CAUSAL DISCOVERY FROM DIVERSE DATA

Anonymous authors
Paper under double-blind review

ABSTRACT

Causal discovery is increasingly applied across heterogeneous environments whose data-generating processes may differ substantially. Unlike standard multi-domain formulations based on hard or soft interventions, we do not assume that the domains arise from interventions on a common baseline graph: their causal graphs may differ structurally, including through edge additions, deletions, or reversals. We ask a fundamental question: *how much causal knowledge can be learned by jointly exploiting such heterogeneous datasets?* We show that mechanisms that remain stable across domains can induce cross-domain distributional invariances, even when other mechanisms and the underlying causal graphs differ. By systematically leveraging these invariances together with within-domain conditional independences, we identify the distributional constraints that such invariances impose on tuples of domain-specific causal graphs and characterize the resulting inv-Markov equivalence class. In particular, cross-domain invariances can reveal causal orientations that are not identifiable from any domain in isolation. Our algorithm can leverage cross-domain invariances to learn cause-effect relations beyond what is learnable in each domain, even when the underlying causal graphs are arbitrarily different, such as having reversed edges. Building on this characterization, we develop Diverse-FCI, a constraint-based algorithm that learns domain-specific partially oriented graphs while exploiting pairwise cross-domain invariances. Under a faithfulness condition ensuring that the relevant cross-domain dependencies remain detectable after latent projection, we prove its soundness. Synthetic experiments show improved orientation recovery over learning each domain independently, with larger gains as structural heterogeneity increases.

1 INTRODUCTION

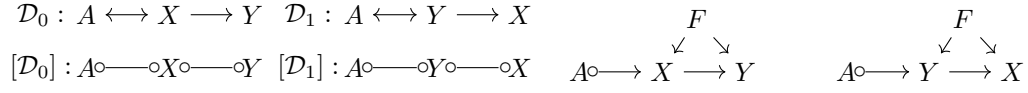
In many scientific and practical applications, data are collected from multiple heterogeneous sources over the same set of variables, such as records from different hospitals (Zech et al., 2018), measurements from different laboratories (Leek et al., 2010), surveys from different regions (Ariely & Davidov, 2012), or observations gathered across time (Guo et al., 2022). Although these datasets concern the same underlying system or closely related systems, they may differ due to changes in populations, environments, measurement processes, or latent conditions (Moreno-Torres et al., 2012). Such differences are rarely generated by controlled experiments or accompanied by a clear intervention interpretation. At the same time, many core mechanisms may remain stable across domains while only a subset vary. This raises a central question: how can we use heterogeneous datasets to distinguish stable causal structure from domain-specific variation?

Observational causal discovery typically identifies only a Markov equivalence class (MEC) of causal graphs under standard assumptions (Spirtes et al., 2001; Zhang, 2008b; Verma & Pearl, 1992; Ali et al., 2009; Meek, 2013b). Stronger identification therefore often relies on additional parametric or functional assumptions, or on interventional data (Shimizu et al., 2006; Eberhardt & Scheines, 2007; Hoyer et al., 2008). This leaves an important gap when multiple heterogeneous domains are available but their differences do not admit a clean interventional interpretation. In such settings, cross-domain mechanism stability may provide causal information beyond learning each domain independently.

Heterogeneous domains can contain causal information unavailable from within-domain observational discovery. Figure 1 illustrates this with two domain-specific acyclic directed mixed graphs (ADMGs), $\mathcal{D}_0 : A \leftrightarrow X \rightarrow Y$ and $\mathcal{D}_1 : A \leftrightarrow Y \rightarrow X$, where the causal direction between X and Y reverses

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.

across domains. Yet within either domain alone, the observational equivalence class leaves the corresponding chain unoriented.



(a) Graph $\mathcal{D}_0$ (b) Graph $\mathcal{D}_1$ (c) Learning of $\mathcal{D}_0$ (d) Learning of $\mathcal{D}_1$

Figure 1: Illustration of learning from cross-domain invariances. Figures (a)–(b) show two domain-specific ADMGs and their within-domain equivalence classes. Figures (c)–(d) show that, by augmenting the domains with an F-node that encodes cross-domain mechanism shifts, the resulting F-adjacencies provide additional orientation information. As a result, the algorithm can recover $X \to Y$ in $\mathcal{D}_0$ and $Y \to X$ in $\mathcal{D}_1$ while they cannot be learned using within-domain independences.

The gain in Figure 1 can be understood directly from the statistical constraints available across the two domains, each represented by an ADMG. Within $\mathcal{D}_0$, the observational CI relation $A \perp\!\!\!\perp Y \mid X$ identifies X as a non-collider on the chain $A - X - Y$, but does not determine the direction of the $X - Y$ edge. Likewise, in $\mathcal{D}_1$, $A \perp\!\!\!\perp X \mid Y$ identifies Y as a non-collider on $A - Y - X$ while leaving the $X - Y$ direction unresolved.

Now suppose that the mechanism of A is stable across the two domains. In this example, its stability is witnessed by the marginal invariance $P^{(0)}(a) = P^{(1)}(a)$. Importantly, this invariance is broken after conditioning on the shifted variables: $P^{(0)}(a \mid x) \neq P^{(1)}(a \mid x)$ and $P^{(0)}(a \mid y) \neq P^{(1)}(a \mid y)$. Thus, although the domain information is marginally separated from A, conditioning on X or Y reveals a dependence. In $\mathcal{D}_0$, the domain-specific skeleton places X between A and this domain information, revealing an arrowhead into X on the $A - X$ edge. Together with the within-domain fact that X is a non-collider on $A - X - Y$, this forces $X \to Y$. In $\mathcal{D}_1$, the same argument applies with X and Y exchanged, yielding $Y \to X$. Hence the same cross-domain invariance information can identify opposite causal directions in the two domains, even though neither direction is identified from within-domain CI information alone.

This example suggests a broader principle: heterogeneous domains may differ in some mechanisms while preserving others, and stable mechanisms leave statistical signatures in the form of cross-domain distributional invariances (Huang et al., 2020). We use F-nodes as a convenient graphical encoding of these pairwise invariance and mechanism-shift patterns (Tian & Pearl, 2001; Kocaoglu et al., 2019). Combining this information with within-domain CI constraints can reveal causal relations that remain unresolved when each domain is analyzed separately (Jaber et al., 2020).

We formalize this idea as causal discovery from diverse data. We observe K domains over a common set of observed variables $\mathbf{V}$, where each domain $i \in [K]$ has its own acyclic causal graph $\mathcal{G}^{(i)} = (\mathbf{V} \cup \mathbf{L}^{(i)}, E^{(i)})$ and induced distribution $P^{(i)}(\mathbf{V})$. We allow domain-specific latent variables and assume no selection bias. Unlike existing multi-domain formulations, we do not require the domains to arise from interventions on a designated baseline graph or to admit a common acyclic causal template. Instead, we characterize the causal information available from jointly exploiting within-domain CI statements and pairwise cross-domain distributional invariances. Our target is therefore different from, rather than uniformly stronger than, standard intervention-based targets: such methods may exploit additional baseline semantics, whereas our framework accommodates domain-specific graph tuples for which such semantics may be unavailable.

We derive cross-domain distributional invariances, which we call *inv-constraints*, by relating heterogeneous domains through a bridge distribution obtained by removing shifted mechanisms. Under an appropriate inv-faithfulness assumption, these constraints impose graphical restrictions on the possible domain graph tuples. We encode them using F-node augmentations (Pearl, 1993), leading to the inv-Markov property, inv-Markov equivalence, and a graphical representation of the resulting equivalence class. Building on this characterization, we develop a constraint-based algorithm that combines within-domain CI tests with cross-domain invariance tests to learn a partially oriented graph for each domain.

Our main contributions are:

- We formulate causal discovery from diverse data, allowing domain-specific causal graphs without assumed intervention semantics, and derive cross-domain inv-constraints.

- We characterize inv-Markov equivalence for domain-specific causal graph tuples using twin augmented MAGs and a compact domain-centric inv-augmented MAG tuple.
- We develop Diverse-FCI, a sound constraint-based algorithm combining within-domain CI tests with cross-domain invariance tests, and evaluate it against multi-domain baselines, showing improved orientation recovery and increasing benefits with structural heterogeneity.

2 Background and Related Work

Causal Bayesian Network (CBN): Given a set of variables $\mathbf{V}$, a directed acyclic graph (DAG) $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ represents causal relations among variables. Equivalently, a structural causal model (SCM) (Pearl, 2009) generates each variable through a structural assignment $V_i := f_i(pa_i, \epsilon_i)$, where pa_i denotes the values of the parents of V_i and ϵ_i is an exogenous noise term. Throughout the paper, we use the term *causal mechanism* to refer to this local data-generating process. In a Markovian CBN, the mechanism is represented by the local conditional distribution $P(v_i \mid pa_i)$. Thus, a mechanism may change across domains either because the parent set changes, or because the structural function or noise distribution changes while the parent set remains the same. A hard intervention $do(X = x)$ sets $X \subseteq \mathbf{V}$ to constants x and removes the causal influence of their parents. A causal graph implies conditional independence (CI) relations through d-separation.

If latent variables are present, we write the causal graph as $\mathcal{D} = (\mathbf{V} \cup \mathbf{L}, \mathbf{E})$, where $\mathbf{V}$ and $\mathbf{L}$ denote observable and latent variables, respectively. Marginalizing latent common causes may induce bidirected edges between observed variables, yielding an Acyclic Directed Mixed Graph (ADMG). Latent variables may also affect the mechanisms of observed variables, so mechanism shifts across domains can be induced by changes involving latent parents or latent confounding.

Ancestral graphs: A Maximal Ancestral Graph (MAG) provides a graphical representation of the observable CI and ancestral relations induced by a causal graph with latent variables (Richardson & Spirtes, 2002; Zhang, 2008a), with separation characterized by m-separation. In a graph $\mathcal{D}$, a triple $\langle X, Y, Z \rangle$ is unshielded if (X, Y) and (Y, Z) are adjacent while X and Z are not. It is an unshielded collider if both edges are oriented into Y. A path $p = \langle X, \ldots, W, Z, Y \rangle$ is a discriminating path for Z if p contains at least three edges, X is not adjacent to Y, and every node between X and Z is a collider on p and a parent of Y. Two MAGs are Markov equivalent if and only if they have the same skeleton, the same unshielded colliders, and agree on the collider status of nodes on common discriminating paths (Ali et al., 2009). A Partial Ancestral Graph (PAG) represents a Markov equivalence class of MAGs. Under faithfulness, FCI learns the corresponding PAG from observational CI statements and is sound and complete (Zhang, 2008b).

Related work: Causal structure learning from observational and interventional distributions has been studied under a variety of assumptions. With latent variables, Kocaoglu et al. (2019) characterized the $\mathcal{I}$-Markov equivalence class for soft interventions using F-node augmentations, Jaber et al. (2020) studied unknown soft interventions through the ψ-MEC, Li et al. (2023) introduced the S-Markov property, and Zhou et al. (2025) considered hard interventions with latent variables. These approaches model domain variation through interventions or related semantics on a common causal system.

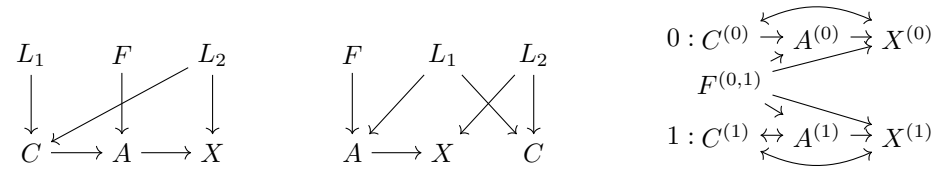
The closest related settings are unknown general interventions and heterogeneous causal discovery. Mascaro & Castelletti (2023) study unknown general interventions that may modify both mechanisms and parent sets, but assume a designated observational baseline and represent other domains relative to it. Our setting instead treats domains symmetrically and exploits pairwise invariances. Appendix E.3 gives a detailed comparison and a concrete identifiability example. Huang et al. (2020) proposed CD-NOD, which allows mechanisms and causal links to vary under a common acyclic causal template, whereas our target is a reference-free tuple of domain-specific latent-variable graphs that need not share such a template. We provide a detailed comparison in Appendix E.2.

Notations: We use upper case letters to denote variables, lower case letters to denote realizations, and bold letters to denote sets. For disjoint sets of variables X, Y, Z, the CI statement that X is independent of Y given Z is denoted by $X \perp\!\!\!\perp Y \mid Z$. In a causal graph $\mathcal{D}$, the corresponding d-separation statement is written as $(X \perp\!\!\!\perp Y \mid Z)_{\mathcal{D}}$. We use $V[\mathcal{D}]$ and $E[\mathcal{D}]$ to denote the node and edge sets of $\mathcal{D}$, respectively. We write $\mathcal{D}_{\overline{X}}$ and $\mathcal{D}_{\underline{X}}$ for the graphs obtained by removing all edges into and out of X, respectively. For $\mathcal{D}_{\overline{X,Y(Z)}}$, the set $Y(Z)$ denotes the subset of Y that are not ancestors of Z in $\mathcal{D}_{\overline{X}}$. In a PAG, a circle mark in an edge $X o\!\!\rightarrow Y$ indicates an endpoint that may be

either an arrowtail or an arrowhead. A star mark in an edge $X *\!\!\rightarrow Y$ is a wildcard and may represent a circle, an arrowhead, or an arrowtail.

3 Combining Datasets Across Domains

We use Pearl's do-calculus (Pearl, 1995; 2009) to relate heterogeneous domains through a common bridge distribution. The required do-calculus rules are recalled in Appendix B. To motivate the mechanism-shift set defined below, consider Figure 2. Only the mechanism of A changes across domains 0 and 1. If $F^{(0,1)}$ were connected only to this structurally changed mechanism, then X would be separable from $F^{(0,1)}$ in each domain, but by different conditioning sets: $\{A, C\}$ in domain 0 and $\{A\}$ in domain 1. Moreover, $\{A, C\}$ does not separate them in domain 1, so no common observed conditioning set works in both domains.



