block ancestral order.

Proof. We proceed by induction on the outer-loop index of Algorithm 2. Let $T = |\mathcal{C}_k|$ be the number of true blocks for client k. For each $t \in [T]$, we view the t-th iteration of the outer loop as the step that recovers the t-th block in the reverse block topological order.

At the beginning of iteration t, we maintain the following induction hypothesis:

(IH1) $\mathcal{I}_{t-1}$ consists of correctly identified blocks, and its image under ψ is closed under the true block-level ancestor relation. Namely, if $C \in \psi(\mathcal{I}_{t-1})$, then $Anc_k^{\mathrm{blk}}(C) \subseteq \psi(\mathcal{I}_{t-1})$.

(IH2) The last $t-1$ recovered block rows of $\hat{\mathbf{Q}}_k$ coincide with the corresponding block rows of some matrix of the form $\mathbf{S}_k\mathbf{P}_{\sigma,k}\mathbf{Q}_k$, where $\mathbf{P}_{\sigma,k}$ is a permutation consistent with the recovered block order and $\mathbf{S}_k$ is a block-diagonal invertible matrix.

The base case $t = 1$ is immediate: $\mathcal{I}_0 = \emptyset$, so (IH1) holds trivially, and (IH2) is vacuous.

Now assume (IH1)–(IH2) hold at the beginning of iteration t. We prove that the t-th iteration correctly identifies one new block, correctly updates its ancestor information, and appends the corresponding correct block row space to $\hat{\mathbf{Q}}_k$.

Step 1: identification of the new block in Steps 4–12 of Algorithm 2. Let

$$W_t \triangleq \langle \hat{\mathbf{q}}_{k,i} : i \in \mathcal{I}_{t-1} \rangle.$$

We first record three claims.

Claim 1. Under Assumptions 1 and 2, let $J = \{j_1, \ldots, j_r\} \subseteq R_{t-1}$. If $\psi(J)$ is not a disjoint union of whole true blocks, then

$$\dim\big(\mathcal{P}_k(J; W_t)\big) > r.$$

*Claim 2.*Under Assumptions 1 and 2, let $J = \{j_1, \ldots, j_r\} \subseteq R_{t-1}$. Suppose that $\psi(J)$ is a union of whole true blocks, but at least one true block parent of $\psi(J)$ is not contained in $\psi(\mathcal{I}_{t-1} \cup J)$. Then

$$\dim\big(\mathcal{P}_k(J; W_t)\big) > r.$$

Claim 3. Under Assumptions 1 and 2, let $C \in \mathcal{C}_k$ satisfy $C \notin \psi(\mathcal{I}_{t-1})$ and

$$anc_k^{\mathrm{blk}}(C) \subseteq \psi(\mathcal{I}_{t-1}).$$

Let $J = \psi^{-1}(C) \subseteq R_{t-1}$. Then

$$\dim\big(\mathcal{P}_k(J; W_t)\big) = |J| = |C|.$$

We postpone the proofs of the three claims and first show how they imply correctness of Steps 4–9.

Since $\psi(\mathcal{I}_{t-1})$ is ancestor-closed by (IH1), there exists at least one true block

$$C^\star \in \mathcal{C}_k \setminus \psi(\mathcal{I}_{t-1})$$

whose true block ancestors are all contained in $\psi(\mathcal{I}_{t-1})$; equivalently, $C^\star$ is minimal in the remaining block partial order. By Claim 3, if $J^\star = \psi^{-1}(C^\star)$, then

$$\dim\big(\mathcal{P}_k(J^\star; W_t)\big) = |J^\star|.$$

Hence the inner loop of Algorithm 2 must terminate.

Now let $\hat{C}$ be the set selected at Step 8 when the inner loop terminates. By construction, $\hat{C}$ is chosen with the smallest cardinality r such that

$$\dim\big(\mathcal{P}_k(\hat{C}; W_t)\big) = r.$$

Claim 1 rules out the possibility that $\psi(\hat{C})$ contains only part of some true block. Claim 2 rules out the possibility that $\psi(\hat{C})$ is a union of whole true blocks whose true block parents have not all been recovered already. Therefore, $\psi(\hat{C})$ must be exactly one whole true block, say C_t, and moreover

$$pa_k^{\mathrm{blk}}(C_t) \subseteq \psi(\mathcal{I}_{t-1}).$$

Since $\psi(\mathcal{I}_{t-1})$ is ancestor-closed by (IH1), this further implies

$$anc_k^{\mathrm{blk}}(C_t) \subseteq \psi(\mathcal{I}_{t-1}).$$

Step 2: identification of the ancestor set in Algorithm 1. We next analyze the call

$$(\{\hat{\mathbf{q}}_{k,i}\}_{i\in\hat{C}}, \mathcal{A}) = \text{ID-Block-Ancestor}\Big(\hat{C}, \{\Theta_k^{(j)}\}_{j\in\{0\}\cup\hat{C}}, \mathcal{I}_{t-1}, \{\hat{\mathbf{q}}_{k,i}\}_{i\in\mathcal{I}_{t-1}}\Big).$$

For each block $B \in \mathcal{I}_{t-1}$, Algorithm 1 removes B temporarily and checks whether

$$\dim\big(\mathcal{P}_k(\hat{C}; W_{-B})\big) = |\hat{C}|, \qquad W_{-B} \triangleq \langle \hat{\mathbf{q}}_{k,i} : i \in \mathcal{I}_{t-1}\backslash B\rangle.$$

We use Claim 2 again. If B is a true block parent of C_t, then after removing B, the projected space still retains an unremoved parent contribution, hence

$$\dim\big(\mathcal{P}_k(\hat{C}; W_{-B})\big) > |\hat{C}|,$$

so B cannot be deleted from $\mathcal{A}$. Therefore,

$$pa_k^{\text{blk}}(C_t) \subseteq \mathcal{A}.$$

We also record one additional claim.

Claim 4. Under Assumptions 1 and 2, if $B \in \mathcal{I}_{t-1}$ is not a true block ancestor of C_t, then

$$\dim\big(\mathcal{P}_k(\hat{C}; W_{-B})\big) = |\hat{C}|,$$

and hence Algorithm 1 removes B from $\mathcal{A}$.

Again postponing its proof, Claim 4 implies

$$\mathcal{A} \subseteq anc_k^{\text{blk}}(C_t).$$

Combining this with the previous inclusion yields

$$pa_k^{\text{blk}}(C_t) \subseteq \mathcal{A} \subseteq anc_k^{\text{blk}}(C_t).$$

Since $\psi(\mathcal{I}_{t-1})$ is ancestor-closed by (IH1), every true ancestor of every block in $\mathcal{I}_{t-1}$ has already been recovered. Therefore, once Algorithm 1 returns $\mathcal{A}$, the update rule in Step 14 of Algorithm 2 correctly inserts all block-level ancestral relations between the newly found block C_t and the previously recovered true ancestor blocks. In particular, the recovered partial order remains correct after adding $\hat{C}$.

Step 3: update of the block row space. Let

$$W \triangleq \langle \hat{\mathbf{q}}_{k,i} : i \in \mathcal{A}\rangle.$$

By the conclusion of Step 2, W is exactly the span of the previously recovered block row spaces corresponding to true block ancestors of C_t. Hence

$$\mathcal{P}_k(\hat{C}; W)$$

is precisely the innovation subspace associated with the block C_t. By the uniqueness statement for the block-RQ decomposition at the level of block row spaces, this innovation subspace is uniquely determined by $\mathbf{H}_k$, the true block order, and the true ancestor structure. Therefore, the normalized vectors $\{\hat{\mathbf{q}}_{k,i}\}_{i\in\hat{C}}$ constructed at Step 10 of Algorithm 1 span exactly the block row space corresponding to C_t in some representation of the form $\mathbf{S}_k\mathbf{P}_{\sigma,k}\mathbf{Q}_k$.

After Step 15 of Algorithm 2, the last t recovered block rows of $\hat{\mathbf{Q}}_k$ therefore coincide with the last t block rows of some $\mathbf{S}_k\mathbf{P}_{\sigma,k}\mathbf{Q}_k$. This proves (IH2) for the next iteration. Moreover, after Step 16, the set $\mathcal{I}_t$ consists of correctly recovered blocks and remains ancestor-closed under the true block-level partial order, proving (IH1) for the next iteration.

This completes the induction. After T iterations, all true blocks of client k have been recovered, the recovered block-level ancestral relation is correct, and the recovered $\hat{\mathbf{Q}}_k$ agrees blockwise with $\mathbf{Q}_k$ up to the intrinsic permutation and blockwise invertible transformation ambiguity. Hence the conclusion of the theorem follows. □

J.1 PROOF OF *Claim 1.*

Proof. Let

$$J = \{i_1, \dots, i_r\} \subseteq R_{t-1},$$

and suppose that the corresponding targets involve exactly the true blocks

$$C_{k,1}, \dots, C_{k,q} \in \mathcal{C}_k,$$

with at least one of these blocks only partially selected by J. For each $s \in [q]$, let p_s denote the length of the shortest block-parent path from $C_{k,s}$ to a root block in the true block partial order. Among all partially selected blocks, choose one, denoted by $C_{k,u}$, such that p_u is maximal.

Define

$$\mathcal{B}_1 \triangleq \{j \in J : C_k(j) = C_{k,s} \text{ for some } s \text{ with } p_s \le p_u\}, \qquad \mathcal{B}_2 \triangleq J \setminus \mathcal{B}_1.$$

Then $C_{k,u}$ is involved in $\mathcal{B}_1$, every block involved in $\mathcal{B}_2$ is fully selected, and no block involved in $\mathcal{B}_1$ is a descendant of $C_{k,u}$. We will show

$$\dim\big(\mathcal{P}_k(\mathcal{B}_1; W_t)\big) > |\mathcal{B}_1|, \qquad \dim\Big(\Pi_{W'^{\perp}}\big(\mathcal{P}_k(\mathcal{B}_2; W_t)\big)\Big) \ge |\mathcal{B}_2|$$

for a suitable subspace W' containing $\mathcal{P}_k(\mathcal{B}_1; W_t)$. This will imply

$$\dim\big(\mathcal{P}_k(J; W_t)\big) > |\mathcal{B}_1| + |\mathcal{B}_2| = r.$$

Step 1: the contribution of $\mathcal{B}_1$. By the induction hypothesis (IH2), the subspace

$$W_t = \langle \hat{\mathbf{q}}_{k,i} : i \in \mathcal{I}_{t-1} \rangle$$

coincides with the span of the corresponding true block rows of $\mathbf{Q}_k$, up to the intrinsic blockwise invertible transformation. Hence, for the purpose of dimension counting, we may work directly with the corresponding block rows of $\mathbf{Q}_k$ and suppress this equivalence in the notation.

Set

$$S_1 \triangleq \rho_k\Big(\bigcup_{j \in \mathcal{B}_1} Anc_k^{\mathrm{blk}}(C_k(j)) \setminus \psi(\mathcal{I}_{t-1})\Big).$$

By Lemma A.5,

$$\mathcal{P}_k(\mathcal{B}_1; W_t)$$

is contained in the span of the vectors

$$[\bar{\mathbf{H}}_k]_{S_1,:}^T (\mathbf{I} - [\mathbf{A}_k]_{S_1,S_1})^T [\mathbf{\Sigma}_k]_{S_1,S_1}^{-1} [\mathbf{e}_{\rho_k(t_j)}]_{S_1},$$

and

$$[\bar{\mathbf{H}}_k]_{S_1,:}^T \Big((\mathbf{I} - [\mathbf{A}_k]_{S_1,S_1})^T [\mathbf{\Sigma}_k]_{S_1,S_1}^{-1} [\mathbf{u}_k^{(j)}]_{S_1} + [\mathbf{d}_k^{(j)}]_{S_1}\Big), \qquad j \in \mathcal{B}_1.$$

Moreover, by the block-RQ property, the projected rows $[\bar{\mathbf{H}}_k]_{S_1,:}$ remain row independent on the residual innovation subspace. Therefore it suffices to prove that

$$\dim\Big\langle (\mathbf{I} - [\mathbf{A}_k]_{S_1,S_1})^T [\mathbf{\Sigma}_k]_{S_1,S_1}^{-1} [\mathbf{e}_{\rho_k(t_j)}]_{S_1}, \\ (\mathbf{I} - [\mathbf{A}_k]_{S_1,S_1})^T [\mathbf{\Sigma}_k]_{S_1,S_1}^{-1} [\mathbf{u}_k^{(j)}]_{S_1} + [\mathbf{d}_k^{(j)}]_{S_1} \;\Big|\; j \in \mathcal{B}_1 \Big\rangle > |\mathcal{B}_1|. \tag{18}$$

The first family,

$$(\mathbf{I} - [\mathbf{A}_k]_{S_1,S_1})^T [\mathbf{\Sigma}_k]_{S_1,S_1}^{-1} [\mathbf{e}_{\rho_k(t_j)}]_{S_1}, \qquad j \in \mathcal{B}_1,$$

already consists of $|\mathcal{B}_1|$ linearly independent vectors, since the standard basis vectors $[\mathbf{e}_{\rho_k(t_j)}]_{S_1}$ are distinct and

$$(\mathbf{I} - [\mathbf{A}_k]_{S_1,S_1})^T [\mathbf{\Sigma}_k]_{S_1,S_1}^{-1}$$

is invertible.