(a) Augmentation in domain 0 (b) Augmentation in domain 1 (c) Pair augmentation

Figure 2: Illustration of our augmentation. Only the mechanism of A changes. Under the naive augmentation $F^{(0,1)} \rightarrow A$, X is separated from $F^{(0,1)}$ by $\{A, C\}$ in domain 0 and by $\{A\}$ in domain 1, but no common conditioning set works in both. Hence $X \in I^{(0,1)}$ despite its unchanged mechanism, and the formal augmented pair graph connects $F^{(0,1)}$ to both copies of A and X.

At the observed level, consider two domains with distributions $P^{(0)}$ and $P^{(1)}$, generated by causal graphs $\mathcal{G}^{(0)}$ and $\mathcal{G}^{(1)}$ over the same observed variables $\mathbf{V}$, possibly with latent variables. A variable $X \in \mathbf{V}$ has a stable mechanism across the two domains if its causal mechanism is unchanged and there exists a common $W \subseteq \mathbf{V} \setminus \{X\}$ that witnesses this stability through the cross-domain invariance $P^{(0)}(x \mid w) = P^{(1)}(x \mid w)$. Otherwise, X belongs to the mechanism-shift set $I^{(0,1)} \subseteq \mathbf{V}$, abbreviated as I when the domain pair is clear. Thus, I contains every variable whose mechanism changes and may additionally contain a variable with an unchanged mechanism for which no common observed conditioning set witnesses stability. Figure 2 illustrates why this enlargement is necessary.

Every variable whose causal mechanism differs across domains belongs to I. Hard-intervening on I therefore removes these differences and yields a common post-intervention model $\mathcal{G}_I := \mathcal{G}^{(0)}_{\overline{I}} = \mathcal{G}^{(1)}_{\overline{I}}$, with corresponding distribution P_I. We use P_I as a bridge between the two domain distributions, leading to the following graphical criterion for cross-domain invariance.

Proposition 3.1. *Let $\mathcal{G}^{(0)}$ and $\mathcal{G}^{(1)}$ be two causal graphs over $\mathbf{V}$, possibly with latent variables, and let I be the mechanism-shift set. Let hard intervention on I yield a common post-intervention graph $\mathcal{G}_I = \mathcal{G}^{(0)}_{\overline{I}} = \mathcal{G}^{(1)}_{\overline{I}}$. Then for any distributions $P^{(0)}$ and $P^{(1)}$ implied by $\mathcal{G}^{(0)}$ and $\mathcal{G}^{(1)}$ respectively, and disjoint sets $Y, W \subseteq \mathbf{V}$ with $Y \cap I = \emptyset$, $P^{(0)}(y \mid w) = P^{(1)}(y \mid w)$, if $(Y \perp\!\!\!\perp W_I \mid W \setminus W_I)_{\mathcal{G}^{(0)}_{\underline{W_I}}} \wedge (Y \perp\!\!\!\perp R \mid W)_{\mathcal{G}^{(0)}_{\overline{R(W)}}} \wedge (Y \perp\!\!\!\perp W_I \mid W \setminus W_I)_{\mathcal{G}^{(1)}_{\underline{W_I}}} \wedge (Y \perp\!\!\!\perp R \mid W)_{\mathcal{G}^{(1)}_{\overline{R(W)}}}$ where $W_I = W \cap I$, $R = I \setminus W$.*

Proposition 3.1 gives graphical sufficient conditions for cross-domain equalities $P^{(0)}(y \mid w) = P^{(1)}(y \mid w)$, which we call inv-constraints. These constrain tuples of domain distributions rather than any single domain. Under the inv-faithfulness assumption introduced in Section 6, observed inv-constraints can conversely restrict the possible causal graph tuples. We next characterize the resulting inv-Markov equivalence class.

4 inv-Markov Equivalence Class

We now characterize the graphical equivalence induced by the inv-constraints. Let the data consist of K domains with distributions $\mathbf{P} = (P^{(i)})_{i=1}^{K}$ over the same observed variables $\mathbf{V}$, generated by

causal graphs $\mathbf{G} = (\mathcal{G}^{(i)})_{i=1}^{K}$, possibly with latent variables. For each unordered pair of domains (i, j), let $I^{(i,j)}$ denote the mechanism-shift set defined as in Section 3. Thus, a variable is treated as stable only when its causal mechanism is shared across the two domains and its stability can be witnessed by a common conditioning set over the observed variables. In particular, under latent confounding, an underlying mechanism may be unchanged but still belong to $I^{(i,j)}$ if no common observed conditioning set can witness its stability.

Let $\mathcal{I} = \{I^{(i,j)} : 1 \le i < j \le K\}$ denote the mechanism-shift collection, and let $\mathcal{G}_I^{(i,j)}$ denote the corresponding bridge graph. We treat pair indices as unordered. Since $\mathcal{I}$ depends on cross-domain mechanism and invariance information, it is not determined by the graph tuple $\mathbf{G}$ alone. The inv-Markov property and inv-Markov equivalence are therefore formally defined with respect to $(\mathbf{G}, \mathcal{I})$. We suppress $\mathcal{I}$ from the notation when it is clear from context.

Definition 4.1. A tuple of distributions $(P^{(i)})_{i=1}^{K}$ generated by $\mathbf{G} = (\mathcal{G}^{(i)})_{i=1}^{K}$, with associated mechanism-shift collection $\mathcal{I} = \{I^{(i,j)} : i < j\}$, satisfies the inv-Markov property if the following hold for disjoint $Y, Z, W \subseteq \mathbf{V}$:

1. For each domain $i \in [K]$, $P^{(i)}(y \mid w, z) = P^{(i)}(y \mid w) \quad \text{if } (Y \perp\!\!\!\perp Z \mid W)_{\mathcal{G}^{(i)}}$,
2. For each unordered pair of domains $i < j$, $P^{(i)}(y \mid w) = P^{(j)}(y \mid w) \quad \text{if } (Y \perp\!\!\!\perp W_I \mid W \setminus W_I)_{\mathcal{G}^{(i)}_{\underline{W_I}}} \wedge (Y \perp\!\!\!\perp R \mid W)_{\mathcal{G}^{(i)}_{\overline{R(W)}}} \wedge (Y \perp\!\!\!\perp W_I \mid W \setminus W_I)_{\mathcal{G}^{(j)}_{\underline{W_I}}} \wedge (Y \perp\!\!\!\perp R \mid W)_{\mathcal{G}^{(j)}_{\overline{R(W)}}}$ where $W_I = W \cap I^{(i,j)}$ and $R = I^{(i,j)} \setminus W$.

We denote the set of all such compatible distribution tuples by $\hat{\mathbf{P}}(\mathbf{G}, \mathcal{I}, \mathbf{V})$.

When $K = 1$, Definition 4.1 reduces to the usual observational Markov property. This leads to the corresponding equivalence notion. Two collections of causal graphs are said to be inv-Markov equivalent if they impose the same constraints on the distribution tuples.

Definition 4.2. Two causal graph tuples with their associated mechanism-shift collections, $(\mathbf{G}_1, \mathcal{I}_1)$ and $(\mathbf{G}_2, \mathcal{I}_2)$, are inv-Markov equivalent if $\hat{\mathbf{P}}(\mathbf{G}_1, \mathcal{I}_1, \mathbf{V}) = \hat{\mathbf{P}}(\mathbf{G}_2, \mathcal{I}_2, \mathbf{V})$.

Definition 4.1 involves separation statements in both the domain graphs and the bridge graphs. We package these constraints into an augmented graph indexed by a pair of domains. Likewise, the augmented graphs below depend on the associated mechanism-shift collection $\mathcal{I}$, which we suppress from their notation.

Definition 4.3 (Augmented Pair Graph)**.** Given a collection of causal graphs $\mathbf{G} = (\mathcal{G}^{(i)})_{i=1}^{K}$ and a pair of domains $i, j \in [K]$ with associated mechanism-shift set $I^{(i,j)}$ as defined above, the augmented pair graph $\mathrm{Aug}^{(i,j)}(\mathbf{G})$ is obtained by taking disjoint copies of the two domain graphs and adding one auxiliary node $F^{(i,j)}$: $\mathrm{Aug}^{(i,j)}(\mathbf{G}) = \left(\mathbf{V}^{(i)} \cup \mathbf{V}^{(j)} \cup \{F^{(i,j)}\}, \mathbf{E}^{(i)} \cup \mathbf{E}^{(j)} \cup \mathcal{E}_F\right)$, where $\mathbf{E}^{(i)} = E[\mathcal{G}^{(i)}]$, $\mathbf{E}^{(j)} = E[\mathcal{G}^{(j)}]$, and $\mathcal{E}_F = \{F^{(i,j)} \to S^{(i)},\ F^{(i,j)} \to S^{(j)} : S \in I^{(i,j)}\}$. If latent variables are present, the construction is understood on the corresponding ADMGs. We omit the collection of causal graphs when it is clear from context.

F-nodes provide a compact observable encoding of the mechanism-shift set, as illustrated in Figure 2c. The adjacencies need not all map to direct structural mechanism changes before latent projection.

Proposition 4.4. *For a pair of causal graphs $\mathcal{G}^{(i)}, \mathcal{G}^{(j)}$ on domains $i, j \in [K]$, construct the augmented pair graph $\mathrm{Aug}^{(i,j)}$. Then for disjoint $Y, Z, W \subseteq \mathbf{V}$, and $k \in \{i, j\}$,*

$$(Y \perp\!\!\!\perp Z|W)_{\mathcal{G}^{(k)}} \iff (Y^{(k)} \perp\!\!\!\perp Z^{(k)}|W^{(k)}, F^{(i,j)})_{\mathrm{Aug}^{(i,j)}} \tag{1}$$

For disjoint $Y, W \subseteq \mathbf{V}$

$$\begin{cases} (Y \perp\!\!\!\perp R|\mathbf{W})_{\mathcal{G}^{(i)}_{\overline{R(W)}}} \\ (Y \perp\!\!\!\perp W_I|W \setminus W_I)_{\mathcal{G}^{(i)}_{\underline{W_I}}} \\ (Y \perp\!\!\!\perp R|\mathbf{W})_{\mathcal{G}^{(j)}_{\overline{R(W)}}} \\ (Y \perp\!\!\!\perp W_I|W \setminus W_I)_{\mathcal{G}^{(j)}_{\underline{W_I}}} \end{cases} \iff \begin{cases} (F^{(i,j)} \perp\!\!\!\perp Y^{(i)}|W^{(i)})_{\mathrm{Aug}^{(i,j)}} \\ (F^{(i,j)} \perp\!\!\!\perp Y^{(j)}|W^{(j)})_{\mathrm{Aug}^{(i,j)}} \end{cases} \tag{2}$$

Different augmented pair graphs may still encode the same constraints. To characterize equivalence, we therefore pass to their corresponding MAGs. Under the definition of the mechanism-shift set, the adjacency of the F-node is already symmetric across the two domain copies: a mechanism-shifted variable is adjacent to F in both copies, whereas a stable variable admits a common conditioning witness and is separable from F in both copies.

Definition 4.5 (Twin Augmented MAG)**.** Given a tuple of causal graphs $\mathbf{G} = (\mathcal{G}^{(i)})_{i=1}^{K}$ and a pair of domains $i, j \in [K]$, the twin augmented MAG is defined as $\mathrm{Twin}^{(i,j)}(\mathbf{G}) := \mathrm{MAG}\Big(\mathrm{Aug}^{(i,j)}(\mathbf{G})\Big)$.

By the definition of the mechanism-shift set, the F-neighborhood in $\mathrm{Twin}^{(i,j)}(\mathbf{G})$ is symmetric across the two domain copies. In particular, if $X \in I^{(i,j)}$, the augmented pair graph contains both $F^{(i,j)} \to X^{(i)}$ and $F^{(i,j)} \to X^{(j)}$, so both adjacencies remain present in the MAG. If $X \notin I^{(i,j)}$, the stability of X admits a common observed conditioning set for distributional invariances. By Proposition 4.4, this corresponds to a paired F-separation in the two copies. Hence $F^{(i,j)}$ is adjacent either to both copies of a variable or to neither.

The term "twin" refers to the two domain copies represented jointly in this MAG. We can now characterize inv-Markov equivalence graphically using the twin augmented MAGs.

Theorem 4.6. *Let $\mathbf{G}_1 = (\mathcal{G}_1^{(i)})_{i=1}^{K}$ and $\mathbf{G}_2 = (\mathcal{G}_2^{(i)})_{i=1}^{K}$ be two tuples of causal graphs over $\mathbf{V}$, with associated mechanism-shift collections $\mathcal{I}_1$ and $\mathcal{I}_2$. Then $(\mathbf{G}_1, \mathcal{I}_1)$ and $(\mathbf{G}_2, \mathcal{I}_2)$ are inv-Markov equivalent if and only if, for every pair of domains $i, j \in [K]$, the twin augmented MAGs $\mathrm{Twin}^{(i,j)}(\mathbf{G}_1)$ and $\mathrm{Twin}^{(i,j)}(\mathbf{G}_2)$ satisfy:*

1. *they have the same skeleton;*
2. *they have the same unshielded colliders;*
3. *if a path p is a discriminating path for a node Y in both graphs, then Y is a collider on p in one graph if and only if it is a collider on p in the other.*

Theorem 4.6 reduces inv-Markov equivalence to ordinary MAG equivalence on the twin augmented graphs. This representation will be used in the next section to combine the domain-specific MAGs and define the graphical object returned by our learning algorithm.

5 INV-AUGMENTED MAG

The twin augmented MAGs characterize the inv-Markov equivalence class pairwise across domains. However, with K domains, this requires inspecting $\binom{K}{2}$ such graphs. Since each domain participates in many pairwise comparisons, it is preferable to aggregate this information into one graph per domain. We therefore introduce a compact representation for the diverse-data setting.

Definition 5.1 (inv-augmented MAG)**.** Given a tuple of causal graphs $\mathbf{G} = (\mathcal{G}^{(i)})_{i=1}^{K}$, with associated mechanism-shift collection $\mathcal{I}$, construct all twin augmented MAGs $\mathrm{Twin}^{(i,j)}(\mathbf{G})$ for unordered pairs $i, j \in [K]$. For each domain $i \in [K]$, define the *inv-augmented MAG* $\mathrm{Aug}^{(i)}(\mathbf{G}) = \big(\mathbf{V}^{(i)} \cup \mathbf{F}^{(i)},\ \mathbf{E}^{(i)} \cup \mathcal{E}^{(i)}\big)$, where $\mathbf{F}^{(i)} = \{F^{(i,j)} : j \neq i\}$, $\mathbf{E}^{(i)} = E[\mathrm{MAG}(\mathcal{G}^{(i)})]$, and $\mathcal{E}^{(i)} = \big\{(F^{(i,j)}, X^{(i)}) : (F^{(i,j)}, X^{(i)}) \in E[\mathrm{Twin}^{(i,j)}],\ j \neq i\big\}$. Equivalently, $\mathrm{Aug}^{(i)}(\mathbf{G})$ is the graph union of the induced subgraphs of $\mathrm{Twin}^{(i,j)}$ on $\mathbf{V}^{(i)} \cup \{F^{(i,j)}\}$ over all $j \neq i$. The *inv-augmented MAG tuple* is $\mathcal{N}(\mathbf{G}) = \big(\mathrm{Aug}^{(i)}(\mathbf{G})\big)_{i=1}^{K}$. We omit the collection of causal graphs when it is clear from the context for simplicity.

Each $\mathrm{Aug}^{(i)}$ preserves all m-separation statements involving the i-th domain that are encoded by the twin augmented MAGs $\mathrm{Twin}^{(i,j)}$, $j \neq i$. Hence, the inv-augmented MAG tuple contains the same pairwise invariance information, but in a domain-centered and more compact form. It is therefore a natural target for learning from multiple heterogeneous datasets.

The next proposition shows that the inv-augmented MAG tuple is equivalent to the family of twin augmented MAGs for characterizing inv-Markov equivalence.

Proposition 5.2. *Let $\mathbf{G}_1 = (\mathcal{G}_1^{(i)})_{i=1}^K$ and $\mathbf{G}_2 = (\mathcal{G}_2^{(i)})_{i=1}^K$ be two tuples of causal graphs over the same observed variables V, with associated mechanism-shift collections $\mathcal{I}_1$ and $\mathcal{I}_2$. Construct their inv-augmented MAG tuples $\mathcal{N}(\mathbf{G}_1)$ and $\mathcal{N}(\mathbf{G}_2)$ according to Definition 5.1. Then $(\mathbf{G}_1, \mathcal{I}_1)$ and $(\mathbf{G}_2, \mathcal{I}_2)$ are inv-Markov equivalent if and only if, for each $i \in [K]$, the graphs $\mathrm{Aug}^{(i)}(\mathbf{G}_1)$ and $\mathrm{Aug}^{(i)}(\mathbf{G}_2)$ satisfy the three conditions in Theorem 4.6.*

Proposition 5.2 shows that the inv-augmented MAG tuple is a unified graphical representation for causal discovery from diverse data. In the next section, we use this to develop a constraint-based learning algorithm that combines within-domain CI tests with cross-domain invariance tests.

6 Learning from Diverse Data

The inv-augmented MAG tuple provides a compact representation of the causal information shared across heterogeneous domains. Learning requires the converse direction from observed distributional constraints to graphical structure. Conceptually, our assumption has two parts. First, in the underlying domain-specific augmented causal model, genuine dependencies induced by mechanism changes are not canceled by special parameter choices. Second, after latent variables are marginalized, F-dependencies induced by mechanism changes and preserved by latent projection remain detectable from the observed domain distributions. Together, these conditions make the F-separations encoded by the augmented graphs testable through cross-domain invariances.

Assumption 6.1 (inv-faithfulness)**.** For a tuple of causal graphs $\mathbf{G} = (\mathcal{G}^{(i)})_{i=1}^K$ with associated mechanism-shift collection $\mathcal{I}$, a tuple of distributions $\left(P^{(i)}\right)_{i=1}^K$ over a set of variables $\mathbf{V}$ is said to be *inv-faithful* to $(\mathbf{G}, \mathcal{I})$ if the converse for each of the conditions in Definition 4.1 holds. When $\mathcal{I}$ is clear from context, we omit it for simplicity and say that the distribution tuple is inv-faithful to $\mathbf{G}$.

The inv-faithfulness assumption strengthens ordinary within-domain faithfulness by additionally requiring observed cross-domain invariances to correspond to the graphical conditions encoded by the F-augmented graphs. Equivalently, dependencies induced by mechanism differences and preserved under latent projection remain detectable in the observed distributions.

Definition 6.2 (inv-augmented graph and inv-augmented PAG)**.** Given a tuple of causal graphs $\mathbf{G} = (\mathcal{G}^{(i)})_{i=1}^K$ with associated mechanism-shift collection $\mathcal{I}$, for each $i \in [K]$, let $\mathcal{M}^{(i)} = \mathrm{Aug}^{(i)}(\mathbf{G})$. Let $\mathbf{M}$ denote the collection of all tuples of inv-augmented MAGs induced by pairs $(\mathbf{G}', \mathcal{I}')$ that are inv-Markov equivalent to $(\mathbf{G}, \mathcal{I})$. For any $i \in [K]$, an *inv-augmented graph*, denoted by $\mathcal{P}^{(i)}(\mathbf{G})$ (or $\mathcal{P}^{(i)}$ when the graph tuple is clear from context), is a graph such that:

1. $\mathcal{P}^{(i)}$ has the same adjacencies as every i-th inv-augmented MAG in $\mathbf{M}$; and
2. every non-circle mark in $\mathcal{P}^{(i)}$ is shared by every i-th inv-augmented MAG in $\mathbf{M}$.

The inv-augmented graph tuple is $\mathcal{L}(\mathbf{G}) = \left(\mathcal{P}^{(i)}\right)_{i=1}^K$. If, in addition, every circle mark in each $\mathcal{P}^{(i)}$ is non-invariant across the equivalence class, then $\mathcal{L}(\mathbf{G})$ is called the *inv-augmented PAG tuple*, and each $\mathcal{P}^{(i)}$ is called an *inv-augmented PAG*. We omit the graph collection when it is clear from context.

Definition 6.2 defines the inv-augmented PAG tuple as the full equivalence-class representation. Under Assumption 6.1, we develop Algorithm 1 to recover a sound partial representation of this target. The algorithm follows the spirit of FCI, and we call it Diverse-FCI. It first learns the domain-specific graph skeletons from within-domain CI tests, then adds the F-nodes using pairwise cross-domain invariance tests, and finally applies orientation rules. We apply the seven standard FCI orientation rules used in the no-selection-bias setting, which we relabel as Rules 1–7 in Appendix C. In addition, Rule 8 orients edges outward from all F-nodes.

The pairwise invariance search and the additional F-node orientation routine are given in Appendix C. Intuitively, the algorithm uses $F^{(i,j)}$-adjacencies and non-adjacencies to encode pairwise mechanism-shift and stability information, and then propagates this information through the standard FCI orientation logic within each domain-centric augmented graph.

Theorem 6.3. *Consider a tuple of distributions $(P^{(i)})_{i\in[K]}$ that are inv-faithful to $(\mathbf{G}, \mathcal{I})$. Then Algorithm 1 is sound.*

Algorithm 1 Main Causal Discovery Algorithm

Input: Distributions from different domains $(P^{(i)})_{i\in[K]}$, observable variables $\mathbf{V}$
Initialize $\mathcal{L}$ as an empty tuple, $SepSet$ as an empty set;
Phase I: Initialize with Complete Graphs;
for i **in** $[K]$ **do**
  Duplicate $\mathbf{V}$ to create $\mathbf{V}^{(i)}$, run Algorithm 2 on i, K to get $\mathbf{F}^{(i)}$ as F nodes;
  Put a circle edge (o—o) between every pair of $X \in \mathbf{V}^{(i)}$ and $Y \in \mathbf{V}^{(i)} \cup \mathbf{F}^{(i)}$;
Phase II: Learning the Skeleton and Separating Sets;
for i **in** $[K]$ **do**
  $E[\mathcal{P}^{(i)}] \leftarrow \emptyset$, $V[\mathcal{P}^{(i)}] \leftarrow \mathbf{V}^{(i)} \cup \mathbf{F}^{(i)}$;
  for j **in** $[K]$ **do**
    Run Algorithm 3 on $i, j, (P^{(i)})_{i\in[K]}, \mathbf{V}, \mathbf{F}^{(i)}$ to get $\mathcal{E}^{(j)}, SepSet^{(j)}$;
    $E[\mathcal{P}^{(i)}] \leftarrow E[\mathcal{P}^{(i)}] \cup \mathcal{E}^{(j)}$, $SepSet \leftarrow SepSet \cup SepSet^{(j)}$;
  $\mathcal{L} \leftarrow \mathcal{L} \cup \mathcal{P}^{(i)}$;
Phase III: Apply Orientation Rules to each $\mathcal{P}^{(i)}, i \in [K]$;
Rule 0: For every unshielded triple $\langle X^{(i)}, Y^{(i)}, Z^{(i)}\rangle$ in $\mathcal{P}^{(i)}, i \in [K]$, orient it as $X^{(i)} *\!\!\rightarrow Y^{(i)} \leftarrow *\!\!Z^{(i)}$ if $Y \notin SepSet(X^{(i)}, Z^{(i)})$.
Rule 8: For any edge adjacent to an F node, orient the edge out of the F node.
First apply Rule 0, then apply 7 FCI rules in Zhang (2008b) together with the following additional rule to each $\mathcal{P}^{(i)}$ until none applies.
Output: inv-augmented graph tuple $\mathcal{L}$

Theorem 6.3 establishes soundness of Diverse-FCI. The procedure need not be complete under latent confounding. Appendix E.4 gives a counterexample and discussion. Because F-adjacencies are pair-specific, Diverse-FCI requires neither a designated reference domain nor a mechanism-shift pattern shared by all domains. There are $\binom{K}{2}$ distinct unordered domain pairs, with the corresponding pairwise invariance information incorporated into each domain-centric graph.

7 Experiments

Table 1: Synthetic comparison for $K = 2$ domains and $m = 12$ cross-domain modifications. Results are domain-averaged mean $\pm$ standard error over 50 independent replications. The best value in each column and panel is shown in bold.

(a) $n = 10$

Method	Skeleton SHD ↓	Partial-graph SHD ↓	Definite endpoint P/R ↑	Runtime (s) ↓
Diverse-FCI	$\mathbf{4.67} \pm 0.29$	$\mathbf{19.83} \pm 0.65$	$0.65/0.39$	0.66 ± 0.01
Per-domain FCI	4.73 ± 0.30	20.80 ± 0.71	$\mathbf{0.66}/0.36$	$\mathbf{0.13} \pm 0.01$
Pooled FCI	17.52 ± 0.80	47.43 ± 1.73	$0.29/\mathbf{0.40}$	1.24 ± 0.11
ψ-FCI	7.45 ± 0.33	24.82 ± 0.69	$0.56/0.29$	0.23 ± 0.00
FCI-JCI	6.92 ± 0.32	24.35 ± 0.74	$0.51/0.38$	1.64 ± 0.02
CD-NOD	6.88 ± 0.32	25.62 ± 1.02	$0.43/0.36$	0.50 ± 0.03

(b) $n = 30$

Method	Skeleton SHD ↓	Partial-graph SHD ↓	Definite endpoint P/R ↑	Runtime (s) ↓
Diverse-FCI	$\mathbf{15.82} \pm 0.63$	$\mathbf{59.12} \pm 1.39$	$\mathbf{0.66}/0.39$	20.00 ± 0.57
Per-domain FCI	16.07 ± 0.62	63.10 ± 1.29	$0.64/0.36$	$\mathbf{0.65} \pm 0.02$
Pooled FCI	40.57 ± 1.86	108.85 ± 3.81	$0.35/0.35$	325.41 ± 126.14
ψ-FCI	20.07 ± 0.73	66.40 ± 1.63	$0.64/0.33$	19.74 ± 0.67
FCI-JCI	19.20 ± 0.72	64.93 ± 1.44	$0.60/0.37$	108.00 ± 2.55
CD-NOD	18.57 ± 0.78	65.52 ± 1.74	$0.54/\mathbf{0.40}$	3.89 ± 0.20

We evaluate Diverse-FCI on controlled synthetic data with known domain-specific causal graphs. We ask whether cross-domain inv-constraints improve recovery over independent per-domain learning and how performance changes with the number of domains. Additional synthetic and real-data experiments are reported in Appendix D.

Setup. We generate multi-domain binary causal models with latent confounding and construct additional domains by applying random structural or mechanism modifications. Unless stated otherwise, we draw 20,000 samples per domain and report results over 50 independent replications. We compare Diverse-FCI with per-domain FCI, pooled FCI, ψ-FCI, FCI-JCI, and CD-NOD. All methods are evaluated on recovery of the domain-specific MAGs over the observed variables using skeleton SHD, partial-graph SHD, definite-endpoint precision and recall, and runtime. Full data-generation details, testing parameters, and evaluation conventions are given in Appendix D.1.

Comparison with baselines. Table 1 reports representative results for $K = 2$ and $m = 12$. Diverse-FCI closely matches per-domain FCI in skeleton recovery but consistently improves partial-graph recovery. For $n = 30$, the partial-graph SHD decreases from 63.10 to 59.12, while definite-endpoint recall increases from 0.36 to 0.39 and precision remains comparable. The same pattern appears for $n = 10$. Thus, the main benefit of the cross-domain inv-constraints is additional orientation information rather than a large change in skeleton recovery. In this structurally heterogeneous setting, Diverse-FCI also obtains the lowest partial-graph SHD among all evaluated methods, although it is substantially slower than learning the domains independently.

Table 2: Scaling with the number of domains for $n = 30$ and $m = 12$. Results are domain-averaged mean $\pm$ standard error over 50 independent replications. Pooled FCI and FCI-JCI were not completed for $K > 2$ under the computational budget.