It remains to show that at least one vector from the second family is not in the span of the first family. Since $C_{k,u}$ is only partially selected by J, there exists an index

$$b \in \rho_k(C_{k,u})$$

such that b is not the target coordinate of any intervention in $\mathcal{B}_1 \cap \psi^{-1}(C_{k,u})$. By construction of $\mathcal{B}_1$, no descendant block of $C_{k,u}$ is involved in $\mathcal{B}_1$. Hence, after reordering the blocks in S_1 so that $C_{k,u}$ appears first, the block upper-triangular structure of $\mathbf{A}_k$ and the block-diagonal structure of $\mathbf{\Sigma}_k$ imply that for every $j \in \mathcal{B}_1 \cap \psi^{-1}(C_{k,u})$,

$$\Big[[\mathbf{\Sigma}_k]_{S_1,S_1}(\mathbf{I} - [\mathbf{A}_k]^T_{S_1,S_1})^{-1}[\mathbf{d}_k^{(j)}]_{S_1} + [\mathbf{u}_k^{(j)}]_{S_1}\Big]_b$$
$$= \Big[[\mathbf{\Sigma}_k]_{\rho_k(C_{k,u}),\rho_k(C_{k,u})}(\mathbf{I} - [\mathbf{A}_k]^T_{\rho_k(C_{k,u}),\rho_k(C_{k,u})})^{-1}[\mathbf{d}_k^{(j)}]_{\rho_k(C_{k,u})} + [\mathbf{u}_k^{(j)}]_{\rho_k(C_{k,u})}\Big]_b. \tag{19}$$

By the third clause of Assumption 2, there exists some

$$j^\star \in \mathcal{B}_1 \cap \psi^{-1}(C_{k,u})$$

such that the quantity in equation 19 is nonzero at the unselected index b.

Now suppose, for contradiction, that

$$(\mathbf{I} - [\mathbf{A}_k]_{S_1,S_1})^T[\mathbf{\Sigma}_k]^{-1}_{S_1,S_1}[\mathbf{u}_k^{(j^\star)}]_{S_1} + [\mathbf{d}_k^{(j^\star)}]_{S_1}$$

lies in the span of the first $|\mathcal{B}_1|$ vectors. Left-multiplying by the invertible matrix

$$[\mathbf{\Sigma}_k]_{S_1,S_1}(\mathbf{I} - [\mathbf{A}_k]^T_{S_1,S_1})^{-1}$$

would imply that

$$[\mathbf{u}_k^{(j^\star)}]_{S_1} + [\mathbf{\Sigma}_k]_{S_1,S_1}(\mathbf{I} - [\mathbf{A}_k]^T_{S_1,S_1})^{-1}[\mathbf{d}_k^{(j^\star)}]_{S_1}$$

lies in the span of the basis vectors $[\mathbf{e}_{\rho_k(t_j)}]_{S_1}$ for $j \in \mathcal{B}_1$. But every such basis vector vanishes at the unselected coordinate b, whereas by construction and Assumption 2, the vector above has a nonzero b-th entry. This is impossible. Therefore equation 18 holds, and hence

$$\dim\big(\mathcal{P}_k(\mathcal{B}_1; W_t)\big) > |\mathcal{B}_1|.$$

Step 2: the contribution of $\mathcal{B}_2$. Define

$$W' \triangleq \Big\langle [\mathbf{Q}_k]_{i,:} \;\Big|\; i \in \rho_k\Big(\bigcup_{j\in\mathcal{B}_1} Anc_k^{\mathrm{blk}}(C_k(j))\Big) \cup \mathcal{I}_{t-1}\Big\rangle.$$

By construction, $\mathcal{P}_k(\mathcal{B}_1; W_t) \subseteq W'$, and therefore

$$\Pi_{W'^\perp}\big(\mathcal{P}_k(\mathcal{B}_2; W_t)\big) = \mathcal{P}_k(\mathcal{B}_2; W').$$

Let

$$S_2 \triangleq \rho_k\Big(\bigcup_{j\in\mathcal{B}_2} Anc_k^{\mathrm{blk}}(C_k(j)) \setminus \Big(\bigcup_{j\in\mathcal{B}_1} Anc_k^{\mathrm{blk}}(C_k(j)) \cup \psi(\mathcal{I}_{t-1})\Big)\Big).$$

Applying Lemma A.5 again, it suffices to study the vectors

$$(\mathbf{I} - [\mathbf{A}_k]_{S_2,S_2})^T[\mathbf{\Sigma}_k]^{-1}_{S_2,S_2}[\mathbf{e}_{\rho_k(t_j)}]_{S_2}, \qquad j \in \mathcal{B}_2.$$

Since the blocks involved in $\mathcal{B}_2$ are all completely selected, these vectors are linearly independent; indeed, the corresponding basis vectors $[\mathbf{e}_{\rho_k(t_j)}]_{S_2}$ are distinct, and

$$(\mathbf{I} - [\mathbf{A}_k]_{S_2,S_2})^T[\mathbf{\Sigma}_k]^{-1}_{S_2,S_2}$$

is invertible. Hence

$$\dim\Big(\Pi_{W'^\perp}\big(\mathcal{P}_k(\mathcal{B}_2; W_t)\big)\Big) \geq |\mathcal{B}_2|.$$

Step 3: conclusion. Since $\mathcal{P}_k(\mathcal{B}_1; W_t) \subseteq W'$, we have

$$\dim\big(\mathcal{P}_k(J; W_t)\big) \geq \dim\big(\mathcal{P}_k(\mathcal{B}_1; W_t)\big) + \dim\Big(\Pi_{W'^\perp}\big(\mathcal{P}_k(\mathcal{B}_2; W_t)\big)\Big).$$

Combining the bounds established above gives

$$\dim\big(\mathcal{P}_k(J; W_t)\big) > |\mathcal{B}_1| + |\mathcal{B}_2| = r.$$

This proves Claim 1. □

J.2 PROOF OF *Claim 2.*

Proof. Let

$$J = \{i_1, \dots, i_r\} \subseteq R_{t-1},$$

and suppose that $\psi(J)$ is the disjoint union of whole true blocks

$$C_{k,1}, \dots, C_{k,q} \in \mathcal{C}_k.$$

Assume further that there exists a true block parent of one of these blocks which is not yet contained in $\psi(\mathcal{I}_{t-1})$ and is not among the selected blocks themselves. We will prove

$$\dim\big(\mathcal{P}_k(J; W_t)\big) > r.$$

Set

$$S \triangleq \rho_k\Big(Anc_k^{\mathrm{blk}}(C_{k,1} \cup \cdots \cup C_{k,q}) \setminus \psi(\mathcal{I}_{t-1})\Big).$$

By Lemma A.5, it suffices to show that

$$\dim\Big\langle (\mathbf{I} - [\mathbf{A}_k]_{S,S})^T [\mathbf{\Sigma}_k]_{S,S}^{-1} [\mathbf{e}_{\rho_k(t_j)}]_S, \\ (\mathbf{I} - [\mathbf{A}_k]_{S,S})^T [\mathbf{\Sigma}_k]_{S,S}^{-1} [\mathbf{u}_k^{(j)}]_S + [\mathbf{d}_k^{(j)}]_S \;\Big|\; j \in J \Big\rangle > r. \tag{20}$$

The first family,

$$(\mathbf{I} - [\mathbf{A}_k]_{S,S})^T [\mathbf{\Sigma}_k]_{S,S}^{-1} [\mathbf{e}_{\rho_k(t_j)}]_S, \qquad j \in J,$$

already consists of r linearly independent vectors, since the basis vectors $[\mathbf{e}_{\rho_k(t_j)}]_S$ are distinct and

$$(\mathbf{I} - [\mathbf{A}_k]_{S,S})^T [\mathbf{\Sigma}_k]_{S,S}^{-1}$$

is invertible. Therefore, to prove equation 20, it is enough to show that at least one vector from the second family is not contained in the span of the first family.

Choose a block $C_{k,i} \in \{C_{k,1}, \dots, C_{k,q}\}$ and a true block parent

$$C_{k,j} \in Pa_k^{\mathrm{blk}}(C_{k,i})$$

such that

$$C_{k,j} \notin \psi(\mathcal{I}_{t-1}) \cup \{C_{k,1}, \dots, C_{k,q}\}.$$

We claim that there exist

$$j^\star \in \psi^{-1}(C_{k,i}) \qquad \text{and} \qquad b \in \rho_k(C_{k,j})$$

such that

$$\Big([\mathbf{\Sigma}_k]_{S,S}(\mathbf{I} - [\mathbf{A}_k]_{S,S}^T)^{-1}[\mathbf{d}_k^{(j^\star)}]_S + [\mathbf{u}_k^{(j^\star)}]_S\Big)_b \neq 0. \tag{21}$$

We prove this by contradiction. Suppose instead that for every

$$j \in \psi^{-1}(C_{k,i}) \qquad \text{and every} \qquad b \in \rho_k(C_{k,j}),$$

one has

$$\Big([\mathbf{\Sigma}_k]_{S,S}(\mathbf{I} - [\mathbf{A}_k]_{S,S}^T)^{-1}[\mathbf{d}_k^{(j)}]_S + [\mathbf{u}_k^{(j)}]_S\Big)_b = 0. \tag{22}$$

Given the block-diagonal structure of $\mathbf{\Sigma}_k$ and the fact that $C_{k,j}$ is a block-level parent of $C_{k,i}$, it follows that

$$\Big([\mathbf{\Sigma}_k]_{S,S}(\mathbf{I} - [\mathbf{A}_k]_{S,S}^T)^{-1}[\mathbf{d}_k^{(j)}]_S + [\mathbf{u}_k^{(j)}]_S\Big)_b \\ = \Big([\mathbf{\Sigma}_k]_{\rho_k(C_{k,j}),\rho_k(C_{k,j})}(\mathbf{I} - [\mathbf{A}_k]_{S,S}^T)^{-1}_{\rho_k(C_{k,j}),\rho_k(C_{k,i}\cup C_{k,j})}[\mathbf{d}_k^{(j)}]_{\rho_k(C_{k,i}\cup C_{k,j}} + [\mathbf{u}_k^{(j)}]_{\rho_k(C_{k,i}\cup C_{k,j})}\Big)_b.$$

Enumerate the nodes in $C_{k,i}$ according to the true topological order within the block:

$$v_1, \dots, v_m.$$

We will show inductively that no node in $C_{k,j}$ can be an ancestor of any v_s, which contradicts the assumption that $C_{k,j} \in Pa_k^{\mathrm{blk}}(C_{k,i})$.

For the first node v_1, using Lemma A.4, by the support characterization of $\mathbf{d}_k^{(v_1)}$, there is no support in the coordinates corresponding to $\rho_k(C_{k,i})$, since v_1 has no parent inside $C_{k,i}$. Moreover, for $b \in \rho_k(C_{k,j})$, the term $[\mathbf{u}_k^{(v_1)}]_b$ vanishes by the support description of $\mathbf{u}_k^{(v_1)}$. Hence equation 22 implies

$$[\mathbf{\Sigma}_k]_{\rho_k(C_{k,j}),\rho_k(C_{k,j})}(\mathbf{I} - [\mathbf{A}_k]_{S,S}^T)^{-1}_{\rho_k(C_{k,j}),\rho_k(C_{k,j})}[\mathbf{d}_k^{(v_1)}]_{\rho_k(C_{k,j})} = \mathbf{0}.$$

Since

$$[\mathbf{\Sigma}_k]_{\rho_k(C_{k,j}),\rho_k(C_{k,j})}(\mathbf{I} - [\mathbf{A}_k]_{S,S}^T)^{-1}_{\rho_k(C_{k,j}),\rho_k(C_{k,j})}$$

is invertible, we obtain

$$[\mathbf{d}_k^{(v_1)}]_{\rho_k(C_{k,j})} = \mathbf{0}.$$

By Assumption 2 (the first part concerning $\mathbf{d}_k^{(j)}$), this implies that no node in $C_{k,j}$ is a parent of v_1. Hence no node in $C_{k,j}$ is an ancestor of v_1, and therefore

$$\left[(\mathbf{I} - [\mathbf{A}_k]_{S,S})^{-1}\right]_{\rho_k(C_{k,j}),\, \rho_k(v_1)} = 0. \tag{23}$$

Now suppose inductively that for some $s \geq 2$, we have already shown that for every $a < s$,

$$\left[(\mathbf{I} - [\mathbf{A}_k]_{S,S})^{-1}\right]_{\rho_k(C_{k,j}),\, \rho_k(v_a)} = 0.$$

We consider the node v_s. By the support characterization of $\mathbf{d}_k^{(v_s)}$, the coordinates of

$$[\mathbf{d}_k^{(v_s)}]_{\rho_k(C_{k,i})}$$

can only lie on earlier nodes $v_1, \ldots, v_{s-1}$ inside the same block. The only new within-block support that may appear relative to previous interventions is therefore on coordinates corresponding to $v_1, \ldots, v_{s-1}$. By the induction hypothesis, the columns of $(\mathbf{I} - [\mathbf{A}_k]_{S,S})^{-1}$ indexed by these earlier nodes vanish on the row block $\rho_k(C_{k,j})$. Hence these within-block contributions do not affect the coordinates in $\rho_k(C_{k,j})$.

Therefore, for $b \in \rho_k(C_{k,j})$, the identity equation 22 again reduces to

$$[\mathbf{\Sigma}_k]_{\rho_k(C_{k,j}),\rho_k(C_{k,j})}(\mathbf{I} - [\mathbf{A}_k]_{S,S}^T)^{-1}_{\rho_k(C_{k,j}),\rho_k(C_{k,j})}[\mathbf{d}_k^{(v_s)}]_{\rho_k(C_{k,j})} = \mathbf{0}.$$

Since the matrix multiplying $[\mathbf{d}_k^{(v_s)}]_{\rho_k(C_{k,j})}$ is invertible, we conclude

$$[\mathbf{d}_k^{(v_s)}]_{\rho_k(C_{k,j})} = \mathbf{0}.$$

Again by Assumption 2, no node in $C_{k,j}$ is a parent of v_s, and any node within $C_{k,i}$ that serves as an ancestor of v_s possesses no ancestors in $C_{k,j}$. Hence, no node in $C_{k,j}$ is an ancestor of v_s. Therefore

$$\left[(\mathbf{I} - \mathbf{A}_k)^{-1}\right]_{\rho_k(C_{k,j}),\, \rho_k(v_s)} = 0.$$

This completes the induction.

We have thus shown that no node in $C_{k,j}$ is an ancestor of any node in $C_{k,i}$, contradicting the assumption that $C_{k,j} \in pa_k^{\mathrm{blk}}(C_{k,i})$. Hence the contradiction assumption equation 22 is false, and equation 21 must hold.

Now fix $j^\star$ and b as in equation 21. Since $b \in \rho_k(C_{k,j})$ and the block $C_{k,j}$ is not selected, every basis vector

$$[\mathbf{e}_{\rho_k(t_j)}]_S, \qquad j \in J,$$

vanishes at the coordinate b. Consequently,

$$[\mathbf{\Sigma}_k]_{S,S}(\mathbf{I} - [\mathbf{A}_k]_{S,S}^T)^{-1}[\mathbf{d}_k^{(j^\star)}]_S + [\mathbf{u}_k^{(j^\star)}]_S$$

cannot belong to the span of $\{[\mathbf{e}_{\rho_k(t_j)}]_S : j \in J\}$. Equivalently, after multiplying back by the invertible matrix

$$(\mathbf{I} - [\mathbf{A}_k]_{S,S})^T[\mathbf{\Sigma}_k]_{S,S}^{-1},$$

the vector

$$(\mathbf{I} - [\mathbf{A}_k]_{S,S})^T [\mathbf{\Sigma}_k]_{S,S}^{-1} [\mathbf{u}_k^{(j^\star)}]_S + [\mathbf{d}_k^{(j^\star)}]_S$$

cannot lie in the span of the first r vectors in equation 20. Therefore the whole family in equation 20 has rank strictly larger than r, and hence

$$\dim\big(\mathcal{P}_k(J; W_t)\big) > r.$$

This proves Claim 2. □

J.3 Proof of *Claim 3.*

Proof. Let

$$\hat{C} = \{i_1, \dots, i_r\} \subseteq R_{t-1},$$

and suppose that $\psi(\hat{C})$ is exactly one whole true block, denoted by $C_{k,i}$. Assume moreover that

$$anc_k^{\text{blk}}(C_{k,i}) \subseteq \psi(\mathcal{I}_{t-1}).$$

We show that

$$\dim\big(\mathcal{P}_k(\hat{C}; W_t)\big) = |\hat{C}| = r.$$

Since all true block ancestors of $C_{k,i}$ have already been recovered at the beginning of iteration t, the residual ancestor set appearing in Lemma A.5 reduces exactly to the block $C_{k,i}$ itself. More precisely,

$$S \triangleq \rho_k\Big(Anc_k^{\text{blk}}(C_{k,i}) \setminus \psi(\mathcal{I}_{t-1})\Big) = \rho_k(C_{k,i}).$$

Therefore, by Lemma A.5, the space

$$\mathcal{P}_k(\hat{C}; W_t)$$

is contained in the span of the vectors

$$[\bar{\mathbf{H}}_k]_{S,:}^T (\mathbf{I} - [\mathbf{A}_k]_{S,S})^T [\mathbf{\Sigma}_k]_{S,S}^{-1} [\mathbf{e}_{\rho_k(t_j)}]_S,$$

and

$$[\bar{\mathbf{H}}_k]_{S,:}^T \Big((\mathbf{I} - [\mathbf{A}_k]_{S,S})^T [\mathbf{\Sigma}_k]_{S,S}^{-1} [\mathbf{u}_k^{(j)}]_S + [\mathbf{d}_k^{(j)}]_S\Big), \qquad j \in \hat{C}.$$

We first consider the first family. Since $\hat{C}$ contains exactly all interventions corresponding to the block $C_{k,i}$, the coordinates $\rho_k(t_j)$ for $j \in \hat{C}$ exhaust all indices in $S = \rho_k(C_{k,i})$. Hence the vectors

$$[\mathbf{e}_{\rho_k(t_j)}]_S, \qquad j \in \hat{C},$$

form the standard basis of $\mathbb{R}^{|S|} = \mathbb{R}^r$. Because

$$(\mathbf{I} - [\mathbf{A}_k]_{S,S})^T [\mathbf{\Sigma}_k]_{S,S}^{-1}$$

is invertible, it follows that

$$(\mathbf{I} - [\mathbf{A}_k]_{S,S})^T [\mathbf{\Sigma}_k]_{S,S}^{-1} [\mathbf{e}_{\rho_k(t_j)}]_S, \qquad j \in \hat{C},$$

are r linearly independent vectors. Since $[\bar{\mathbf{H}}_k]_{S,:}$ has full row rank on the residual innovation subspace, the corresponding projected vectors

$$[\bar{\mathbf{H}}_k]_{S,:}^T (\mathbf{I} - [\mathbf{A}_k]_{S,S})^T [\mathbf{\Sigma}_k]_{S,S}^{-1} [\mathbf{e}_{\rho_k(t_j)}]_S, \qquad j \in \hat{C},$$

also span an r-dimensional space. Hence

$$\dim\big(\mathcal{P}_k(\hat{C}; W_t)\big) \geq r.$$

On the other hand, every vector appearing in the second family above belongs to the same r-dimensional ambient coordinate space indexed by $S = \rho_k(C_{k,i})$. Equivalently, before multiplying by $[\bar{\mathbf{H}}_k]_{S,:}^T$, all coefficient vectors lie in $\mathbb{R}^S$, whose dimension is exactly $|S| = r$. Since the first family already forms a basis of this space, each vector

$$(\mathbf{I} - [\mathbf{A}_k]_{S,S})^T [\mathbf{\Sigma}_k]_{S,S}^{-1} [\mathbf{u}_k^{(j)}]_S + [\mathbf{d}_k^{(j)}]_S, \qquad j \in \hat{C},$$

must be a linear combination of the first r vectors

$$(\mathbf{I} - [\mathbf{A}_k]_{S,S})^T [\mathbf{\Sigma}_k]_{S,S}^{-1} [\mathbf{e}_{\rho_k(t_j)}]_S, \qquad j \in \hat{C}.$$

After left multiplication by $[\bar{\mathbf{H}}_k]_{S,:}^T$, the same linear dependence relation is preserved. Therefore, the entire projected family still spans a space of dimension at most r. Thus

$$\dim\big(\mathcal{P}_k(\hat{C}; W_t)\big) \le r.$$

Combining the lower and upper bounds yields

$$\dim\big(\mathcal{P}_k(\hat{C}; W_t)\big) = r.$$

This proves Claim 3. □

J.4 Proof of *Claim 4.*

Proof. Let $\hat{C}$ be the block identified in Step 1 of the current iteration, and let $C_t = \psi(\hat{C})$ be the corresponding true block. Let

$$U_{\hat{C}} \triangleq \operatorname{row}\Big(\{\mathbf{\Theta}_k^{(j)} - \mathbf{\Theta}_k^{(0)}\}_{j \in \hat{C}}\Big).$$

Fix a block $B \in \mathcal{I}_{t-1}$ such that

$$B \notin Anc_k^{\mathrm{blk}}(C_t).$$

Recall that Algorithm 1 defines

$$W_t = \langle \hat{\mathbf{q}}_{k,i} : i \in \mathcal{I}_{t-1} \rangle, \qquad W_{-B} = \langle \hat{\mathbf{q}}_{k,i} : i \in \mathcal{I}_{t-1} \backslash B \rangle.$$

We will show that

$$\dim\big(\mathcal{P}_k(\hat{C}; W_{-B})\big) = |\hat{C}|.$$

By the induction hypothesis (IH2), the recovered block row spaces in $\hat{\mathbf{Q}}_k$ coincide with the corresponding true block row spaces of $\mathbf{Q}_k$, up to the intrinsic permutation and blockwise invertible transformation ambiguity. Hence, for the purpose of dimension counting, we may identify W_t and W_{-B} with the corresponding spans of true block rows of $\mathbf{Q}_k$.