K	Method	Partial-graph SHD $\downarrow$	Endpoint recall $\uparrow$	Runtime (s) $\downarrow$
2	Diverse-FCI	$\mathbf{59.12 \pm 1.39}$	0.39 ± 0.01	20.00 ± 0.57
	Per-domain FCI	63.10 ± 1.29	0.36 ± 0.01	$\mathbf{0.65 \pm 0.02}$
	Pooled FCI	108.85 ± 3.81	0.35 ± 0.01	325.41 ± 126.14
	ψ-FCI	66.40 ± 1.63	0.33 ± 0.01	19.74 ± 0.67
	FCI-JCI	64.93 ± 1.44	0.37 ± 0.01	108.00 ± 2.55
	CD-NOD	65.52 ± 1.74	$\mathbf{0.40 \pm 0.01}$	3.89 ± 0.20
3	Diverse-FCI	$\mathbf{58.23 \pm 1.16}$	$\mathbf{0.40 \pm 0.01}$	72.57 ± 1.46
	Per-domain FCI	63.01 ± 1.10	0.36 ± 0.01	$\mathbf{0.99 \pm 0.03}$
	ψ-FCI	66.21 ± 1.52	0.34 ± 0.01	38.44 ± 0.36
	CD-NOD	72.11 ± 1.35	0.39 ± 0.01	19.42 ± 1.05
4	Diverse-FCI	$\mathbf{57.34 \pm 1.15}$	$\mathbf{0.41 \pm 0.01}$	147.35 ± 1.92
	Per-domain FCI	62.73 ± 1.09	0.37 ± 0.01	$\mathbf{1.34 \pm 0.04}$
	ψ-FCI	66.55 ± 1.45	0.34 ± 0.01	73.52 ± 0.59
	CD-NOD	79.85 ± 1.48	0.38 ± 0.01	48.45 ± 2.25

Effect of the number of domains. We next vary the number of domains while fixing $n = 30$ and $m = 12$. Table 2 shows that Diverse-FCI benefits from additional domains: its partial-graph SHD decreases from 59.12 at $K = 2$ to 57.34 at $K = 4$, while definite-endpoint recall increases from 0.39 to 0.41. Per-domain FCI cannot exploit the additional cross-domain information, whereas CD-NOD's partial-graph SHD increases substantially as K grows. Runtime increases because the current implementation compares all $\binom{K}{2}$ domain pairs. Additional results on the effect of structural heterogeneity are reported in Appendix D.2.

8 Conclusion

We study causal discovery from diverse domains whose differences need not admit an interventional interpretation. Cross-domain inv-constraints allow us to characterize inv-Markov equivalence using twin augmented MAGs and a compact domain-centric inv-augmented MAG tuple. Building on this characterization, we develop Diverse-FCI, which combines within-domain CI tests with pairwise cross-domain invariance tests and is sound under inv-faithfulness.

Synthetic experiments show that these invariances provide additional orientation information and become increasingly useful as structural heterogeneity grows. Diverse-FCI is not complete in general under latent confounding. Appendix E.4 shows that additional tuple-level orientation rules may be required. Overall, the framework provides a reference-free approach to learning domain-specific causal structure from heterogeneous datasets.

AI USE STATEMENT

Generative AI tools were used in a supporting role during the preparation of this work. They assisted with drafting, language editing, and organization, internal review of the manuscript for clarity, consistency, typographical errors, and missing cross-references, and the routine execution and organization of repeated computational experiments on simulated datasets. For real-world datasets, they assisted with experiment design and causal feature extraction. All the reported experimental results were produced using author-developed code and were independently inspected and verified by the authors.

The research problem, core methodological ideas, theoretical results and proofs, experimental design, implementation of the proposed method, analysis of the results, and scientific conclusions were developed by the authors. All AI-assisted suggestions were critically reviewed and revised as necessary, and the authors take full responsibility for the content of the paper.

REPRODUCIBILITY STATEMENT

We provide the assumptions, definitions, and complete proofs of the theoretical results in the main text and Appendix B. The proposed Diverse-FCI procedure is specified in Algorithm 1, with the pairwise invariance search and auxiliary routines given in Appendix C. Section 7 summarizes the synthetic experimental setup, baselines, and evaluation metrics, while Appendix D provides the full data-generation procedure, conditional-independence tests, hyperparameters, evaluation conventions, and number of independent replications. Appendix D further provides the dataset construction, preprocessing, variable selection, sampling procedures, and experimental settings for the SERGIO, SEER, and MIMIC-IV experiments, together with the computational resources used. These details are provided to facilitate reproduction of the theoretical and experimental results.

REFERENCES

Jayadev Acharya, Arnab Bhattacharyya, Constantinos Daskalakis, and Saravanan Kandasamy. Learning and testing causal models with interventions. *Advances in Neural Information Processing Systems*, 31, 2018.

Raghavendra Addanki, Andrew McGregor, and Cameron Musco. Intervention efficient algorithms for approximate learning of causal graphs. In *Algorithmic Learning Theory*, pp. 151–184. PMLR, 2021.

R. Ayesha Ali, Thomas S. Richardson, and Peter Spirtes. Markov equivalence for ancestral graphs. *The Annals of Statistics*, 37(5B):2808–2837, 2009. doi: 10.1214/08-AOS626.

Gal Ariely and Eldad Davidov. Assessment of measurement equivalence with cross-national and longitudinal surveys in political science. *European Political Science*, 11(3):363–377, 2012.

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

Payam Dibaeinia and Saurabh Sinha. Sergio: a single-cell expression simulator guided by gene regulatory networks. *Cell systems*, 11(3):252–271, 2020.

Kemi M Doll, Alfred Rademaker, and Julie A Sosa. Practical guide to surgical data sets: surveillance, epidemiology, and end results (seer) database. *JAMA surgery*, 153(6):588–589, 2018.

Frederick Eberhardt and Richard Scheines. Interventions and causal inference. *Philosophy of science*, 74(5):981–995, 2007.

Patrick Forré and Joris M Mooij. Constraint-based causal discovery for non-linear structural causal models with cycles and latent confounders. *arXiv preprint arXiv:1807.03024*, 2018.

Tian Gao, Songtao Lu, Junkyu Lee, Elliot Nelson, Debarun Bhattacharjya, Yue Yu, and Miao Liu. Meta-D^{2}AG: Causal graph learning with interventional dynamic data. In *Advances in Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=YFZOIra4OE.

AmirEmad Ghassami, Saber Salehkaleybar, Negar Kiyavash, and Kun Zhang. Learning causal structures using regression invariance. *Advances in Neural Information Processing Systems*, 30, 2017.

Lin Lawrence Guo, Stephen R Pfohl, Jason Fries, Alistair EW Johnson, Jose Posada, Catherine Aftandilian, Nigam Shah, and Lillian Sung. Evaluation of domain generalization and adaptation on improving model robustness to temporal dataset shift in clinical medicine. *Scientific reports*, 12 (1):2726, 2022.

Alain Hauser and Peter Bühlmann. Characterization and greedy learning of interventional markov equivalence classes of directed acyclic graphs. *The Journal of Machine Learning Research*, 13(1): 2409–2464, 2012.

Christina Heinze-Deml, Jonas Peters, and Nicolai Meinshausen. Invariant causal prediction for nonlinear models. *Journal of Causal Inference*, 6(2):20170016, 2018.

Benjamin Herdeanu, Juan Nathaniel, Carla Roesch, Jatan Buch, Gregor Ramien, Johannes Haux, and Pierre Gentine. CausalDynamics: A large-scale benchmark for structural discovery of dynamical causal models. In *Advances in Neural Information Processing Systems*, 2025. URL `https://mlanthology.org/neurips/2025/herdeanu2025neurips-causaldynamics/`.

Patrik Hoyer, Dominik Janzing, Joris M Mooij, Jonas Peters, and Bernhard Schölkopf. Nonlinear causal discovery with additive noise models. *Advances in neural information processing systems*, 21, 2008.

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

Amin Jaber, Murat Kocaoglu, Karthikeyan Shanmugam, and Elias Bareinboim. Causal discovery from soft interventions with unknown targets: Characterization and learning. *Advances in neural information processing systems*, 33:9551–9561, 2020.

Yibo Jiang and Bryon Aragam. Learning nonparametric latent causal graphs with unknown interventions. *Advances in Neural Information Processing Systems*, 36:60468–60513, 2023.

Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Brian Gow, Benjamin Moody, Steven Horng, Leo Anthony Celi, and Roger Mark. MIMIC-IV. *PhysioNet*, July 2024. doi: 10.13026/hxp0-hg59. URL `https://doi.org/10.13026/hxp0-hg59`. Version 3.0.

Nan Rosemary Ke, Olexa Bilaniuk, Anirudh Goyal, Stefan Bauer, Hugo Larochelle, Bernhard Schölkopf, Michael C Mozer, Chris Pal, and Yoshua Bengio. Learning neural causal models from unknown interventions. *arXiv preprint arXiv:1910.01075*, 2019.

Murat Kocaoglu, Amin Jaber, Karthikeyan Shanmugam, and Elias Bareinboim. Characterization and learning of causal graphs with latent variables from soft interventions. *Advances in Neural Information Processing Systems*, 32, 2019.

Jeffrey T Leek, Robert B Scharpf, Héctor Corrada Bravo, David Simcha, Benjamin Langmead, W Evan Johnson, Donald Geman, Keith Baggerly, and Rafael A Irizarry. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*, 11(10): 733–739, 2010.

Adam Li, Amin Jaber, and Elias Bareinboim. Causal discovery from observational and interventional data across multiple environments. *Advances in Neural Information Processing Systems*, 36: 16942–16956, 2023.

Romain Lopez, Jan-Christian Hütter, Jonathan Pritchard, and Aviv Regev. Large-scale differentiable causal discovery of factor graphs. *Advances in Neural Information Processing Systems*, 35: 19290–19303, 2022.

Alessandro Mascaro and Federico Castelletti. Bayesian causal discovery from unknown general interventions. *arXiv preprint arXiv:2312.00509*, 2023.

Christopher Meek. Causal inference and causal explanation with background knowledge. *arXiv preprint arXiv:1302.4972*, 2013a.

Christopher Meek. Strong completeness and faithfulness in bayesian networks. *arXiv preprint arXiv:1302.4973*, 2013b.

Joris M Mooij, Sara Magliacane, and Tom Claassen. Joint causal inference from multiple contexts. *Journal of machine learning research*, 21(99):1–108, 2020.

Jose G Moreno-Torres, Troy Raeder, Rocío Alaiz-Rodríguez, Nitesh V Chawla, and Francisco Herrera. A unifying view on dataset shift in classification. *Pattern recognition*, 45(1):521–530, 2012.

National Cancer Institute. Surveillance, epidemiology, and end results (seer) program. *Cancer Statistics, SEER Data & Software, Registry Operations*, 2018.

J Pearl. Aspects of graphical models connected with causality. *Proceedings of the 49th Session of the international Statistical Institute, Italy*, pp. 399–401, 1993.

Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Ronan Perry, Julius Von Kügelgen, and Bernhard Schölkopf. Causal discovery in heterogeneous environments under the sparse mechanism shift hypothesis. *Advances in Neural Information Processing Systems*, 35:10904–10917, 2022.

Jonas Peters, Peter Bühlmann, and Nicolai Meinshausen. Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(5):947–1012, 2016.

Parjanya Prashant, Ignavier Ng, Kun Zhang, and Biwei Huang. Differentiable causal discovery for latent hierarchical causal models. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=Bp0HBaMNRl.

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

Shohei Shimizu, Patrik O Hoyer, Aapo Hyvärinen, Antti Kerminen, and Michael Jordan. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(10), 2006.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, prediction, and search*. MIT press, 2001.

Peter L Spirtes, Christopher Meek, and Thomas S Richardson. Causal inference in the presence of latent variables and selection bias. *arXiv preprint arXiv:1302.4983*, 2013.

Jin Tian and Judea Pearl. Causal discovery from changes. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pp. 512–521, 2001.

Panagiotis Tigas, Yashas Annadani, Andrew Jesson, Bernhard Schölkopf, Yarin Gal, and Stefan Bauer. Interventions, where and how? experimental design for causal models at scale. *Advances in neural information processing systems*, 35:24130–24143, 2022.

Thomas Verma and Judea Pearl. An algorithm for deciding if a set of observed independencies has a causal explanation. In *Uncertainty in artificial intelligence*, pp. 323–330. Elsevier, 1992.

Yuhao Wang, Liam Solus, Karren Yang, and Caroline Uhler. Permutation-based causal inference algorithms with interventions. *Advances in Neural Information Processing Systems*, 30, 2017.

Menghua Wu, Umesh Padia, Sean H. Murphy, Regina Barzilay, and Tommi Jaakkola. Identifying biological perturbation targets through causal differential networks. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 67537–67561. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/wu25v.html.

Karren Yang, Abigail Katcoff, and Caroline Uhler. Characterizing and learning equivalence classes of causal dags under interventions. In *International Conference on Machine Learning*, pp. 5541–5550. PMLR, 2018.

Binghua Yao and Joris Marten Mooij. σ-maximal ancestral graphs. In *Proceedings of the 41st Conference on Uncertainty in Artificial Intelligence*, volume 286 of *Proceedings of Machine Learning Research*, pp. 4775–4805. PMLR, 2025.

John R Zech, Marcus A Badgeley, Manway Liu, Anthony B Costa, Joseph J Titano, and Eric Karl Oermann. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS medicine*, 15(11):e1002683, 2018.

Jiji Zhang. Causal reasoning with ancestral graphs. *Journal of Machine Learning Research*, 9(7), 2008a.

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, 2008b.

Zihan Zhou, Muhammad Qasim Elahi, and Murat Kocaoglu. Sample efficient bayesian learning of causal graphs from interventions. *Advances in Neural Information Processing Systems*, 37: 9178–9204, 2024.

Zihan Zhou, Muhammad Qasim Elahi, and Murat Kocaoglu. Characterization and learning of causal graphs from hard interventions. *Advances in Neural Information Processing Systems*, 38, 2025.

A Detailed Related Work

Equivalence classes. Causal discovery algorithms use statistical constraints in data to restrict the set of causal diagrams compatible with the underlying data-generating process. In general, these constraints are not sufficient to identify a unique graph. The standard target is therefore an equivalence class (EC) of causal diagrams that agree on the constraints observable from the available distributions. With purely observational data, this equivalence class is the Markov equivalence class (MEC), consisting of graphs that entail the same set of d-separation statements over the observed variables (Spirtes et al., 2001; Verma & Pearl, 1992; Spirtes et al., 2013; Meek, 2013a). Under causal sufficiency, interventional data can refine this class: prior work characterizes the $\mathcal{I}$-MEC induced by a collection of intervention targets (Hauser & Bühlmann, 2012; Yang et al., 2018). Related work studies equivalence under local changes to causal mechanisms (Tian & Pearl, 2001).

When latent variables are present, the equivalence-class structure becomes more subtle. For soft interventions with known targets, prior work characterizes the corresponding $\mathcal{I}$-MEC in the presence of latent confounding (Kocaoglu et al., 2019). When soft-intervention targets are unknown, the corresponding object is the ψ-MEC, an equivalence class over both causal graphs and intervention targets (Jaber et al., 2020). More recently, causal discovery from observational and interventional data across multiple environments has been studied using selection diagrams, leading to the S-MEC (Li et al., 2023). In contrast to these intervention- or selection-based formulations, our setting does not require domains to correspond to interventions or to be represented through a selection diagram. Instead, we consider diverse datasets over a shared observed variable set and aim to recover sound domain-specific partially oriented graphs together with pairwise mechanism-shift information.

Learning from combined datasets. There is a large literature on learning causal structure from multiple datasets, environments, or experiments. Many approaches exploit invariances across environments, but they often rely on causal sufficiency, Markovianity, specific functional restrictions such as linearity, or explicit observational/interventional semantics (Peters et al., 2016; Ghassami et al., 2017; Heinze-Deml et al., 2018; Huang et al., 2020; Perry et al., 2022). Joint Causal Inference (JCI) takes a different approach by pooling data across contexts and augmenting the variable set with context variables, so that structure learning can be performed on a single combined dataset (Mooij et al., 2020). Neural and differentiable approaches have also been proposed for combining observational and interventional data, including settings with unknown intervention targets (Ke et al., 2019; Brouillard et al., 2020; Lopez et al., 2022). These methods can perform well empirically, especially in large-scale or high-dimensional settings, but they typically do not provide the same equivalence-class characterization or soundness guarantees as constraint-based graphical approaches.

Several works study the design or use of interventions for causal discovery. Existing results analyze the sample and intervention complexity required to test whether an unknown causal Bayesian network matches a given model (Acharya et al., 2018), as well as the problem of choosing a minimum-cost intervention set to orient causal or ancestral relations (Addanki et al., 2021). Adaptive experimental design methods choose interventions sequentially to improve graph recovery (Tigas et al., 2022). Score- and permutation-based methods such as GIES (Hauser & Bühlmann, 2012) and IGSP (Wang et al., 2017) learn a single DAG from observational and interventional data, typically under causal sufficiency and known intervention targets. For unknown general interventions, Mascaro & Castelletti (2023) allow both the mechanisms and parent sets of unknown targets to change, characterize the corresponding interventional equivalence classes, and develop Bayesian structure learning for this setting. Bayesian approaches have also been developed for other causally sufficient intervention settings (Zhou et al., 2024). In parallel, representation-learning work studies identifiability of causal representations under unknown interventions and weak distributional assumptions (Jiang & Aragam, 2023).

Recent studies characterize and learn equivalence classes induced by hard interventions with latent variables (Zhou et al., 2025), study dynamic DAG discovery from interventional and non-stationary time-series domains (Gao et al., 2025), and identify biological perturbation targets by comparing observational and interventional datasets through causal differential networks (Wu et al., 2025). Other recent work develops differentiable causal discovery for latent hierarchical causal models (Prashant et al., 2025) and introduces a large-scale benchmark for structural discovery in dynamical causal systems (Herdeanu et al., 2025). These works highlight growing interest in causal discovery beyond a single homogeneous observational dataset. Our contribution is complementary: we study a constraint-

based FCI-style procedure for heterogeneous domains that share observed variables but need not be interpreted as interventions, and we return domain-specific partially oriented graphs together with mechanism-shift candidates.

B PROOFS

B.1 PROOF FOR PROPOSITION 3.1

For completeness, we recall the do-calculus rules used in the proof.

Theorem B.1 (Do-calculus, Theorem 3 in Pearl (1995))**.** *Let $\mathcal{G} = (\mathbf{V} \cup \mathbf{L}, \mathbf{E})$ be a causal graph. Then for any distribution consistent with $\mathcal{G}$:*

$$\textit{Rule 1:} \quad P_{\mathbf{X}}(\mathbf{y} \mid \mathbf{w}, \mathbf{z}) = P_{\mathbf{X}}(\mathbf{y} \mid \mathbf{w}) \qquad \textit{if } (\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{W}, \mathbf{X})_{\mathcal{G}_{\overline{\mathbf{X}}}}, \tag{3}$$

$$\textit{Rule 2:} \quad P_{\mathbf{X},\mathbf{z}}(\mathbf{y} \mid \mathbf{w}) = P_{\mathbf{X}}(\mathbf{y} \mid \mathbf{w}, \mathbf{z}) \qquad \textit{if } (\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{W}, \mathbf{X})_{\mathcal{G}_{\overline{\mathbf{X}},\underline{\mathbf{Z}}}}, \tag{4}$$

$$\textit{Rule 3:} \quad P_{\mathbf{X},\mathbf{z}}(\mathbf{y} \mid \mathbf{w}) = P_{\mathbf{X}}(\mathbf{y} \mid \mathbf{w}) \qquad \textit{if } (\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{W}, \mathbf{X})_{\mathcal{G}_{\overline{\mathbf{X}},\overline{\mathbf{Z}(\mathbf{W})}}}. \tag{5}$$

Here $\mathbf{Z}(\mathbf{W})$ denotes the subset of $\mathbf{Z}$ that is not ancestral to $\mathbf{W}$ in $\mathcal{G}_{\overline{\mathbf{X}}}$.

Proof. To show that $P^{(0)}(y|w) = P^{(1)}(y|w)$, if $(Y \perp\!\!\!\perp W_I \mid W \setminus W_I)_{\mathcal{G}^{(0)}_{\underline{W_I}}} \wedge (Y \perp\!\!\!\perp R \mid W)_{\mathcal{G}^{(0)}_{\overline{R(W)}}} \wedge (Y \perp\!\!\!\perp W_I \mid W \setminus W_I)_{\mathcal{G}^{(1)}_{\underline{W_I}}} \wedge (Y \perp\!\!\!\perp R \mid W)_{\mathcal{G}^{(1)}_{\overline{R(W)}}}$, it is sufficient to show that $P^{(0)}(y|w) = P_I(y|w)$ if $(Y \perp\!\!\!\perp W_I \mid W \setminus W_I)_{\mathcal{G}^{(0)}_{\underline{W_I}}} \wedge (Y \perp\!\!\!\perp R \mid W)_{\mathcal{G}^{(0)}_{\overline{R(W)}}}$ and $P^{(1)}(y|w) = P_I(y|w)$ if $(Y \perp\!\!\!\perp W_I \mid W \setminus W_I)_{\mathcal{G}^{(1)}_{\underline{W_I}}} \wedge (Y \perp\!\!\!\perp R \mid W)_{\mathcal{G}^{(1)}_{\overline{R(W)}}}$. Without loss of generality, we show this holds for $P^{(0)}$.

We first show that $P^{(0)}(y|w) = P^{(0)}(y|w \setminus w_i, do(w_i))$ if $(Y \perp\!\!\!\perp W_I \mid W \setminus W_I)_{\mathcal{G}^{(0)}_{\underline{W_I}}}$. This holds directly from do-calculus Rule 2. Applying Rule 3 gives $P^{(0)}(y|w, do(w_i)) = P^{(0)}(y|w, do(i))$ if $(Y \perp\!\!\!\perp R \mid W)_{\mathcal{G}^{(0)}_{\underline{W_I}\overline{R(W)}}}$. By definition, $P^{(0)}(y \mid w, \mathrm{do}(i)) = P_I(y \mid w)$. Accordingly, we have $P^{(0)}(y|w) = P_I(y|w)$ if $(Y \perp\!\!\!\perp W_I \mid W \setminus W_I)_{\mathcal{G}^{(0)}_{\underline{W_I}}} \wedge (Y \perp\!\!\!\perp R \mid W)_{\mathcal{G}^{(0)}_{\underline{W_I}\overline{R(W)}}}$. We then show that this is equivalent to the condition in the statement with the following lemma.

Lemma B.2. *For a causal graph $\mathcal{G}$ on observable $\mathbf{V}$ and $Y, W, I \subseteq \mathbf{V}$, Y and W disjoint, $(Y \perp\!\!\!\perp W_I \mid W \setminus W_I)_{\mathcal{G}_{\underline{W_I}}} \wedge (Y \perp\!\!\!\perp R \mid W)_{\mathcal{G}_{\underline{W_I}\overline{R(W)}}}$ is equivalent to $(Y \perp\!\!\!\perp W_I \mid W \setminus W_I)_{\mathcal{G}_{\underline{W_I}}} \wedge (Y \perp\!\!\!\perp R \mid W)_{\mathcal{G}_{\overline{R(W)}}}$.*

Proof. ($\Leftarrow$) Since $\mathcal{G}_{\overline{R(W)}}$ is a supergraph of $\mathcal{G}_{\underline{W_I}}$, any d-separation statement that holds in $\mathcal{G}_{\overline{R(W)}}$ should also hold in $\mathcal{G}_{\underline{W_I}\overline{R(W)}}$. Therefore this direction is obvious.

($\Rightarrow$) For the sake of contradiction, suppose that $(Y \perp\!\!\!\perp W_I \mid W \setminus W_I)_{\mathcal{G}_{\underline{W_I}}} \wedge (Y \perp\!\!\!\perp R \mid W)_{\mathcal{G}_{\underline{W_I}\overline{R(W)}}}$ holds while $(Y \perp\!\!\!\perp W_I \mid W \setminus W_I)_{\mathcal{G}_{\underline{W_I}}} \wedge (Y \perp\!\!\!\perp R \mid W)_{\mathcal{G}_{\overline{R(W)}}}$ does not hold. Notice that the only difference between the two statements is the edges outgoing from W_I in $\mathcal{G}_{\overline{R(W)}}$. There exists a path p that is active between $U \in Y$ and $T \in R$ given W in $\mathcal{G}_{\overline{R(W)}}$. p is blocked in $\mathcal{G}_{\underline{W_I}\overline{R(W)}}$. There are two cases, either p goes through an edge that is removed in $\mathcal{G}_{\underline{W_I}\overline{R(W)}}$ or the removed edges deactivate a collider on p.

For the first case, if removing outgoing edges around W_I breaks p, then conditioning on W_I should also block p, and thus the supposition does not hold.

For the second case, suppose there is a collider $C \in An(W_I)$ on p. Removing outgoing edges from W_I will not break the directed path from C to W_I and thus C is still activated on p in $\mathcal{G}_{\underline{W_I}\overline{R(W)}}$. □

Due to Lemma B.2, we can show the equivalence between the two statements. Thus, we have $P^{(0)}(y|w) = P_I(y|w)$ if $(Y \perp\!\!\!\perp W_I \mid W \setminus W_I)_{\mathcal{G}^{(0)}_{\underline{W_I}}} \wedge (Y \perp\!\!\!\perp R \mid W)_{\mathcal{G}^{(0)}_{\overline{R(W)}}}$. Similarly, we can show

$P^{(1)}(y|w) = P_I(y|w)$ if $(Y \perp\!\!\!\perp W_I \mid W \setminus W_I)_{\mathcal{G}^{(1)}_{\underline{W_I}}} \wedge (Y \perp\!\!\!\perp R \mid W)_{\mathcal{G}^{(1)}_{\overline{R(W)}}}$. This concludes the proof.

□

Remark B.3 (Interpretation of the mechanism-shift set). The mechanism-shift set I is defined in Section 3 using both causal-mechanism stability and observable conditional invariance. Hence, I contains every variable whose causal mechanism differs across the domains, and may additionally contain a variable whose causal mechanism is unchanged when no conditioning set over $\mathbf{V}$ witnesses its stability across the two domains. Therefore, I need not be a minimal set of structurally changed mechanisms. Intervening on such additional variables removes the same mechanism in both domains and therefore does not destroy the common bridge. The do-calculus argument in Proposition 3.1 requires only that intervention on I yields the common bridge model and distribution P_I.

B.2 Proof for Proposition 4.4

Proof. Consider a pair of domains (i, j) and write $I = I^{(i,j)}$ for its mechanism-shift set. We first show the equivalence in Equation 1 for within-domain separations. Consider the right hand side of the equation, since F nodes are all source nodes without incoming edges, conditioning on the F nodes will block all the paths that go through F nodes while not affecting other paths. Therefore the two sides are equivalent.

Next, we show the equivalence for Equation 2. Notice that it is sufficient to show:

$$\begin{cases} (Y \perp\!\!\!\perp R|\mathbf{W})_{\mathcal{G}^{(i)}_{\overline{R(W)}}} \\ (Y \perp\!\!\!\perp W_I|W \setminus W_I)_{\mathcal{G}^{(i)}_{\underline{W_I}}} \end{cases} \iff (F^{(i,j)} \perp\!\!\!\perp Y^{(i)}|W^{(i)})_{\mathrm{Aug}^{(i,j)}} \tag{6}$$

and

$$\begin{cases} (Y \perp\!\!\!\perp R|\mathbf{W})_{\mathcal{G}^{(j)}_{\overline{R(W)}}} \\ (Y \perp\!\!\!\perp W_I|W \setminus W_I)_{\mathcal{G}^{(j)}_{\underline{W_I}}} \end{cases} \iff (F^{(i,j)} \perp\!\!\!\perp Y^{(j)}|W^{(j)})_{\mathrm{Aug}^{(i,j)}} \tag{7}$$

Below we show the equivalence in Equation 6 for domain i.