Now define

$$V_{anc} \triangleq \Big\langle [\mathbf{Q}_k]_{\rho_k(i),:} \,\Big|\, i \in C',\ C' \in Anc_k^{\mathrm{blk}}(C_t) \Big\rangle.$$

By Lemma 2, the perturbation row space generated by the interventions in $\hat{C}$ satisfies

$$U_{\hat{C}} \subseteq V_{anc}.$$

Consider the orthogonal projection maps

$$T_t : U_{\hat{C}} \to W_t^{\perp}, \qquad x \mapsto \Pi_{W_t^{\perp}}(x),$$

and

$$T_{-B} : U_{\hat{C}} \to W_{-B}^{\perp}, \qquad x \mapsto \Pi_{W_{-B}^{\perp}}(x).$$

By definition,

$$\mathcal{P}_k(\hat{C}; W_t) = \mathrm{Im}(T_t), \qquad \mathcal{P}_k(\hat{C}; W_{-B}) = \mathrm{Im}(T_{-B}).$$

Therefore,

$$\dim\big(\mathcal{P}_k(\hat{C}; W_t)\big) = \dim(\mathrm{Im}(T_t)), \qquad \dim\big(\mathcal{P}_k(\hat{C}; W_{-B})\big) = \dim(\mathrm{Im}(T_{-B})).$$

We next identify the kernels of these two maps. Since orthogonal projection onto $W^{\perp}$ vanishes exactly on W, we have

$$\ker(T_t) = U_{\hat{C}} \cap W_t, \qquad \ker(T_{-B}) = U_{\hat{C}} \cap W_{-B}.$$

Thus, by the rank-nullity theorem, it is enough to prove

$$U_{\hat{C}} \cap W_t = U_{\hat{C}} \cap W_{-B}.$$

To see this, note first that $W_{-B} \subseteq W_t$, so it suffices to show that removing the block B does not remove any direction lying in $U_{\hat{C}}$. Since $B \notin Anc_k^{\mathrm{blk}}(C_t)$, the block row space associated with B does not enter the ancestor-generated subspace V_{anc} for C_t. Hence

$$W_t \cap V_{anc} = W_{-B} \cap V_{anc}.$$

Because $U_{\hat{C}} \subseteq V_{anc}$, it follows that

$$U_{\hat{C}} \cap W_t = U_{\hat{C}} \cap (W_t \cap V_{anc}) = U_{\hat{C}} \cap (W_{-B} \cap V_{anc}) = U_{\hat{C}} \cap W_{-B}.$$

Therefore,

$$\ker(T_t) = \ker(T_{-B}).$$

Applying rank-nullity to T_t and T_{-B} gives

$$\dim(\mathrm{Im}(T_t)) = \dim(U_{\hat{C}}) - \dim(\ker T_t) = \dim(U_{\hat{C}}) - \dim(\ker T_{-B}) = \dim(\mathrm{Im}(T_{-B})).$$

Hence

$$\dim\big(\mathcal{P}_k(\hat{C}; W_t)\big) = \dim\big(\mathcal{P}_k(\hat{C}; W_{-B})\big).$$

Finally, by Claim 3,

$$\dim\big(\mathcal{P}_k(\hat{C}; W_t)\big) = |\hat{C}|.$$

Therefore,

$$\dim\big(\mathcal{P}_k(\hat{C}; W_{-B})\big) = |\hat{C}|.$$

This proves Claim 4. □

K Proof of Corollary 1

Corollary A.1 (Corollary 1 in the main text)**.** *Assume Assumptions 1 and 2 hold. Let $\hat{\mathbf{Q}}_k$ be the output of Algorithm 2, and define*

$$\hat{Z}_k = \hat{\mathbf{Q}}_k X_k.$$

For each true block $C \in \mathcal{C}_k$, let $\hat{C} = \psi^{-1}(C)$ denote the corresponding estimated block. Then, for every $i \in \hat{C}$, the estimated latent variable $\hat{Z}_{k,(i)}$ lies in the span of the true latent variables indexed by the block ancestors of C, namely

$$\hat{Z}_{k,(i)} \in \mathrm{span}\{Z_{(j)} : j \in C', C' \in Anc_k^{\mathrm{blk}}(C)\}.$$

Thus, up to the relabeling induced by ψ, each estimated latent variable is identified up to block-level ancestor mixing.

Proof. By Theorem 1, there exist a block-diagonal invertible matrix $\mathbf{S}_k$ and a block permutation matrix $\mathbf{P}_{\sigma,k}$ such that

$$\hat{\mathbf{Q}}_k = \mathbf{S}_k \mathbf{P}_{\sigma,k} \mathbf{Q}_k.$$

Therefore,

$$\hat{Z}_k = \hat{\mathbf{Q}}_k X_k = \mathbf{S}_k \mathbf{P}_{\sigma,k} \mathbf{Q}_k X_k.$$

Next, recall that the block-RQ decomposition of $\mathbf{H}_k$ is

$$\mathbf{H}_k = \mathbf{R}_k \mathbf{Q}_k.$$

Since $\mathbf{H}_k X_k = Z_k$, we have

$$\mathbf{R}_k \mathbf{Q}_k X_k = Z_k.$$

Because $\mathbf{R}_k$ is invertible, it follows that

$$\mathbf{Q}_k X_k = \mathbf{R}_k^{-1} Z_k.$$

Substituting this into the previous display yields

$$\hat{Z}_k = \mathbf{S}_k \mathbf{P}_{\sigma,k} \mathbf{R}_k^{-1} Z_k.$$

Now fix a block $C_{k,i}$. By Lemma A.6,

$$[\mathbf{R}_k^{-1}]_{C_{k,i},C_{k,j}} = 0 \qquad \text{whenever } C_{k,j} \notin Anc_k^{\mathrm{blk}}(C_{k,i}).$$

Hence the block $(\mathbf{R}_k^{-1} Z_k)_{C_{k,i}}$ depends only on the blocks $Z_{k,C_{k,j}}$ with

$$C_{k,j} \in Anc_k^{\mathrm{blk}}(C_{k,i}).$$

Equivalently, $(\mathbf{R}_k^{-1} Z_k)_{C_{k,i}}$ is a linear combination of the latent blocks associated with the block ancestors of $C_{k,i}$, including $C_{k,i}$ itself.

Finally, left multiplication by the block-diagonal invertible matrix $\mathbf{S}_k$ only applies an invertible linear transformation within each recovered block, while $\mathbf{P}_{\sigma,k}$ only permutes the block order. Therefore, neither operation changes the ancestor-restricted block dependence pattern; they only change the within-block coordinates and the block ordering. It follows that each recovered block $\hat{Z}_{k,(i)}$ is a linear combination of the true latent blocks corresponding to the block ancestors of its matched true block (including itself), exactly as claimed. □

L Proof of Theorem 2

Theorem A.2 (Theorem 2 in the main text)**.** *Assume Assumptions 1–3 hold. Let* $\{(\hat{\mathcal{C}}_k, \hat{\prec}_k^{\mathrm{blk}})\}_{k=1}^K$ *be the outputs of Algorithm 2 and the input to Algorithm 3. Let* $\hat{\prec}$ *be the output of Algorithm 3. Then* $(\mathcal{V}, \hat{\prec})$ *recovers* $(\mathcal{V}, \prec)$ *up to relabeling under the bijection* $\psi : j \mapsto t_j$.

Proof. We prove that, under the bijection $\psi : j \mapsto t_j$, the output $\hat{\prec}$ of Algorithm 3 coincides exactly with the true global ancestral relation $\prec$. It suffices to show the two inclusions

$$\prec \subseteq \hat{\prec}, \qquad \hat{\prec} \subseteq \prec .$$

We first prove $\prec \subseteq \hat{\prec}$. Take any two global nodes $u, v \in \mathcal{V}$ such that

$$u \prec v.$$

If $u = v$, then the claim is trivial. Assume $u \neq v$. Since $\prec$ is the ancestral relation of the global DAG, there exists a directed path from u to v. Equivalently, there exists a sequence of nodes

$$u = v_0 \to v_1 \to \cdots \to v_m = v$$

such that each edge $v_{r-1} \to v_r$ is a cover edge in the Hasse diagram of $\prec$, namely there is no node w satisfying

$$v_{r-1} \prec w \prec v_r.$$

Fix any such cover edge $v_{r-1} \to v_r$. By Assumption 3, there exists some client k such that both v_{r-1} and v_r appear as singleton blocks in the client-specific block structure, and moreover

$$\{v_{r-1}\} \in pa_k^{\mathrm{blk}}(\{v_r\}).$$

By Theorem 1, Algorithm 2 recovers the client-level block structure and block-level ancestral relation correctly. Hence its output satisfies

$$\{v_{r-1}\} \in \hat{pa}_k^{\mathrm{blk}}(\{v_r\}).$$

Therefore, Algorithm 3 adds

$$v_{r-1} \tilde{\prec} v_r$$

to $\tilde{\prec}$ at Lines 2–3. Since this holds for every cover edge along the above path, we obtain

$$v_0 \tilde{\prec} v_1, \quad v_1 \tilde{\prec} v_2, \quad \ldots, \quad v_{m-1} \tilde{\prec} v_m.$$

Algorithm 3 then defines $\hat{\prec}$ as the transitive closure of $\tilde{\prec}$. Hence

$$u = v_0 \hat{\prec} v_1 \hat{\prec} \cdots \hat{\prec} v_m = v,$$

which implies

$$u \hat{\prec} v.$$

Since $u \prec v$ was arbitrary, we conclude that

$$\prec \subseteq \hat{\prec}.$$

Next we prove $\hat{\prec} \subseteq \prec$. Take any two nodes $u, v \in \mathcal{V}$ such that

$$u \hat{\prec} v.$$

Since $\hat{\prec}$ is the transitive closure of $\tilde{\prec}$, there exists a sequence of nodes

$$u = v_0,\ v_1,\ \ldots,\ v_m = v$$

such that

$$v_{r-1} \tilde{\prec} v_r, \qquad r = 1, \ldots, m.$$

It therefore suffices to show that every relation inserted into $\tilde{\prec}$ is a true global ancestral relation.

Fix any pair $a, b \in \mathcal{V}$ such that

$$a \tilde{\prec} b.$$

By the construction of Algorithm 3, there exists some client k such that

$$\{a\}, \{b\} \in \hat{\mathcal{C}}_k, \qquad \{a\} \in \hat{pa}_k^{\mathrm{blk}}(\{b\}).$$

Again by Theorem 1, the estimated client-level structure is correct, so in the true client-specific structure we also have

$$\{a\} \in pa_k^{\mathrm{blk}}(\{b\}).$$

Since both blocks are singletons, Proposition 2 implies that this block-level parent relation corresponds to a genuine node-level parent-child relation in client k. In particular,

$$a \prec_k b,$$

where $\prec_k$ denotes the node-level ancestral relation in the induced graph for client k. On the other hand, the proof of Proposition 1 shows that $\prec_k$ is exactly the restriction of the global ancestral relation $\prec$ to O_k. Therefore

$$a \prec_k b \quad \Longrightarrow \quad a \prec b.$$

This proves that

$$\tilde{\prec} \subseteq \prec .$$

Since $\prec$ is transitive and $\hat{\prec}$ is the transitive closure of $\tilde{\prec}$, it follows immediately that

$$\hat{\prec} \subseteq \prec .$$

Combining the two inclusions yields

$$\hat{\prec} = \prec .$$

Therefore, under the bijection $\psi : j \mapsto t_j$, the recovered structure $(\mathcal{V}, \hat{\prec})$ coincides exactly with the true global ancestral structure $(\mathcal{V}, \prec)$, up to relabeling. □

M AUXILIARY LEMMAS

Lemma A.3. *Let $X = \mathbf{G}Z$, where $\mathbf{G} \in \mathbb{R}^{n \times p}$ has full column rank. Assume that $Cov(Z)$ is nonsingular, and define*

$$\mathbf{K} \triangleq Cov(Z)^{-1}, \qquad \mathbf{H} \triangleq \mathbf{G}^{\dagger}.$$

Then

$$Cov(X)^{\dagger} = \mathbf{H}^T \mathbf{K} \mathbf{H}.$$

Proof. Since $X = \mathbf{G}Z$, its covariance matrix is

$$Cov(X) = \mathbf{G}\, Cov(Z)\, \mathbf{G}^T.$$

Because $\mathbf{G}$ has full column rank, $\mathbf{G}^T$ has full row rank. Moreover, $Cov(Z)$ is invertible by assumption. According to Greville (1966), we can apply the standard pseudoinverse identity for a matrix product to obtain

$$Cov(X)^{\dagger} = \left(G\, Cov(Z)\, G^T\right)^{\dagger} = (G^T)^{\dagger} Cov(Z)^{\dagger} G^{\dagger}.$$

Using $Cov(Z)^{\dagger} = Cov(Z)^{-1} = \mathbf{K}$ and $(\mathbf{G}^T)^{\dagger} = (\mathbf{G}^{\dagger})^T = \mathbf{H}^T$, we conclude that

$$Cov(X)^{\dagger} = \mathbf{H}^T \mathbf{K} \mathbf{H}.$$

□

Lemma A.4 (Support of $\mathbf{d}_k^{(j)}$ and $\mathbf{u}_k^{(j)}$)**.** *Under the setup of Proposition 1, for each client $k \in [K]$ and intervention $j \in \mathcal{J}_k$ targeting $t_j \in O_k$, let $r = \rho_k(t_j)$. Then the vectors $\mathbf{d}_k^{(j)}$ and $\mathbf{u}_k^{(j)}$ defined in Assumption 2 satisfy*

$$\operatorname{supp}(\mathbf{d}_k^{(j)}) \subseteq \rho_k\big(pa_k(t_j)\big), \qquad \operatorname{supp}(\mathbf{u}_k^{(j)}) \subseteq \rho_k\big(sp_k(t_j)\big).$$

Equivalently, for any $v \in O_k$,

$$v \notin pa_k(t_j) \implies [\mathbf{d}_k^{(j)}]_{\rho_k(v)} = 0, \qquad v \notin sp_k(t_j) \implies [\mathbf{u}_k^{(j)}]_{\rho_k(v)} = 0.$$