($\Rightarrow$) For the sake of contradiction, suppose the left hand side holds while the right hand side does not hold, i.e., there is an active minimal path p between F and Y given W in Aug. Here we omit the superscripts for simplicity. There are two cases, p either goes through R or W_I.

If p goes through $R(W)$, p starts at $U \in R(W)$ with either an arrowhead or arrowtail at U. If it is an arrowhead, U becomes a collider on p. Since U is not an ancestor of any node in W, p is blocked in this case. Therefore, p has to start at U with an outgoing edge. If p is a directed path, there does not exist any node in W on p, because otherwise p would be blocked by conditioning on W. However, the same directed path would also appear in $\mathcal{G}^{(i)}_{\overline{R(W)}}$ which contradicts with $(Y \perp\!\!\!\perp R|\mathbf{W})_{\mathcal{G}^{(i)}_{\overline{R(W)}}}$. Thus there is at least one collider on p. Consider the collider C that is closest to F on p. In this case, to make p active, C has to be an ancestor of a node T in W. However, that would create a directed path from U to T while U is defined as non-ancestor of any node in W. Therefore, p cannot start at any node in $R(W)$. On the other hand, consider such active p through $U \in R \setminus R(W)$. However, if p is active, the path p' constructed by removing $F \to U$ in $\mathcal{G}^{(i)}_{\overline{R(W)}}$ would also be active. To see this, p' can only be inactive due to the removal of an edge into $R(W)$. Denote $T \in R(W)$ that is closest to Y. In this case, there exists another path p'' that starts with $F \to T$ and attaches to the subpath between T and Y on p. p'' would also be active while we have already shown that such path does not exist.

If p goes through $U \in W_I$, p starts at U with either an arrowhead or arrowtail. If p starts with an arrowtail at U, conditioning on W would block p. Thus p can only start at U with an arrowhead. To make p blocked or broken in $\mathcal{G}^{(i)}_{\underline{W_I}}$, there is either a node in W_I on p with outgoing edges or a collider

that is not an ancestor of any node in W. If it is the former case, then conditioning on W will block p in Aug. If it is the latter case, p will be blocked by the collider. Therefore, such p does not exist.

Since either case cannot hold, the supposition is wrong.

($\Leftarrow$) For the sake of contradiction, suppose that the left hand side of Equation 6 does not hold, i.e., at least one of the two statements does not hold. We discuss each of them respectively.

If $(Y \perp\!\!\!\perp R|\mathbf{W})_{\mathcal{G}^{(i)}_{\overline{R(W)}}}$ does not hold, there is an active minimal path p between Y and R given W in $\mathcal{G}^{(i)}_{\overline{R(W)}}$. Consider $U \in R(W)$ as an end of p. Since all edges into $R(W)$ are removed, p has to start with an outgoing edge from U. Since U is not ancestor of any node in W, p has to be a directed path to Y since any collider on p would block p. Also, there is no node in W on the directed path because conditioning on W would block it. However, if such a path p exists, $F \to U$ can be attached to p in Aug to create an active path, contradicting the given condition. Consider $U \in R \setminus R(W)$ as an end of p. In this case, the path constructed by attaching $F \to U$ to p is also active. To witness, if p starts with an arrowhead at U, since U is an ancestor of some node in W, U will be an activated collider on the path in Aug and thus the path would be active if p is active. Otherwise, if p starts with an arrowtail at U, the path in Aug would also be active if p is active. Thus the supposition is wrong.

If $(Y \perp\!\!\!\perp W_I|W \setminus W_I)_{\mathcal{G}^{(i)}_{\underline{W_I}}}$ does not hold, there is an active minimal path p between Y and W_I given W in $\mathcal{G}^{(i)}_{\underline{W_I}}$. Since edges that go out of W_I are removed, p has to start with an arrowhead at U. However, if p is active, the path constructed by attaching $F \to U$ in Aug is also active. Since $U \in W$, conditioning on U will activate U on the path while the rest of the path is the same as p. Therefore, the supposition does not hold.

Since either statement cannot be violated, the supposition is wrong.

This concludes the proof.

□

B.3 Proof for Theorem 4.6

We first introduce a few auxiliary facts. Consider an unordered pair of domains (i, j), and write $F = F^{(i,j)}$. For a twin augmented MAG $\mathcal{M} = \mathrm{Twin}^{(i,j)}(\mathbf{G})$, define the set of testable m-separation statements

$$\begin{aligned}\mathcal{T}(\mathcal{M}) :=& \{(X^{(i)} \perp\!\!\!\perp Y^{(i)} \mid Z^{(i)}, F)_{\mathcal{M}} : X, Y \subseteq V,\ Z \subseteq \mathbf{V} \setminus (X \cup Y)\} \\ &\cup \{(X^{(j)} \perp\!\!\!\perp Y^{(j)} \mid Z^{(j)}, F)_{\mathcal{M}} : X, Y \subseteq \mathbf{V},\ Z \subseteq \mathbf{V} \setminus (X \cup Y)\} \\ &\cup \{(F \perp\!\!\!\perp Y^{(i)} \mid W^{(i)})_{\mathcal{M}} \wedge (F \perp\!\!\!\perp Y^{(j)} \mid W^{(j)})_{\mathcal{M}} : Y \subseteq \mathbf{V},\ W \subseteq \mathbf{V} \setminus Y\}.\end{aligned}$$

By Proposition 4.4, the first two classes correspond to within-domain CI constraints, while the third class corresponds to cross-domain inv-constraints witnessed by a common observed conditioning set in the two domains. This requirement on common conditioning set is precisely the one used in the definition of the mechanism-shift set $I^{(i,j)}$.

Lemma B.4. *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two twin augmented MAGs for the same pair of domains (i, j). If $\mathcal{M}_1$ and $\mathcal{M}_2$ do not satisfy the three MAG equivalence conditions in Theorem 4.6, then*

$$\mathcal{T}(\mathcal{M}_1) \neq \mathcal{T}(\mathcal{M}_2).$$

Proof. If the three MAG equivalence conditions fail, then by the standard MAG equivalence characterization, $\mathcal{M}_1$ and $\mathcal{M}_2$ imply different m-separation relations. We show that such a difference can always be witnessed by a statement in $\mathcal{T}$.

We first show that $\mathcal{M}_1$ and $\mathcal{M}_2$ should have the same skeleton.

We begin by showing that they have the same skeleton on $\mathbf{V}^{(i)}$. Suppose otherwise, in $\mathcal{M}_1$, $X^{(i)}$ and $Y^{(i)}$ are adjacent but they are non-adjacent in $\mathcal{M}_2$. Then it implies that $(X^{(i)} \not\perp\!\!\!\perp Y^{(i)} \mid W, F)_{\mathcal{M}_1} \wedge (X^{(i)} \perp\!\!\!\perp Y^{(i)} \mid W^{(i)}, F)_{\mathcal{M}_2}$ for some $W \subseteq \mathbf{V}$ which is a testable separation statement. Similarly, we can show that $\mathcal{M}_1$ and $\mathcal{M}_2$ have the same skeleton in $\mathbf{V}^{(j)}$.

Next, we show that F has the same adjacencies in the two graphs. Suppose otherwise. Without loss of generality, let F be adjacent to $X^{(i)}$ in $\mathcal{M}_1$ but non-adjacent to $X^{(i)}$ in $\mathcal{M}_2$. By Lemma B.5, the adjacency of F is symmetric across the two domain copies. Hence F is adjacent to both $X^{(i)}$ and $X^{(j)}$ in $\mathcal{M}_1$, while it is non-adjacent to both in $\mathcal{M}_2$.

We use the following consequence of the definition of the mechanism-shift set.

Lemma B.5. *Consider $\mathcal{M} = \mathrm{Twin}^{(i,j)}(\mathbf{G})$ with associated mechanism-shift set $I^{(i,j)}$. For every $X \in \mathbf{V}$,*

$$F^{(i,j)} \to X^{(i)} \iff F^{(i,j)} \to X^{(j)} \iff X \in I^{(i,j)}. \tag{8}$$

Moreover, if $X \notin I^{(i,j)}$, there exists a common $W \subseteq \mathbf{V} \setminus \{X\}$ such that

$$(F^{(i,j)} \perp\!\!\!\perp X^{(i)} \mid W^{(i)})_{\mathcal{M}} \wedge (F^{(i,j)} \perp\!\!\!\perp X^{(j)} \mid W^{(j)})_{\mathcal{M}}. \tag{9}$$

Proof. If $X \in I^{(i,j)}$, then by Definition 4.3, the augmented pair graph contains both $F^{(i,j)} \to X^{(i)}$ and $F^{(i,j)} \to X^{(j)}$. These direct adjacencies remain present in its MAG.

If $X \notin I^{(i,j)}$, then by definition the causal mechanism of X is shared across the two domains and its stability admits a common observed conditioning witness $W \subseteq \mathbf{V} \setminus \{X\}$. By Proposition 4.4, this witness corresponds to

$$(F^{(i,j)} \perp\!\!\!\perp X^{(i)} \mid W^{(i)}) \wedge (F^{(i,j)} \perp\!\!\!\perp X^{(j)} \mid W^{(j)}) \tag{10}$$

in the augmented pair graph. These separation relations are preserved by the MAG representation. Hence $F^{(i,j)}$ is non-adjacent to both copies of X. □

Applying Lemma B.5 to $\mathcal{M}_2$, there exists a common $W \subseteq \mathbf{V} \setminus \{X\}$ such that

$$(F \perp\!\!\!\perp X^{(i)} \mid W^{(i)})_{\mathcal{M}_2} \wedge (F \perp\!\!\!\perp X^{(j)} \mid W^{(j)})_{\mathcal{M}_2}. \tag{11}$$

Thus this paired separation belongs to $\mathcal{T}(\mathcal{M}_2)$. In $\mathcal{M}_1$, however, F is adjacent to both copies of X, so neither adjacency can be m-separated by any conditioning set. Hence the same statement does not belong to $\mathcal{T}(\mathcal{M}_1)$, contradicting $\mathcal{T}(\mathcal{M}_1) = \mathcal{T}(\mathcal{M}_2)$. Therefore the F-adjacencies agree, and $\mathcal{M}_1$ and $\mathcal{M}_2$ have the same skeleton.

Now suppose the skeletons agree but the unshielded colliders differ. By the standard proof of MAG equivalence, a difference in the collider status of an unshielded triple implies a difference in an m-separation statement between the two endpoints. If the triple is contained in one domain copy, this gives a statement in the first or second class of $\mathcal{T}$. If the triple contains F, then F cannot be the middle vertex because all edges adjacent to F have tails at F. Hence F is an endpoint, and the same argument reduces the difference to a paired F-separation statement in the third class of $\mathcal{T}$.

Finally, suppose the skeletons and unshielded colliders agree, but there is a common discriminating path whose discriminated vertex has different collider status. Again, the standard MAG equivalence argument yields an m-separation that differs between the two graphs. If the path does not contain F, the difference is within one domain copy and is in the first or second class of $\mathcal{T}$. If F is on the path, then F can only be an endpoint: it cannot be an internal collider because no edge has an arrowhead into F. Thus the difference again reduces to whether F is m-separated from the opposite endpoint under the projected conditioning set, yielding a statement in the third class of $\mathcal{T}$. Therefore $\mathcal{T}(\mathcal{M}_1) \neq \mathcal{T}(\mathcal{M}_2)$. □

Lemma B.6 (Meek's path lemma)**.** *(Meek, 2013b) Consider a causal DAG $\mathcal{D} = (\mathbf{V}, \mathbf{E})$, and suppose $(A \not\perp\!\!\!\perp B \mid C)_{\mathcal{D}}$. Let $\mathcal{D}_s = (\mathbf{V}_s, \mathbf{E}_s)$ be the subgraph containing all nodes on an active path that witnesses $(A \not\perp\!\!\!\perp B \mid C)_{\mathcal{D}}$. Then any distribution over $\mathbf{V}_s$ in which every adjacent pair of variables in $\mathcal{D}_s$ is dependent satisfies*

$$A \not\perp\!\!\!\perp B \mid C.$$

Lemma B.7. *Fix an unordered pair of domains (i, j). Let $(\mathbf{G}_1, \mathcal{I}_1)$ and $(\mathbf{G}_2, \mathcal{I}_2)$ be two causal graph tuples with their associated mechanism-shift collections, where $\mathcal{I}_r = \{I_r^{(a,b)} : 1 \le a < b \le K\}$ for $r \in \{1, 2\}$. Let*

$$\mathcal{M}_1 = \mathrm{Twin}^{(i,j)}(\mathbf{G}_1), \qquad \mathcal{M}_2 = \mathrm{Twin}^{(i,j)}(\mathbf{G}_2). \tag{12}$$

If there exists a testable statement

$$\tau \in \mathcal{T}(\mathcal{M}_1) \setminus \mathcal{T}(\mathcal{M}_2), \tag{13}$$

then there exists a tuple of domain distributions in $\hat{\mathbf{P}}(\mathbf{G}_2, \mathcal{I}_2, \mathbf{V})$ *that is not in* $\hat{\mathbf{P}}(\mathbf{G}_1, \mathcal{I}_1, \mathbf{V})$.

Proof. Since $\tau \notin \mathcal{T}(\mathcal{M}_2)$, the corresponding m-separation fails in $\mathcal{M}_2$. We construct a distribution tuple compatible with $(\mathbf{G}_2, \mathcal{I}_2)$ that violates the constraint τ imposed by $(\mathbf{G}_1, \mathcal{I}_1)$.

First suppose that τ is a within-domain CI statement. For example,

$$(X^{(i)} \perp\!\!\!\perp Y^{(i)} \mid Z^{(i)}, F^{(i,j)})_{\mathcal{M}_1} \quad \text{but} \quad (X^{(i)} \not\perp\!\!\!\perp Y^{(i)} \mid Z^{(i)}, F^{(i,j)})_{\mathcal{M}_2}. \tag{14}$$

Since conditioning on $F^{(i,j)}$ blocks paths that pass through the F-node, this difference is witnessed by an m-connection between $X^{(i)}$ and $Y^{(i)}$ given $Z^{(i)}$ inside the i-th domain copy. By the latent-projection property of MAGs, there exists a corresponding d-connecting path between X and Y given Z in the underlying causal graph $\mathcal{G}_2^{(i)}$. Let $\mathcal{D}_{\text{path}}$ denote the subgraph containing the observed and latent variables on such a path.