Proof. Fix $k \in [K]$ and $j \in \mathcal{J}_k$. For notational simplicity, write

$$O = O_k, \qquad M = M_k, \qquad t = t_j, \qquad r = \rho_k(t),$$

and let

$$\mathbf{B} \coloneqq (\mathbf{I} - [\mathbf{A}]_{M,M})^{-1}.$$

Since intervention j acts only on the structural equation of node t in the global latent SEM, after partitioning according to (O, M), the only changed blocks of the structural matrix are the r-th row blocks in $[\,\cdot\,]_{O,O}$ and $[\,\cdot\,]_{O,M}$. More precisely,

$$[\widetilde{\mathbf{A}}^{(j)}]_{O,O} = [\mathbf{A}]_{O,O} + \mathbf{e}_r[\Delta\mathbf{a}^{(j)}]_O^\top, \qquad [\widetilde{\mathbf{A}}^{(j)}]_{O,M} = [\mathbf{A}]_{O,M} + \mathbf{e}_r[\Delta\mathbf{a}^{(j)}]_M^\top,$$

whereas

$$[\widetilde{\mathbf{A}}^{(j)}]_{M,O} = [\mathbf{A}]_{M,O}, \qquad [\widetilde{\mathbf{A}}^{(j)}]_{M,M} = [\mathbf{A}]_{M,M}.$$

By the expression of the induced client-specific SEM in Proposition 1,

$$\mathbf{A}_k = [\mathbf{A}]_{O,O} + [\mathbf{A}]_{O,M}(\mathbf{I} - [\mathbf{A}]_{M,M})^{-1}[\mathbf{A}]_{M,O},$$

and similarly,

$$\mathbf{A}_k^{(j)} = [\widetilde{\mathbf{A}}^{(j)}]_{O,O} + [\widetilde{\mathbf{A}}^{(j)}]_{O,M}(\mathbf{I} - [\widetilde{\mathbf{A}}^{(j)}]_{M,M})^{-1}[\widetilde{\mathbf{A}}^{(j)}]_{M,O}.$$

Substituting the above block relations yields

$$\mathbf{A}_k^{(j)} = \mathbf{A}_k + \mathbf{e}_r\Big([\Delta\mathbf{a}^{(j)}]_O^\top + [\Delta\mathbf{a}^{(j)}]_M^\top\mathbf{B}[\mathbf{A}]_{M,O}\Big).$$

Hence

$$\mathbf{A}_k^{(j)} = \mathbf{A}_k + \mathbf{e}_r(\mathbf{d}_k^{(j)})^\top, \qquad (\mathbf{d}_k^{(j)})^\top = [\Delta\mathbf{a}^{(j)}]_O^\top + [\Delta\mathbf{a}^{(j)}]_M^\top\mathbf{B}[\mathbf{A}]_{M,O}.$$

Therefore, $\mathbf{d}_k^{(j)}$ parameterizes the perturbation of the r-th row of $\mathbf{A}_k$, namely the perturbation of the incoming directed coefficients of node t in the induced client graph. By definition of the induced graph $\mathcal{G}_k$, for any $v \in O_k$,

$$v \in pa_k(t) \iff [\mathbf{A}_k]_{r,\rho_k(v)} \neq 0.$$

It follows that an admissible nonzero entry of $\mathbf{d}_k^{(j)}$ can only appear at coordinates indexed by $\rho_k(pa_k(t))$, i.e.,

$$\operatorname{supp}(\mathbf{d}_k^{(j)}) \subseteq \rho_k\big(pa_k(t)\big).$$

Next consider $\mathbf{u}_k^{(j)}$. By Proposition 1, the induced noise covariance satisfies

$$\mathbf{\Sigma}_k = \mathbf{I} + [\mathbf{A}]_{O,M}(\mathbf{I} - [\mathbf{A}]_{M,M})^{-1}(\mathbf{I} - [\mathbf{A}]_{M,M})^{-T}[\mathbf{A}]_{O,M}^\top.$$

Under intervention j, only the r-th row of $[\mathbf{A}]_{O,M}$ is perturbed, while the $M \times M$ block remains unchanged. Therefore the perturbation of $\mathbf{\Sigma}_k$ has the form

$$\mathbf{\Sigma}_k^{(j)} = \mathbf{\Sigma}_k + \mathbf{e}_r(\mathbf{u}_k^{(j)})^\top + \mathbf{u}_k^{(j)}\mathbf{e}_r^\top + \eta_k^{(j)}\mathbf{e}_r\mathbf{e}_r^\top,$$

for some scalar $\eta_k^{(j)}$, where

$$\mathbf{u}_k^{(j)} = [\mathbf{A}]_{O,M}(\mathbf{I} - [\mathbf{A}]_{M,M})^{-1}(\mathbf{I} - [\mathbf{A}]_{M,M})^{-T}[\Delta\mathbf{a}^{(j)}]_M.$$

Thus $\mathbf{u}_k^{(j)}$ parameterizes the off-diagonal perturbation in the r-th row and column of $\mathbf{\Sigma}_k$, namely the perturbation of the bidirected part incident to node t in the induced mixed graph. By definition of $\mathcal{G}_k$, for any $v \in O_k$,

$$v \in sp_k(t) \quad \iff \quad [\mathbf{\Sigma}_k]_{r,\rho_k(v)} \neq 0 \quad \iff \quad [\mathbf{\Sigma}_k]_{\rho_k(v),r} \neq 0.$$

Hence a nonzero entry of $\mathbf{u}_k^{(j)}$ can only appear at coordinates indexed by $\rho_k(sp_k(t))$, and therefore

$$\operatorname{supp}(\mathbf{u}_k^{(j)}) \subseteq \rho_k\big(sp_k(t)\big).$$

Combining the two parts proves the claim. □

Lemma A.5. *Fix a client $k \in [K]$. Let $\mathcal{I} \subseteq \mathcal{C}_k$ be a collection of blocks that is closed under block ancestors, namely*

$$C \in \mathcal{I},\ C' \in Anc_k^{\mathrm{blk}}(C) \quad \Longrightarrow \quad C' \in \mathcal{I}.$$

Define

$$W \ \triangleq\ \Big\langle [\mathbf{Q}_k]_{\rho_k(i),:} \ \Big|\ i \in \bigcup_{C\in\mathcal{I}} C \Big\rangle.$$

Let $j_1, \dots, j_\ell$ be intervention indices such that

$$C_k(j_r) \notin \mathcal{I}, \qquad r = 1, \dots, \ell,$$

where $C_k(j_r)$ denotes the block in $\mathcal{C}_k$ containing the target t_{j_r}. Define

$$\mathcal{S}_{\mathcal{I}} \ \triangleq\ \Big(\bigcup_{r=1}^{\ell} Anc_k^{\mathrm{blk}}\big(C_k(j_r)\big)\Big) \setminus \mathcal{I}, \qquad S \ \triangleq\ \rho_k\Big(\bigcup_{C\in\mathcal{S}_{\mathcal{I}}} C\Big).$$

For each $s \in S$, let $[\bar{\mathbf{H}}_k]_{s,:}$ be the projection of $[\mathbf{H}_k]_{s,:}$ onto $W^\perp$, and let $[\bar{\mathbf{H}}_k]_{S,:}$ be the matrix formed by stacking these projected rows.

Then

$$\Pi_{W^\perp}\Big(\operatorname{row}\Big(\{\mathbf{\Theta}_k^{(j)} - \mathbf{\Theta}_k^{(0)}\}_{j\in\{j_1,\dots,j_\ell\}}\Big)\Big)$$

is contained in

$$\Big\langle [\bar{\mathbf{H}}_k]_{S,:}^T (\mathbf{I} - [\mathbf{A}_k]_{S,S})^T [\mathbf{\Sigma}_k]_{S,S}^{-1} [\mathbf{e}_{\rho_k(t_j)}]_S,\ [\bar{\mathbf{H}}_k]_{S,:}^T \Big((\mathbf{I} - [\mathbf{A}_k]_{S,S})^T [\mathbf{\Sigma}_k]_{S,S}^{-1} [\mathbf{u}_k^{(j)}]_S + [\mathbf{d}_k^{(j)}]_S\Big) \ \Big|\ j \in \{j_1, \dots, j_\ell\}\Big\rangle.$$

Proof. For each intervention index j, define

$$\boldsymbol{\alpha}_k^{(j)} \ \triangleq\ (\mathbf{I} - \mathbf{A}_k)^T \mathbf{\Sigma}_k^{-1} \mathbf{e}_{\rho_k(t_j)}, \qquad \boldsymbol{\beta}_k^{(j)} \ \triangleq\ (\mathbf{I} - \mathbf{A}_k)^T \mathbf{\Sigma}_k^{-1} \mathbf{u}_k^{(j)} + \mathbf{d}_k^{(j)}.$$

By Lemma 1,

$$\operatorname{row}(\mathbf{\Theta}_k^{(j)} - \mathbf{\Theta}_k^{(0)}) \subseteq \langle \mathbf{H}_k^T \boldsymbol{\alpha}_k^{(j)},\ \mathbf{H}_k^T \boldsymbol{\beta}_k^{(j)} \rangle.$$

Hence

$$\operatorname{row}\Big(\{\mathbf{\Theta}_k^{(j)} - \mathbf{\Theta}_k^{(0)}\}_{j\in\{j_1,\dots,j_\ell\}}\Big) \subseteq \Big\langle \mathbf{H}_k^T \boldsymbol{\alpha}_k^{(j)},\ \mathbf{H}_k^T \boldsymbol{\beta}_k^{(j)} \ \Big|\ j \in \{j_1, \dots, j_\ell\}\Big\rangle.$$

Since orthogonal projection is linear, it suffices to analyze

$$\Pi_{W^\perp}(\mathbf{H}_k^T \boldsymbol{\alpha}_k^{(j)}) \qquad \text{and} \qquad \Pi_{W^\perp}(\mathbf{H}_k^T \boldsymbol{\beta}_k^{(j)})$$

for each $j \in \{j_1, \dots, j_\ell\}$.

Now fix such a j, and write

$$T_j \ \triangleq\ \rho_k\Big(\bigcup_{C\in Anc_k^{\mathrm{blk}}(C_k(j))} C\Big).$$

By the support characterization of $\mathbf{u}_k^{(j)}$ and $\mathbf{d}_k^{(j)}$ in Lemma A.4, together with the block upper-triangular structure of $\mathbf{A}_k$ and the block-diagonal structure of $\mathbf{\Sigma}_k$, we have

$$\operatorname{supp}(\mathbf{e}_{\rho_k(t_j)}) \subseteq T_j, \qquad \operatorname{supp}(\mathbf{u}_k^{(j)}) \subseteq T_j, \qquad \operatorname{supp}(\mathbf{d}_k^{(j)}) \subseteq T_j,$$

and therefore

$$\mathrm{supp}(\boldsymbol{\alpha}_k^{(j)}) \subseteq T_j, \qquad \mathrm{supp}(\boldsymbol{\beta}_k^{(j)}) \subseteq T_j.$$

Consequently,

$$\mathbf{H}_k^T \boldsymbol{\alpha}_k^{(j)} = [\mathbf{H}_k]_{T_j,:}^T [\boldsymbol{\alpha}_k^{(j)}]_{T_j}, \qquad \mathbf{H}_k^T \boldsymbol{\beta}_k^{(j)} = [\mathbf{H}_k]_{T_j,:}^T [\boldsymbol{\beta}_k^{(j)}]_{T_j}.$$

Thus, rows of $\mathbf{H}_k$ indexed outside T_j do not contribute at all, regardless of whether we project onto $W^\perp$.

Next, because $\mathcal{I}$ is closed under block ancestors and

$$W = \Big\langle [\mathbf{Q}_k]_{\rho_k(i),:} \;\Big|\; i \in \bigcup_{C \in \mathcal{I}} C \Big\rangle,$$

the block-RQ construction implies that every row of $\mathbf{H}_k$ indexed by

$$\rho_k\Big(\bigcup_{C \in \mathcal{I}} C \Big)$$

already lies in W. Hence these rows vanish after projection onto $W^\perp$. On the other hand, for rows indexed by $T_j \setminus \rho_k(\cup_{C \in \mathcal{I}} C)$, the projection may change the row vectors, but only through their projections onto $W^\perp$. Therefore,

$$\Pi_{W^\perp}(\mathbf{H}_k^T \boldsymbol{\alpha}_k^{(j)}) = [\bar{\mathbf{H}}_k]_{S,:}^T [\boldsymbol{\alpha}_k^{(j)}]_S, \qquad \Pi_{W^\perp}(\mathbf{H}_k^T \boldsymbol{\beta}_k^{(j)}) = [\bar{\mathbf{H}}_k]_{S,:}^T [\boldsymbol{\beta}_k^{(j)}]_S.$$

In words, the orthogonal projection acts only on the $\mathbf{H}_k$-side of the generators from Lemma 1, while the coefficient vectors remain unchanged. Moreover, rows indexed by blocks in $\mathcal{I}$ vanish after projection because they already lie in the previously extracted subspace W, whereas rows outside the relevant block ancestors never contribute in the first place by the support characterization above.

It remains to identify the restricted coefficient vectors on S. We claim that

$$[\boldsymbol{\alpha}_k^{(j)}]_S = (\mathbf{I} - [\mathbf{A}_k]_{S,S})^T [\boldsymbol{\Sigma}_k]_{S,S}^{-1} [\mathbf{e}_{\rho_k(t_j)}]_S,$$

and

$$[\boldsymbol{\beta}_k^{(j)}]_S = (\mathbf{I} - [\mathbf{A}_k]_{S,S})^T [\boldsymbol{\Sigma}_k]_{S,S}^{-1} [\mathbf{u}_k^{(j)}]_S + [\mathbf{d}_k^{(j)}]_S.$$

Indeed, by construction of S, the vectors $\mathbf{e}_{\rho_k(t_j)}$, $\mathbf{u}_k^{(j)}$, and $\mathbf{d}_k^{(j)}$ have no support outside the union of blocks in $Anc_k^{\mathrm{blk}}(C_k(j))$, and their coordinates on blocks in $\mathcal{I}$ do not contribute after projection onto $W^\perp$. Moreover, $\boldsymbol{\Sigma}_k$ is block diagonal, so applying $\boldsymbol{\Sigma}_k^{-1}$ does not mix coordinates across distinct blocks. In addition, $\mathbf{A}_k$ is block upper triangular with respect to the block topological order, so for rows indexed by S, the matrix $(\mathbf{I} - \mathbf{A}_k)^T$ only collects contributions from the same block or its block ancestors. Since S already contains precisely those relevant ancestor blocks outside $\mathcal{I}$, no coordinates from S^c contribute to the restriction on S. Therefore, when evaluating the S-coordinates of $\boldsymbol{\alpha}_k^{(j)}$ and $\boldsymbol{\beta}_k^{(j)}$, one may replace $(\mathbf{I} - \mathbf{A}_k)^T \boldsymbol{\Sigma}_k^{-1}$ by its restriction to S, which yields the two identities above.

Substituting these identities into the previous display yields

$$\Pi_{W^\perp}\Big(\mathrm{row}\Big(\{\boldsymbol{\Theta}_k^{(j)} - \boldsymbol{\Theta}_k^{(0)}\}_{j \in \{j_1,\ldots,j_\ell\}}\Big)\Big)$$

contained in

$$\Big\langle [\bar{\mathbf{H}}_k]_{S,:}^T (\mathbf{I} - [\mathbf{A}_k]_{S,S})^T [\boldsymbol{\Sigma}_k]_{S,S}^{-1} [\mathbf{e}_{\rho_k(t_j)}]_S,\ [\bar{\mathbf{H}}_k]_{S,:}^T \Big((\mathbf{I} - [\mathbf{A}_k]_{S,S})^T [\boldsymbol{\Sigma}_k]_{S,S}^{-1} [\mathbf{u}_k^{(j)}]_S + [\mathbf{d}_k^{(j)}]_S \Big) \;\Big|\; j \in \{j_1, \ldots, j_\ell\} \Big\rangle.$$

This proves the claim. □

Lemma A.6. *Let $\mathbf{R}_k$ be the matrix in the block-RQ decomposition of $\mathbf{H}_k$, where the blocks are ordered according to a block topological order. Suppose that*

$$[\mathbf{R}_k]_{C_{k,i}, C_{k,j}} = 0 \qquad \textit{whenever } C_{k,j} \notin Anc_k^{\mathrm{blk}}(C_{k,i}).$$

Then the inverse matrix $\mathbf{R}_k^{-1}$ satisfies the same block support property:

$$[\mathbf{R}_k^{-1}]_{C_{k,i}, C_{k,j}} = 0 \qquad \textit{whenever } C_{k,j} \notin Anc_k^{\mathrm{blk}}(C_{k,i}).$$

Proof. Write

$$\mathbf{R}_k = \mathbf{D}_k + \mathbf{N}_k,$$

where $\mathbf{D}_k$ is the block-diagonal part of $\mathbf{R}_k$, and $\mathbf{N}_k := \mathbf{R}_k - \mathbf{D}_k$ is the off-diagonal part. By assumption, $[\mathbf{N}_k]_{C_{k,i},C_{k,j}}$ can be nonzero only if $C_{k,j} \in anc_k^{\text{blk}}(C_{k,i})$.

Since $\mathbf{R}_k$ is invertible, each diagonal block of $\mathbf{R}_k$ is invertible, and hence so is $\mathbf{D}_k$. Therefore,

$$\mathbf{R}_k = \mathbf{D}_k(\mathbf{I} + \mathbf{D}_k^{-1}\mathbf{N}_k).$$

Let

$$\mathbf{M}_k := \mathbf{D}_k^{-1}\mathbf{N}_k.$$

Then $\mathbf{M}_k$ has zero diagonal blocks and satisfies

$$[\mathbf{M}_k]_{C_{k,i},C_{k,j}} = 0 \qquad \text{whenever } C_{k,j} \notin anc_k^{\text{blk}}(C_{k,i})\}.$$

Since the blocks are arranged in a block topological order, $\mathbf{M}_k$ is strictly block triangular, and hence nilpotent. In particular, if m_k is the number of blocks, then

$$\mathbf{M}_k^{m_k} = 0.$$

It follows that

$$(\mathbf{I} + \mathbf{M}_k)^{-1} = \mathbf{I} - \mathbf{M}_k + \mathbf{M}_k^2 - \cdots + (-1)^{m_k-1}\mathbf{M}_k^{m_k-1},$$

and thus

$$\mathbf{R}_k^{-1} = (\mathbf{I} + \mathbf{M}_k)^{-1}\mathbf{D}_k^{-1}.$$

It remains to verify that every term in the above expression has the same block support pattern. First, $\mathbf{D}_k^{-1}$ is block diagonal, so its (i,j) block can be nonzero only when $i = j$, equivalently when $C_{k,j} \in Anc_k^{\text{blk}}(C_{k,i})$. Next, $\mathbf{M}_k$ satisfies the same ancestor-restricted support property as above. By transitivity of the block ancestor relation, the product of two matrices with this support property again has the same support property. Hence each power $\mathbf{M}_k^s$ also satisfies

$$[\mathbf{M}_k^s]_{C_{k,i},C_{k,j}} = 0 \qquad \text{whenever } C_{k,j} \notin Anc_k^{\text{blk}}(C_{k,i}).$$

Therefore the finite sum $(\mathbf{I} + \mathbf{M}_k)^{-1}$ satisfies the same property, and so does its product with $\mathbf{D}_k^{-1}$. Consequently,

$$[\mathbf{R}_k^{-1}]_{C_{k,i},C_{k,j}} = 0 \qquad \text{whenever } C_{k,j} \notin Anc_k^{\text{blk}}(C_{k,i}),$$

as claimed. □

N ADDITIONAL FINITE-SAMPLE DETAILS

This appendix collects the implementation details omitted from Section 5, including the finite-sample counterparts of the population algorithms, the full simulation setup, the exact evaluation metrics, and the hyperparameter tuning protocol.

N.1 FINITE-SAMPLE ALGORITHMS

Algorithm 5 is the finite-sample counterpart of Algorithm 1. It replaces exact row-space dimension equalities by a singular-value ratio criterion controlled by the threshold τ.

Algorithm 6 is the finite-sample counterpart of Algorithm 2. It searches for the smallest subset whose projected perturbation matrix is approximately rank-r under the same thresholding rule.

N.2 ADDITIONAL SIMULATION DETAILS

Data generation. We consider three simulation settings with different global causal structures and client-specific missing patterns as follows. Each setting is specified by a global causal coefficient matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$, a collection of used latent node sets $\{O_k\}_{k=1}^K$ with $O_k \subseteq [p]$, and client-specific observation dimensions $\{n_k\}_{k=1}^K$.

Algorithm 5 ID-Block-Ancestor-FS

Input: $\hat{C} \subseteq \mathcal{J}_k$, $\{\hat{\Theta}_k^{(j)}\}_{j\in\{0\}\cup\hat{C}}$, $\mathcal{I}$, $\{\hat{\mathbf{q}}_{k,i}\}_{i\in\mathcal{I}}$
Output: $\{\hat{\mathbf{q}}_{k,i}\}_{i\in\hat{C}}$, $\mathcal{A}$

Initialize $\mathcal{A} \leftarrow \mathcal{I}$
Form

$$\tilde{\mathbf{M}}_k(\hat{C}) \leftarrow \begin{bmatrix} \hat{\mathbf{\Theta}}_k^{(j_1)} - \hat{\mathbf{\Theta}}_k^{(0)} \\ \hat{\mathbf{\Theta}}_k^{(j_2)} - \hat{\mathbf{\Theta}}_k^{(0)} \\ \vdots \\ \hat{\mathbf{\Theta}}_k^{(j_{|\hat{C}|})} - \hat{\mathbf{\Theta}}_k^{(0)} \end{bmatrix}, \qquad \hat{C} = \{j_1, \ldots, j_{|\hat{C}|}\}$$

for each block $B \in \mathcal{I}$ **do**
  Form $\mathbf{W}_{-B}$ by stacking the row vectors $\{\hat{\mathbf{q}}_{k,i}\}_{i\in\mathcal{I}\setminus B}$.
  Compute

$$\mathbf{M}_{-B}(\hat{C}) \leftarrow \tilde{\mathbf{M}}_k(\hat{C})(\mathbf{I} - \mathbf{W}_{-B}^T(\mathbf{W}_{-B}\mathbf{W}_{-B}^T)^{-1}\mathbf{W}_{-B}).$$

  Compute

$$\rho(\mathbf{M}_{-B}(\hat{C})) = \frac{\sum_{i=1}^{|\hat{C}|}\sigma_i(\mathbf{M}_{-B}(\hat{C}))}{\sum_i \sigma_i(\mathbf{M}_{-B}(\hat{C}))},$$

  where $\sigma_i(\cdot)$ denotes the i-th largest singular value.
  if $\rho(\mathbf{M}_{-B}(\hat{C})) > 1 - \tau$ **then**
    Remove B from $\mathcal{A}$
  end if
end for
Form $\mathbf{W}_{\text{anc}}$ by stacking the row vectors $\{\hat{\mathbf{q}}_{k,i}\}_{i\in\mathcal{A}}$.
Compute

$$\mathbf{M} \leftarrow \tilde{\mathbf{M}}_k(\hat{C})(\mathbf{I} - \mathbf{W}_{\text{anc}}^T(\mathbf{W}_{\text{anc}}\mathbf{W}_{\text{anc}}^T)^{-1}\mathbf{W}_{\text{anc}}).$$

Set $\{\hat{\mathbf{q}}_{k,i}\}_{i\in\hat{C}}$ to be the right singular vectors of $\mathbf{M}$ corresponding to its top $|\hat{C}|$ singular values.
return $\{\hat{\mathbf{q}}_{k,i}\}_{i\in\hat{C}}$, $\mathcal{A}$

Setting A. Here $p = 5$ and

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0.7 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1.3 & -0.9 \\ 0 & 0 & 0 & 0 & 0.6 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We take $K = 5$ clients, with $O_1 = \{1, 3, 4, 5\}$, $O_2 = \{1, 2, 4, 5\}$, $O_3 = \{1, 2, 3, 4\}$, $O_4 = \{1, 2, 3\}$, $O_5 = \{2, 3, 5\}$, and observation dimensions $(n_1, n_2, n_3, n_4, n_5) = (8, 6, 7, 6, 5)$.

Setting B. Here $p = 8$ and

$$\mathbf{A} = \begin{pmatrix} 0 & -0.8 & 0.1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.6 & 1.2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We take $K = 4$ clients, with $O_1 = \{2, 3, 4, 7, 8\}$, $O_2 = \{1, 2, 3, 5, 6, 7\}$, $O_3 = \{1, 2, 4, 5, 6, 8\}$, $O_4 = \{1, 3, 4, 5, 7, 8\}$, and observation dimensions $(n_1, n_2, n_3, n_4) = (9, 10, 8, 9)$.