Let $T = (T_1, \ldots, T_m)$ be a topological ordering of the variables in $\mathcal{D}_{\text{path}}$. Define

$$T = AT + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \Omega), \tag{15}$$

where $A_{ab} \neq 0$ only if $T_b \to T_a$ is an edge in $\mathcal{D}_{\text{path}}$, and Ω is diagonal with positive entries. Equivalently,

$$T = (I - A)^{-1}\epsilon, \qquad \Sigma = (I - A)^{-1}\Omega(I - A)^{-T}. \tag{16}$$

The covariance between every adjacent pair on the path is a polynomial function of the free edge coefficients and noise variances that is not identically zero. Hence we may choose the parameters so that every adjacent pair on $\mathcal{D}_{\text{path}}$ is dependent. By Lemma B.6,

$$X \not\perp\!\!\!\perp Y \mid Z \tag{17}$$

holds under the induced observed distribution $P^{(i)}$.

We extend this construction to all variables and domains using structural equations consistent with $\mathbf{G}_2$ and the associated mechanism-shift collection $\mathcal{I}_2$. For every pair of domains (a, b) and every variable $S \notin I_2^{(a,b)}$, we choose the same causal mechanism in the two domains and preserve a common observed conditioning witness for its stability, as required by the definition of the mechanism-shift set. For $S \in I_2^{(a,b)}$, we may choose its causal mechanisms to differ, which is one valid realization of membership in the mechanism-shift set. The remaining free parameters can be chosen generically while preserving these compatibility conditions and the dependence constructed above. Thus the resulting tuple belongs to $\hat{\mathbf{P}}(\mathbf{G}_2, \mathcal{I}_2, \mathbf{V})$, but violates the within-domain CI constraint imposed by $(\mathbf{G}_1, \mathcal{I}_1)$. Hence it does not belong to $\hat{\mathbf{P}}(\mathbf{G}_1, \mathcal{I}_1, \mathbf{V})$.

Now suppose that τ is a cross-domain inv-constraint. Then $\mathcal{M}_1$ imposes a paired F-separation of the form

$$(F^{(i,j)} \perp\!\!\!\perp Y^{(i)} \mid W^{(i)})_{\mathcal{M}_1} \quad \text{and} \quad (F^{(i,j)} \perp\!\!\!\perp Y^{(j)} \mid W^{(j)})_{\mathcal{M}_1}, \tag{18}$$

whereas at least one of these separations fails in $\mathcal{M}_2$. Without loss of generality, suppose

$$(F^{(i,j)} \not\perp\!\!\!\perp Y^{(i)} \mid W^{(i)})_{\mathcal{M}_2}. \tag{19}$$

By the latent-projection property, this failure is witnessed by a d-connecting path from $F^{(i,j)}$ to $Y^{(i)}$ given $W^{(i)}$ in the underlying augmented causal graph associated with $\text{Aug}^{(i,j)}(\mathbf{G}_2)$. Let $\mathcal{D}_{\text{path}}$ denote the subgraph containing the observed and latent variables on such a path.

We construct a regime-indexed linear Gaussian SCM. Let $F = F^{(i,j)} \sim \text{Bernoulli}(1/2)$, where $F = 0$ represents domain i and $F = 1$ represents domain j. For the variables $T = (T_1, \ldots, T_m)$ in $\mathcal{D}_{\text{path}}$, define

$$T = A(F)T + \epsilon(F), \tag{20}$$

where

$$A(F) = \begin{cases} A^{(i)}, & F = 0, \\ A^{(j)}, & F = 1, \end{cases} \qquad \epsilon(F) \sim \begin{cases} \mathcal{N}(0, \Omega^{(i)}), & F = 0, \\ \mathcal{N}(0, \Omega^{(j)}), & F = 1. \end{cases} \tag{21}$$

The matrices $A^{(i)}$ and $A^{(j)}$ respect the corresponding domain graphs $\mathcal{G}_2^{(i)}$ and $\mathcal{G}_2^{(j)}$.

For every $S \notin I_2^{(i,j)}$, choose the same causal mechanism in the two domains and choose the remaining compatible parameters so that a common observed conditioning set witnesses its stability, as required by the definition of $I_2^{(i,j)}$. For every $S \in I_2^{(i,j)}$, we may choose at least one structural parameter or noise parameter to differ across the two domains. Thus the construction is compatible with the mechanism-shift set $I_2^{(i,j)}$.

Conditional on $F = f$, the distribution is Gaussian with covariance

$$\Sigma_f = (I - A^{(f)})^{-1}\Omega^{(f)}(I - A^{(f)})^{-T}. \tag{22}$$

We choose the remaining free coefficients and variances generically. For every ordinary edge on the d-connecting path, the corresponding adjacent variables can therefore be chosen to be dependent. For every edge $F \to S$ on the path, we choose the mechanisms in the two regimes so that the conditional law of S differs across the regimes, and hence

$$F \not\perp\!\!\!\perp S. \tag{23}$$

Thus every adjacent pair on $\mathcal{D}_{\mathrm{path}}$ can be chosen to be dependent. By Lemma B.6,

$$F \not\perp\!\!\!\perp Y \mid W \tag{24}$$

under the joint regime distribution $P^\star(F, \mathbf{V})$.

Define the two domain distributions by

$$P^{(i)}(\mathbf{V}) = P^\star(\mathbf{V} \mid F = 0), \qquad P^{(j)}(\mathbf{V}) = P^\star(\mathbf{V} \mid F = 1). \tag{25}$$

Since $F \not\perp\!\!\!\perp Y \mid W$, the conditional distribution of Y given W depends on the value of F. Hence, for some y, w,

$$P^{(i)}(y \mid w) \neq P^{(j)}(y \mid w). \tag{26}$$

Thus the cross-domain invariance imposed by $(\mathbf{G}_1, \mathcal{I}_1)$ is violated.

Finally, extend the construction to the remaining variables and domains using structural equations consistent with $\mathbf{G}_2$ while preserving the mechanism-sharing and common-witness conditions associated with $\mathcal{I}_2$. The remaining free parameters can be chosen generically so that the constructed dependence is preserved. The resulting tuple is therefore compatible with $(\mathbf{G}_2, \mathcal{I}_2)$ and satisfies its inv-Markov constraints, while violating τ. Hence

$$(P^{(1)}, \ldots, P^{(K)}) \in \hat{\mathbf{P}}(\mathbf{G}_2, \mathcal{I}_2, \mathbf{V}) \setminus \hat{\mathbf{P}}(\mathbf{G}_1, \mathcal{I}_1, \mathbf{V}). \tag{27}$$

This proves the lemma. □

Lemma B.8. *For every pair of domains $i, j \in [K]$, the twin augmented graph $\mathrm{Twin}^{(i,j)}(\mathbf{G})$ in Definition 4.5 is a MAG. Moreover, for every domain $i \in [K]$, the inv-augmented graph $\mathrm{Aug}^{(i)}(\mathbf{G})$ in Definition 5.1 is also a MAG.*

Proof. By Definition 4.5,

$$\mathrm{Twin}^{(i,j)}(\mathbf{G}) = \mathrm{MAG}\Big(\mathrm{Aug}^{(i,j)}(\mathbf{G})\Big), \tag{28}$$

so $\mathrm{Twin}^{(i,j)}(\mathbf{G})$ is a MAG by construction.

We next consider $\mathrm{Aug}^{(i)}(\mathbf{G})$. Its subgraph over $\mathbf{V}^{(i)}$ is $\mathrm{MAG}(\mathcal{G}^{(i)})$, and every additional $F^{(i,j)}$ is a source node connected to observed variables only through outgoing edges. Adding these source nodes cannot create a directed cycle or change ancestral relations among the observed variables, so $\mathrm{Aug}^{(i)}(\mathbf{G})$ is ancestral.

For maximality, any inducing path between two non-adjacent observed variables that contains no F-node would already be an inducing path in $\mathrm{MAG}(\mathcal{G}^{(i)})$, contradicting its maximality. An F-node cannot occur internally on an inducing path because it cannot be a collider. Thus, any remaining candidate inducing path must have an F-node as an endpoint. If the other endpoint is an observed variable, the same argument as above gives a contradiction: the internal vertex adjacent to the

observed endpoint would have to be both an ancestor of that endpoint and incident to an arrowhead on the corresponding edge. If both endpoints are F-nodes, every internal collider would have to be an ancestor of one of the F-nodes, which is impossible because all F-nodes are sources.

Hence there is no inducing path between any pair of non-adjacent vertices in $\mathrm{Aug}^{(i)}(\mathbf{G})$. It is therefore maximal and ancestral, and thus a MAG. □

We now prove Theorem 4.6.

Proof of Theorem 4.6. **(If.)** Suppose that for every unordered pair (i, j), the twin augmented MAGs $\mathrm{Twin}^{(i,j)}(\mathbf{G}_1)$ and $\mathrm{Twin}^{(i,j)}(\mathbf{G}_2)$ satisfy the three MAG equivalence conditions. Then, by the MAG equivalence theorem, they induce the same m-separation statements. By Proposition 4.4, these m-separation statements are exactly the within-domain CI constraints and cross-domain inv-constraints appearing in Definition 4.1. Hence

$$\hat{\mathbf{P}}(\mathbf{G}_1, \mathcal{I}_1, \mathbf{V}) = \hat{\mathbf{P}}(\mathbf{G}_2, \mathcal{I}_2, \mathbf{V}),$$

so $(\mathbf{G}_1, \mathcal{I}_1)$ and $(\mathbf{G}_2, \mathcal{I}_2)$ are inv-Markov equivalent.

(Only if.) Suppose $(\mathbf{G}_1, \mathcal{I}_1)$ and $(\mathbf{G}_2, \mathcal{I}_2)$ are inv-Markov equivalent, but for some unordered pair (i, j), the twin augmented MAGs

$$\mathcal{M}_1 = \mathrm{Twin}^{(i,j)}(\mathbf{G}_1), \qquad \mathcal{M}_2 = \mathrm{Twin}^{(i,j)}(\mathbf{G}_2)$$

do not satisfy the three MAG equivalence conditions. By Lemma B.4, there exists a testable statement τ that holds in one of $\mathcal{M}_1, \mathcal{M}_2$ but not the other. Without loss of generality, let $\tau \in \mathcal{T}(\mathcal{M}_1) \setminus \mathcal{T}(\mathcal{M}_2)$. By Lemma B.7, there exists a tuple of distributions that belongs to $\hat{\mathbf{P}}(\mathbf{G}_2, \mathcal{I}_2, \mathbf{V})$ but not to $\hat{\mathbf{P}}(\mathbf{G}_1, \mathcal{I}_1, \mathbf{V})$. This contradicts inv-Markov equivalence. Therefore the twin augmented MAGs must satisfy the three MAG equivalence conditions for every pair (i, j). □

B.4 Proof for Proposition 5.2

Proof. For $r \in \{1, 2\}$, all twin augmented MAGs and inv-augmented MAGs constructed from $\mathbf{G}_r$ are understood to use its associated mechanism-shift collection $\mathcal{I}_r$. By Theorem 4.6, it is enough to show that the twin augmented MAGs satisfy the three MAG equivalence conditions for every pair (i, j) if and only if the inv-augmented MAGs satisfy them for every domain i.

(If.) Assume that for every domain i, $\mathrm{Aug}^{(i)}(\mathbf{G}_1)$ and $\mathrm{Aug}^{(i)}(\mathbf{G}_2)$ satisfy the three MAG equivalence conditions. Consider a pair (i, j). By Definitions 4.5 and 5.1, $\mathrm{Twin}^{(i,j)}(\mathbf{G})$ is the union of the induced subgraph of $\mathrm{Aug}^{(i)}(\mathbf{G})$ on $\mathbf{V}^{(i)} \cup \{F^{(i,j)}\}$ and the induced subgraph of $\mathrm{Aug}^{(j)}(\mathbf{G})$ on $\mathbf{V}^{(j)} \cup \{F^{(i,j)}\}$. Thus the skeletons of $\mathrm{Twin}^{(i,j)}(\mathbf{G}_1)$ and $\mathrm{Twin}^{(i,j)}(\mathbf{G}_2)$ agree.

The same argument gives equality of unshielded colliders. Any unshielded triple in a twin augmented MAG is either contained in one of the two induced subgraphs, or has $F^{(i,j)}$ as the middle node and one endpoint in each domain copy. The latter case is always a non-collider because all edges adjacent to $F^{(i,j)}$ have tails at $F^{(i,j)}$. Hence the unshielded colliders agree. For discriminating paths, $F^{(i,j)}$ cannot be an internal collider, since no edge has an arrowhead into $F^{(i,j)}$. Therefore every relevant discriminating path is contained in one of the two induced subgraphs, except possibly for having $F^{(i,j)}$ as an endpoint. In both cases, the collider status is inherited from the corresponding inv-augmented MAG. Hence the twin augmented MAGs satisfy the three conditions.

(Only if.) Assume that for every pair (i, j), the twin augmented MAGs $\mathrm{Twin}^{(i,j)}(\mathbf{G}_1)$ and $\mathrm{Twin}^{(i,j)}(\mathbf{G}_2)$ satisfy the three MAG equivalence conditions. Fix a domain i. Since $\mathrm{Aug}^{(i)}(\mathbf{G})$ is the graph union of the induced subgraphs of $\mathrm{Twin}^{(i,j)}(\mathbf{G})$ on $\mathbf{V}^{(i)} \cup \{F^{(i,j)}\}$ over all $j \neq i$, the skeletons of $\mathrm{Aug}^{(i)}(\mathbf{G}_1)$ and $\mathrm{Aug}^{(i)}(\mathbf{G}_2)$ agree.

For unshielded triples, if at most one endpoint is an F-node, then the triple appears in the relevant twin augmented MAG and therefore has the same collider status in both graphs. If both endpoints are F-nodes, then the triple has the form

$$F^{(i,j)} \to X^{(i)} \leftarrow F^{(i,k)},$$

and is a collider in both graphs by construction. Thus the unshielded colliders agree.