Algorithm 6 ID-Block-Causal-Order-FS

Input: $\{\hat{\Theta}_k^{(j)}\}_{j\in\{0\}\cup\mathcal{J}_k}, p_k$

Output: $\hat{\mathcal{C}}_k, \hat{\preceq}_k^{\text{blk}}, \hat{\mathbf{Q}}_k$

Initialize $\mathcal{I}_0 \leftarrow \emptyset$, $R_0 \leftarrow \mathcal{J}_k$, $\widehat{Q}_k \leftarrow \mathbf{0}_{p_k\times n_k}$, $t \leftarrow 1$
while $R_{t-1} \neq \emptyset$ **do**
  flag $\leftarrow$ True, $r \leftarrow 1$.
  while flag **do**
    Form $\mathbf{W}_t \in \mathbb{R}^{|\mathcal{I}_{t-1}|\times n_k}$ by stacking the row vectors $\{\hat{\mathbf{q}}_{k,i}\}_{i\in\mathcal{I}_{t-1}}$.
    Define $\mathcal{R} = \{C | C \subseteq R_{t-1}, |C| = r\}$.
    for each subset $\tilde{C} \subseteq \mathcal{R}$ with $|\widehat{C}| = r$ **do**
      Form

$$\tilde{\mathbf{M}}_k(\tilde{C}) \leftarrow \begin{bmatrix} \hat{\mathbf{\Theta}}_k^{(j_1)} - \hat{\mathbf{\Theta}}_k^{(0)} \\ \hat{\mathbf{\Theta}}_k^{(j_2)} - \hat{\mathbf{\Theta}}_k^{(0)} \\ \vdots \\ \hat{\mathbf{\Theta}}_k^{(j_r)} - \hat{\mathbf{\Theta}}_k^{(0)} \end{bmatrix}, \qquad \tilde{C} = \{j_1, \ldots, j_r\}$$

    Compute

$$\mathbf{M}_t(\tilde{C}) \leftarrow \tilde{\mathbf{M}}_k(\tilde{C})(\mathbf{I} - \mathbf{W}_t^T(\mathbf{W}_t\mathbf{W}_t^T)^{-1}\mathbf{W}_t).$$

    Compute

$$\rho_r(\mathbf{M}_t(\tilde{C})) = \frac{\sum_{i=1}^r \sigma_i(\mathbf{M}_t(\tilde{C}))}{\sum_i \sigma_i(\mathbf{M}_t(\tilde{C}))},$$

    where $\sigma_i(\cdot)$ denotes the i-th largest singular value.
    end for
    if there exists at least one subset $\tilde{C}$ with $\rho_r(\mathbf{M}_t(\tilde{C})) \geq 1-\tau$ **then**
      Among all such subsets of size r, choose $\hat{C} = \tilde{C}$ which maximizes $\rho_r(\mathbf{M}_t(\hat{C}))$
      Set flag $\leftarrow$ False
    else
      Set $r \leftarrow r+1$
    end if
  end while
  $(\{\hat{\mathbf{q}}_{k,i}\}_{i\in\hat{C}}, \mathcal{A}) \leftarrow$ ID-Block-Ancestor-FS$(\hat{C}, \{\hat{\Theta}_k^{(j)}\}_{j\in\{0\}\cup\hat{C}}, \mathcal{I}_{t-1}, \{\hat{\mathbf{q}}_{k,i}\}_{i\in\mathcal{I}_{t-1}})$
  Add $\hat{C} \hat{\preceq}_k^{\text{blk}} C'$ for any $\tilde{C} \hat{\preceq}_k^{\text{blk}} C', \tilde{C} \in \mathcal{A}$.
  Update $\hat{\mathbf{Q}}_k \leftarrow [\tilde{\mathbf{Q}}_{k,\hat{C}}; \hat{\mathbf{Q}}_k]$ where $\tilde{\mathbf{Q}}_{k,\hat{C}} \in \mathbb{R}^{|\hat{C}|\times n_k}$ is the matrix formed by stacking the row vectors $\{\hat{\mathbf{q}}_{k,i}\}_{i\in\hat{C}}$.
  Update $\mathcal{I}_t \leftarrow \{\hat{C}\} \cup \mathcal{I}_{t-1}$, $R_t \leftarrow R_{t-1} \setminus \hat{C}$, $t \leftarrow t+1$.
end while
Set $\hat{\mathcal{C}}_k \leftarrow \mathcal{I}_{t-1}$
return $\hat{\mathcal{C}}_k, \hat{\preceq}_k^{\text{blk}}, \hat{\mathbf{Q}}_k$.

Setting C. Here $p = 10$ and

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & -0.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0.7 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.2 & 0 & 0 & 0 & 0 & 0 & -0.8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1.1 & 0.9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We take $K = 5$ clients, with $O_1 = \{1, 2, 3, 4, 5, 6, 7, 9\}$, $O_2 = \{3, 5, 6, 7, 8, 10\}$, $O_3 = \{1, 2, 4, 7, 8, 9, 10\}$, $O_4 = \{1, 3, 4, 5, 6, 7, 10\}$, $O_5 = \{1, 2, 3, 6, 8, 9\}$, and observation dimensions $(n_1, n_2, n_3, n_4, n_5) = (12, 10, 9, 11, 8)$.

We consider two types of noise distributions, namely Gaussian noise and Laplace noise. For each network setting and each noise setting, the intervention magnitudes on the intervened entries of $\mathbf{A}$ and $\mathbf{\Sigma}$ are independently drawn from the uniform distribution on $[2, 3]$. By construction, the baseline nonzero entries of $\mathbf{A}$ have magnitudes strictly smaller than the intervention range, which helps separate the observational structural coefficients from the intervention effects. For each client $k \in [K]$, we independently generate a column full-rank matrix $\mathbf{G}_k$ with entries sampled from the uniform distribution on $[-1, 1]$.

Sample sizes. We evaluate the method under finite sample sizes n ranging from 10^3 to 10^6, and also include the population case $n = \infty$, where the estimated precision matrices are replaced by the corresponding population precision matrices.

N.3 Evaluation metrics

For all metrics, for ease of exposition, we first align the row ordering of $\hat{\mathbf{Q}}_k$ with that of $\mathbf{Q}_k$ according to the permutation induced by $\mathbf{P}_\sigma$. This alignment is used only for presentation and does not affect the evaluation result.

Metric E_1: global ancestor recovery. We compare the true partial order $\prec$ and the estimated partial order $\hat{\prec}$ on the node set $\mathcal{V}$, and report their F1-score:

$$E_1 \triangleq \mathrm{F1}(\prec, \hat{\prec}).$$

Metric E_2: block-level subspace error. For each client k and each true block $C \in \mathcal{C}_k$, let $\hat{C} = \psi^{-1}(C)$ be the corresponding estimated block. We compare the row spaces of $[\mathbf{Q}_k]_{\rho_k(C),:}$ and $[\hat{\mathbf{Q}}_k]_{\rho_k(\psi(\hat{C})),:}$ through their projection matrices

$$\mathbf{P}_{k,C} = [\mathbf{Q}_k]_{\rho_k(C),:}[\mathbf{Q}_k]^T_{\rho_k(C),:}, \qquad \hat{\mathbf{P}}_{k,C} = [\hat{\mathbf{Q}}_k]_{\rho_k(\psi(\hat{C})),:}[\hat{\mathbf{Q}}_k]^T_{\rho_k(\psi(\hat{C})),:}.$$

We define

$$E_{k,C} \triangleq \frac{\|\hat{\mathbf{P}}_{k,C} - \mathbf{P}_{k,C}\|_F^2}{\sqrt{2}}, \qquad E_2 \triangleq \frac{1}{K}\sum_{k=1}^{K}\frac{1}{|\mathcal{C}_k|}\sum_{C\in\mathcal{C}_k} E_{k,C}.$$

Metric E_3: support leakage error. Since $\hat{Z}_k = \hat{\mathbf{Q}}_k X_k = \hat{\mathbf{Q}}_k \mathbf{G}_k Z_k$, Corollary 1 implies that $[\hat{\mathbf{Q}}_k \mathbf{G}_k]_{\rho_k(i),\rho_k(j)}$ can be nonzero only if the block containing j is a block-level ancestor of the block containing i. Let

$$\hat{\mathbf{N}}_k \triangleq \hat{\mathbf{Q}}_k \mathbf{G}_k, \qquad supp(\hat{\mathbf{N}}_k) \triangleq \{(i, j) : |[\hat{\mathbf{N}}_k]_{\rho_k(i),\rho_k(j)}| > \epsilon\},$$

where $\epsilon = 10^{-2}$. Define the violation set

$$\mathcal{T}_k \triangleq \{(i, j) : C_k(i) \not\preceq^{\mathrm{blk}}_k C_k(j),\ i, j \in O_k\}.$$

Then the leakage error for client k is

$$E_{3,k} \triangleq \frac{|supp(\hat{\mathbf{N}}_k) \cap \mathcal{T}_k|}{|\mathcal{T}_k|}, \qquad E_3 \triangleq \frac{1}{K}\sum_{k=1}^{K} E_{3,k}.$$

Interpretation. The three metrics quantify complementary aspects of recovery: E_1 measures the accuracy of the global ancestral relation, E_2 measures the recovery of client-specific block-associated subspaces, and E_3 measures whether the estimated latent variables respect the block-level ancestral support implied by Corollary 1. All three metrics take values in $[0, 1]$, where larger E_1 indicates better performance, while smaller E_2 and E_3 indicate more accurate recovery.

N.4 HYPERPARAMETER TUNING

The hyperparameter τ is chosen according to the sample size n. For each n, we conduct 10 additional independent runs of data generation and estimation, and search over

$$\tau \in \{0.01, 0.02, \ldots, 0.10\}.$$

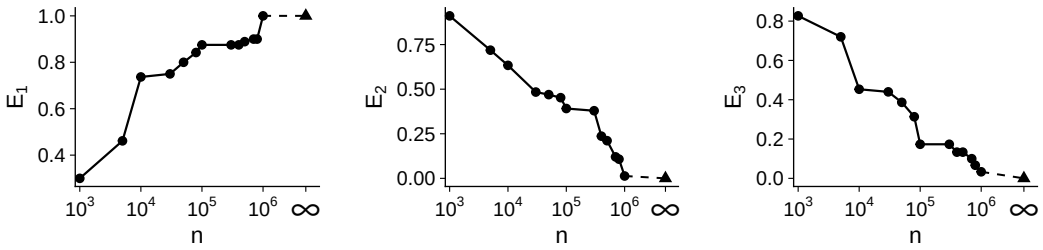
For each candidate value, we compute

$$Score = E_1 - E_2 - E_3,$$

and select the τ that achieves the highest average score over these runs. The resulting value is then used in the final evaluation for that sample size. When such simulation-based calibration is unavailable, we recommend $\tau = 0.05$ as a robust default.

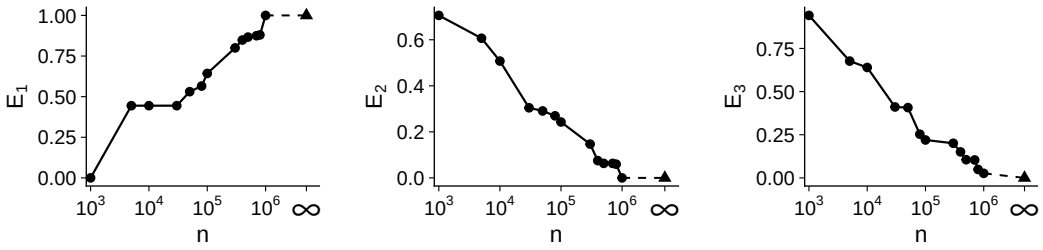
N.5 ADDITIONAL EXPERIMENTAL RESULTS

For the six scenarios induced by three graph settings and two noise settings, we plot the three evaluation metrics as functions of the sample size n in Figure 2 and Figures A.1–A.5. Across all scenarios, recovery improves steadily as n increases. In the population case $n = \infty$, we obtain $E_1 = 1$ and $E_2 = E_3 = 0$, which exactly matches the population-level guarantees in Theorems 1 and 2 together with Corollary 1.



(a) Global ancestral F1. (b) Subspace error of $\{\hat{\mathbf{Q}}_k\}_{k=1}^K$. (c) Support error of $\{\hat{Z}_k\}_{k=1}^K$.

Figure A.1: Finite-sample performance under one graph setting A and laplace noise. As n increases, global ancestral recovery improves while subspace and support errors decrease. The rightmost marker at ∞ corresponds to the population-level precision-matrix input.

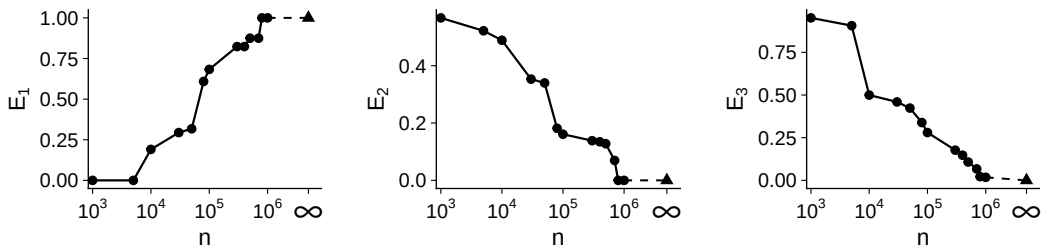


(a) Global ancestral F1. (b) Subspace error of $\{\hat{\mathbf{Q}}_k\}_{k=1}^K$. (c) Support error of $\{\hat{Z}_k\}_{k=1}^K$.

Figure A.2: Finite-sample performance under one graph setting B and gaussian noise. As n increases, global ancestral recovery improves while subspace and support errors decrease. The rightmost marker at ∞ corresponds to the population-level precision-matrix input.

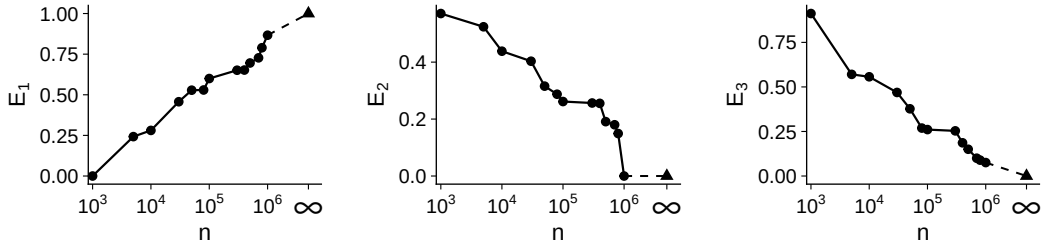
N.6 SENSITIVITY AND SCALABILITY ANALYSIS

We further evaluate sensitivity to the global latent dimension p, the number of clients K, and the client coverage pattern. For each p, we generate an Erdős–Rényi DAG under a fixed topological order, include each admissible edge independently with probability $3/p$, and sample its coefficient from $\mathrm{Unif}[-1, 1]$. Each node is independently included in O_k with probability q, and the observation dimension is set to $n_k = \lceil 1.5p_k \rceil$, where $p_k = |O_k|$. The remaining parameters follow the simulation protocol above. Because p_k, cross-client overlap, and the maximum induced block size are jointly determined by the missingness pattern, we use q as the coverage control: $\mathbb{E}[p_k] = qp$ and $\mathbb{E}[|O_k \cap O_\ell|] = q^2 p$, while larger q generally reduces marginalization-induced block ambiguity.



(a) Global ancestral F1. (b) Subspace error of $\{\hat{\mathbf{Q}}_k\}_{k=1}^K$. (c) Support error of $\{\hat{Z}_k\}_{k=1}^K$.

Figure A.3: Finite-sample performance under one graph setting B and laplace noise. As n increases, global ancestral recovery improves while subspace and support errors decrease. The rightmost marker at ∞ corresponds to the population-level precision-matrix input.



(a) Global ancestral F1. (b) Subspace error of $\{\hat{\mathbf{Q}}_k\}_{k=1}^K$. (c) Support error of $\{\hat{Z}_k\}_{k=1}^K$.

Figure A.4: Finite-sample performance under one graph setting C and gaussian noise. As n increases, global ancestral recovery improves while subspace and support errors decrease. The rightmost marker at ∞ corresponds to the population-level precision-matrix input.

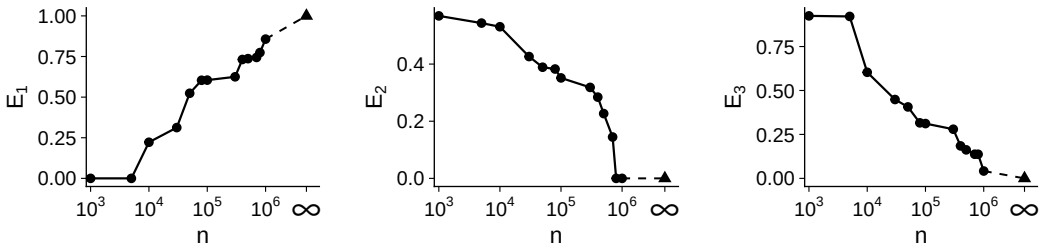
We conduct three controlled experiments: (i) $p \in \{10, 20, 30, 40\}$ with $K = 15$ and $q = 0.6$; (ii) $K \in \{5, 10, 15, 20\}$ with $p = 20$ and $q = 0.6$; and (iii) $q \in \{0.5, 0.6, 0.7, 0.8\}$ with $p = 20$ and $K = 15$. We report E_1, E_2, E_3 for $n \in \{10^4, 5 \times 10^4, 10^5, 5 \times 10^5, 10^6, \infty\}$ in Tables 1–3. Runtime is measured from population precision-matrix inputs and therefore excludes data generation and precision estimation.

Across all settings, recovery improves with the sample size: E_1 increases, while E_2 and E_3 generally decrease. Because the random client subsets do not enforce Assumption 3, global recovery can remain incomplete even at $n = \infty$; for example, the population E_1 decreases from 1.00 to 0.77 as p increases from 10 to 40. Nevertheless, the population client-level errors remain small, with $E_2 \leq 0.05$ and $E_3 \leq 0.02$ across these values of p, confirming that incomplete coverage primarily affects global assembly rather than client-level recovery.

Increasing either the number of clients or their coverage improves global recovery. At $n = \infty$, E_1 increases from 0.34 to 1.00 as K increases from 5 to 20, and from 0.82 to 1.00 as q increases from 0.5 to 0.8, consistent with Assumption 3 being satisfied more frequently. Runtime increases from 0.17 to 4.23 seconds as p grows from 10 to 40, but only from 0.25 to 0.49 seconds as K grows from 5 to 20. This agrees with the fact that computation depends more strongly on the client-level dimensions and block search, while its dependence on K is at most linear. Runtime is not necessarily monotone in q: increasing q enlarges p_k and n_k, but generally reduces the induced block sizes and hence the combinatorial block search.

N.7 Semiconductor Manufacturing Case Study

We further evaluated Jigsaw-CRL using proprietary data from an industry-funded semiconductor-manufacturing project. The system comprises eight latent process states spanning surface preparation, thin-film formation, photoresist coating, exposure and focus, mask-profile development, plasma etching, post-etch profile and damage, and post-anneal recovery. Four clients correspond to surface preparation and deposition, lithography, etching, and post-etch thermal processing. Each client observes and locally intervenes on its accessible states; adjacent clients share boundary states whose



(a) Global ancestral F1. (b) Subspace error of $\{\hat{\mathbf{Q}}_k\}_{k=1}^K$. (c) Support error of $\{\hat{Z}_k\}_{k=1}^K$.

Figure A.5: Finite-sample performance under one graph setting C and laplace noise. As n increases, global ancestral recovery improves while subspace and support errors decrease. The rightmost marker at ∞ corresponds to the population-level precision-matrix input.

Table 1: Sensitivity to the global latent dimension p ($K = 15$, $q = 0.6$). Runtime is measured at $n = \infty$.

p	Metric	10^4	5×10^4	10^5	5×10^5	10^6	∞
10	E_1	0.19	0.45	0.56	0.58	0.81	1.00
	E_2	0.32	0.30	0.24	0.14	0.00	0.00
	E_3	0.57	0.39	0.25	0.18	0.07	0.00
	Runtime (s)	–	–	–	–	–	0.17
20	E_1	0.17	0.49	0.50	0.62	0.85	0.93
	E_2	0.45	0.28	0.27	0.18	0.10	0.00
	E_3	0.48	0.38	0.27	0.14	0.04	0.01
	Runtime (s)	–	–	–	–	–	0.44
30	E_1	0.17	0.38	0.47	0.55	0.71	0.82
	E_2	0.49	0.38	0.30	0.11	0.08	0.03
	E_3	0.50	0.42	0.31	0.09	0.05	0.01
	Runtime (s)	–	–	–	–	–	2.67
40	E_1	0.13	0.41	0.45	0.55	0.69	0.77
	E_2	0.38	0.34	0.25	0.20	0.07	0.05
	E_3	0.49	0.37	0.24	0.14	0.06	0.02
	Runtime (s)	–	–	–	–	–	4.23

identities are aligned by standardized process labels. The sensor, metrology, and processing-result measurements available to each client are modeled as client-specific mixtures of its accessible states.

We used one quarter of factory records sampled at one-minute intervals. After excluding maintenance and transient periods and accounting for temporal dependence, approximately 8×10^5 records were retained to estimate precision matrices for the observational and local intervention environments. An expert-validated partial order, based on process-flow specifications, historical design-of-experiments evidence, and domain knowledge, was used only for evaluation. Jigsaw-CRL achieved $E_1 = \mathbf{0.79}$, with $\hat{\prec} \subseteq \prec$: every recovered ancestral relation agreed with the reference, while the remaining error consisted of missed relations. The true mixing matrices and latent subspaces are unavailable, so E_2 and E_3 cannot be evaluated.

Table 2: Sensitivity to the number of clients K ($p = 20$, $q = 0.6$). Runtime is measured at $n = \infty$.

K	Metric	10^4	5×10^4	10^5	5×10^5	10^6	∞
5	E_1	0.13	0.18	0.19	0.24	0.32	0.34
	E_2	0.42	0.34	0.22	0.18	0.10	0.00
	E_3	0.58	0.48	0.34	0.11	0.00	0.00
	Runtime (s)	–	–	–	–	–	0.25
10	E_1	0.15	0.44	0.46	0.55	0.76	0.81
	E_2	0.36	0.28	0.27	0.14	0.10	0.00
	E_3	0.48	0.39	0.28	0.13	0.08	0.00
	Runtime (s)	–	–	–	–	–	0.37
15	E_1	0.17	0.49	0.50	0.62	0.85	0.93
	E_2	0.45	0.28	0.27	0.18	0.10	0.00
	E_3	0.48	0.38	0.27	0.14	0.04	0.01
	Runtime (s)	–	–	–	–	–	0.44
20	E_1	0.18	0.52	0.54	0.67	0.92	1.00
	E_2	0.50	0.29	0.21	0.14	0.05	0.00
	E_3	0.47	0.36	0.25	0.17	0.04	0.00
	Runtime (s)	–	–	–	–	–	0.49

Table 3: Sensitivity to the client coverage probability q ($p = 20$, $K = 15$). Runtime is measured at $n = \infty$.

q	Metric	10^4	5×10^4	10^5	5×10^5	10^6	∞
0.5	E_1	0.12	0.46	0.46	0.57	0.75	0.82
	E_2	0.55	0.36	0.29	0.15	0.08	0.00
	E_3	0.47	0.39	0.18	0.10	0.08	0.00
	Runtime (s)	–	–	–	–	–	0.53
0.6	E_1	0.17	0.49	0.50	0.62	0.85	0.93
	E_2	0.45	0.28	0.27	0.18	0.10	0.00
	E_3	0.48	0.38	0.27	0.14	0.04	0.01
	Runtime (s)	–	–	–	–	–	0.44
0.7	E_1	0.18	0.54	0.54	0.73	0.91	1.00
	E_2	0.44	0.29	0.22	0.14	0.10	0.00
	E_3	0.49	0.36	0.26	0.15	0.00	0.00
	Runtime (s)	–	–	–	–	–	0.82
0.8	E_1	0.17	0.52	0.59	0.69	0.94	1.00
	E_2	0.41	0.32	0.27	0.18	0.10	0.00
	E_3	0.46	0.38	0.27	0.09	0.00	0.00
	Runtime (s)	–	–	–	–	–	0.77