Finally, consider a discriminating path in both $\mathrm{Aug}^{(i)}(\mathbf{G}_1)$ and $\mathrm{Aug}^{(i)}(\mathbf{G}_2)$. An F-node cannot be an internal collider because all edges adjacent to an F-node have tails at the F-node. Hence the path either contains no F-node, contains one $F^{(i,j)}$ as an endpoint, or has an F-node as the discriminated vertex with the same non-collider status in both graphs. In the first two cases, the path is contained in the corresponding twin augmented MAG, so the collider status is the same by assumption. In the last case, the status is also the same because F-nodes cannot be colliders. Therefore $\mathrm{Aug}^{(i)}(\mathbf{G}_1)$ and $\mathrm{Aug}^{(i)}(\mathbf{G}_2)$ satisfy the three MAG equivalence conditions for every i. The result follows from Theorem 4.6. □

B.5 Proof for Theorem 6.3

Proof. Let $\mathbf{G} = (\mathcal{G}^{(i)})_{i=1}^{K}$ be the true graph tuple with associated mechanism-shift collection $\mathcal{I}$, and let

$$\mathcal{L}(\mathbf{G}) = \Big(\mathcal{P}^{(i)}(\mathbf{G})\Big)_{i=1}^{K} \tag{29}$$

be the true inv-augmented graph tuple. We show that every edge deletion and every orientation made by Algorithm 1 is valid in $\mathcal{L}(\mathbf{G})$.

First consider adjacencies between two observed variables $X^{(i)}, Y^{(i)}$ in the same domain. Algorithm 1 uses the usual FCI adjacency search with oracle CI tests on $P^{(i)}$. By inv-faithfulness, these CI statements are equivalent to the corresponding m-separation statements in the MAG of $\mathcal{G}^{(i)}$. Hence the learned within-domain skeleton is correct by the soundness of FCI.

We next consider the adjacencies involving an F-node. Fix a pair of domains (i, j) and a variable $Y \in \mathbf{V}$. By Lemma B.5, the target F-adjacencies satisfy

$$F^{(i,j)} \sim Y^{(i)} \quad \Longleftrightarrow \quad F^{(i,j)} \sim Y^{(j)} \quad \Longleftrightarrow \quad Y \in I^{(i,j)}. \tag{30}$$

Thus, the target F-neighborhood is symmetric by construction.

The role of inv-faithfulness is instead to ensure that the distributional invariance tests used by Algorithm 1 recover these target adjacencies.

Now consider the pairwise invariance search in Algorithm 1. The algorithm removes the edge between $F^{(i,j)}$ and $Y^{(i)}$ if it finds some $W \subseteq \mathbf{V} \setminus \{Y\}$ such that

$$P^{(i)}(y \mid w) = P^{(j)}(y \mid w). \tag{31}$$

By inv-faithfulness and Proposition 4.4, this implies the paired F-separation above, and hence $Y \notin I^{(i,j)}$. Thus the deleted edge is indeed absent from the corresponding twin augmented MAG and from the domain-centric inv-augmented MAG.

If no such W exists, then by the definition of the mechanism-shift set $Y \in I^{(i,j)}$. By Lemma B.5, the corresponding F-edge is therefore present in the target graph. Hence the oracle pairwise invariance search recovers exactly the F-adjacencies of the true inv-augmented MAG tuple.

It remains to verify the orientations. For unshielded triples involving only observed variables, Rule 0 is sound by the standard FCI argument, since the within-domain skeleton and separating sets are correct.

Now consider an unshielded triple involving $F^{(i,j)}$. Whenever Algorithm 1 removes an $F^{(i,j)}$–$Y^{(i)}$ edge using a conditioning set W, the argument above shows that W gives a valid $F^{(i,j)}$–$Y^{(i)}$ separating set in the corresponding pairwise augmented representation. By Lemma B.5, the corresponding target F-nonadjacency is symmetric across the two domain copies, and the paired separation is preserved in the twin augmented MAG. In the domain-centric graph, the algorithm additionally conditions on the other F-nodes. These are source nodes, so conditioning on them blocks paths through those nodes and does not activate new paths. Therefore the separating sets used by Rule 0 are valid in the true inv-augmented MAG. Hence Rule 0 is sound also for unshielded triples involving an F-node.

The remaining standard FCI orientation rules are sound when applied to the true inv-augmented MAGs with the correct skeleton and separating sets.

Finally, Rule 8 orients every edge incident to an F-node outward from the F-node. This is sound by construction. In the augmented pair graph, F-nodes are auxiliary source nodes, and every adjacency incident to an F-node in the corresponding twin and domain-centric inv-augmented MAGs has a tail at the F-node. Thus no member of the inv-Markov equivalence class contains an arrowhead into an F-node.

Therefore every edge deletion and every orientation made by Algorithm 1 is valid in the true inv-augmented graph tuple $\mathcal{L}(\mathbf{G})$. Hence Algorithm 1 is sound. □

C Algorithms and Orientation Rules

C.1 Algorithms

Algorithm 2 Algorithm for Creating F Nodes

Input: Integer i, K
Initialize $\mathbf{F} \leftarrow \emptyset$;
for j **in** $[K] \setminus \{i\}$ **do**
  $\mathbf{F} \leftarrow \mathbf{F} \cup \{F^{(i,j)}\}$
Output: The set of F nodes $\mathbf{F}$

Algorithm 3 Algorithm for Finding Separating Set

Input: Integers i, j, distributions $(P^{(i)})_{i\in[K]}$, observable nodes $\mathbf{V}$, F nodes $\mathbf{F}^{(i)}$
Initialize $\mathcal{E} \leftarrow \emptyset$;
if $i = j$ **then**
  for X, Y **in** $\mathbf{V}$ **do**
    $SepFlag \leftarrow False$;
    for $\mathbf{W} \subseteq \mathbf{V} \setminus \{X, Y\}$ **do**
      if $P^{(i)}(y|\mathbf{w}, x) = P^{(i)}(y|\mathbf{w})$ **then**
        $SepFlag \leftarrow True$
        $SepSet(X^{(i)}, Y^{(i)}) \leftarrow \mathbf{W}^{(i)} \cup \mathbf{F}^{(i)}$;
    if $SepFlag = False$ **then**
      $\mathcal{E} \leftarrow \mathcal{E} \cup (X^{(i)}, Y^{(i)})$;
else
  for Y **in** $\mathbf{V}$ **do**
    $SepFlag \leftarrow False$;
    for $\mathbf{W} \subseteq \mathbf{V} \setminus \{Y\}$ **do**
      if $P^{(i)}(y|\mathbf{w}) = P^{(j)}(y|\mathbf{w})$ **then**
        $SepFlag \leftarrow True$;
        $SepSet(F^{(i,j)}, Y^{(i)}) \leftarrow \mathbf{W}^{(i)} \cup \mathbf{F}^{(i)} \setminus \{F^{(i,j)}\}$;
        $SepSet(F^{(i,j)}, Y^{(j)}) \leftarrow \mathbf{W}^{(j)} \cup \mathbf{F}^{(j)} \setminus \{F^{(i,j)}\}$;
    if $SepFlag = False$ **then**
      $\mathcal{E} \leftarrow \mathcal{E} \cup (F^{(i,j)}, Y^{(i)})$;
Output: Circle edges $\mathcal{E}$, separating sets $SepSet$

C.2 Orientation Rules

Diverse-FCI applies Rule 0 in Algorithm 1 first to orient unshielded colliders, and then repeatedly applies the orientation rules below until no further orientation is possible. For convenience, we relabel the seven standard FCI orientation rules used in our setting (Zhang, 2008b) as Rules 1–7. These correspond to Rules R1–R4 and R8–R10 in the original FCI numbering. The original Rules R5–R7 concern undirected edges arising from selection bias and are not needed here. Rule 8 is the additional F-node orientation rule used by Diverse-FCI.

A path is *uncovered* if every consecutive triple on the path is unshielded. A path from A to B is *possibly directed* if no edge on the path contains an arrowhead pointing toward A along the path. We use $\mathrm{SepSet}(A, B)$ to denote a separating set for A and B.

Rule 1. If $A{*\!\!\rightarrow}\, Bo\!\!-\!\!*\, C$ and A and C are non-adjacent, orient $Bo\!\!-\!\!*\, C$ as $B \rightarrow C$.

Rule 2. If either $A \rightarrow B{*\!\!\rightarrow}\, C$ or $A{*\!\!\rightarrow}\, B \rightarrow C$, and $A *\!\!-\!\!o C$, orient $A *\!\!-\!\!o C$ as $A{*\!\!\rightarrow}\, C$.

Rule 3. If $A{*\!\!\rightarrow}\, B \leftarrow\!\!* C$, $A *\!\!-\!\!o D o\!\!-\!\!*\, C$, and $D *\!\!-\!\!o B$, where A and C are non-adjacent, orient $D *\!\!-\!\!o B$ as $D{*\!\!\rightarrow}\, B$.

Rule 4. Suppose $p = \langle A, \ldots, B, C, D\rangle$ is a discriminating path for C, and $Co\!\!-\!\!*\, D$. If $C \in \mathrm{SepSet}(A, D)$, orient $Co\!\!-\!\!*\, D$ as $C \rightarrow D$. Otherwise, orient C as a collider on $\langle B, C, D\rangle$, i.e., $B{*\!\!\rightarrow}\, C \leftarrow\!\!* D$.

Rule 5. If either $A \rightarrow B \rightarrow C$ or $A\!\!-\!\!o B \rightarrow C$, and $Ao\!\!\rightarrow C$, orient $Ao\!\!\rightarrow C$ as $A \rightarrow C$.

Rule 6. Suppose $Ao \rightarrow C$ and there exists an uncovered possibly directed path $p = \langle A, B, D, \ldots, C\rangle$ from A to C such that B and C are non-adjacent. Then orient $Ao\!\!\rightarrow C$ as $A \rightarrow C$.

Rule 7. Suppose $Ao\!\!\rightarrow C$ and $B \rightarrow C \leftarrow D$. Let p_1 and p_2 be uncovered possibly directed paths from A to B and from A to D, respectively. Let V_1 be the node adjacent to A on p_1 and V_2 the node adjacent to A on p_2. If $V_1 \neq V_2$ and V_1 and V_2 are non-adjacent, orient $Ao\!\!\rightarrow C$ as $A \rightarrow C$.

Rule 8. For any edge between an F-node and an observed variable, orient the endpoint at the F-node as a tail. Equivalently, $F^{(i,j)} *\!\!-\!\!* X^{(i)}$ is oriented as $F^{(i,j)}\!\!-\!\!*\, X^{(i)}$.

D Additional Experiments

D.1 Synthetic Experiment Details

For the synthetic experiments in Section 7, we first sample a random DAG over n observed binary variables with expected directed degree $d_{\rightarrow}$. We then add random bidirected edges with expected degree $d_{\leftrightarrow}$ to represent latent confounding. Each bidirected edge is realized by an independent latent root parent of its two endpoints.

Starting from this graph, we construct the remaining domain graphs by applying random structural or mechanism modifications while preserving acyclicity within each domain. The possible modifications are: adding or deleting a directed edge, reversing a directed edge, applying a hard or soft intervention to a node, and adding or deleting a latent confounder. Mechanisms not involved in a modification remain shared across the relevant domains.

Unless otherwise specified, we use $d_{\rightarrow} = 2$ and $d_{\leftrightarrow} = 1$, draw 20,000 i.i.d. samples per domain, and consider $m \in \{1, 2, 4, 8, 12, 16\}$ cross-domain modifications. Each reported result is averaged over 50 independent replications.

For CI tests, we use discrete G^2 tests. FCI-based methods use $\alpha = 0.05$ and maximum conditioning-set size 3. For methods that return a single global or context-augmented graph, we restrict the output to the observed system variables and compare the resulting graph with the true MAG in each domain.

Table 3: Domain-averaged partial-graph SHD as the number m of cross-domain structural or mechanism modifications increases ($n = 30$, $K = 2$). Results are mean $\pm$ standard error over 50 independent replications. Lower is better.

m	Diverse-FCI	Per-domain FCI	Pooled FCI	ψ-FCI	FCI-JCI	CD-NOD
1	61.15 ± 1.35	61.73 ± 1.47	58.98 ± 1.67	63.87 ± 1.51	59.10 ± 1.60	$\mathbf{57.05} \pm 1.53$
2	60.57 ± 1.55	62.00 ± 1.61	63.62 ± 1.65	63.85 ± 1.58	59.20 ± 1.60	$\mathbf{57.85} \pm 1.97$
4	59.92 ± 1.53	62.33 ± 1.53	75.67 ± 1.89	63.77 ± 1.62	61.27 ± 1.66	$\mathbf{58.83} \pm 1.91$
8	$\mathbf{59.13} \pm 1.31$	61.90 ± 1.35	93.55 ± 2.43	64.97 ± 1.41	62.83 ± 1.33	61.20 ± 1.68
12	$\mathbf{59.12} \pm 1.39$	63.10 ± 1.29	108.85 ± 3.81	66.40 ± 1.63	64.93 ± 1.44	65.52 ± 1.74
16	$\mathbf{59.80} \pm 1.25$	64.05 ± 1.21	111.02 ± 2.36	67.73 ± 1.41	68.47 ± 1.28	67.37 ± 1.55

Thus, all methods are evaluated on recovery of the domain-specific MAGs over the observed variables. This evaluation does not credit the additional mechanism-shift information encoded by the F-node neighborhoods returned by Diverse-FCI.

Skeleton SHD counts adjacency additions and deletions, whereas partial-graph SHD additionally counts disagreements in endpoint marks. Definite-endpoint precision and recall evaluate predicted tails and arrowheads, with circle endpoints treated as unresolved. Runtime measures the complete run of each method. We first average each metric across domains within a replication and then report the mean and standard error across replications.

D.2 Additional Results on Structural Heterogeneity

We further study how performance changes as the domains become increasingly heterogeneous. We fix $n = 30$ and $K = 2$ and vary the number m of cross-domain structural or mechanism modifications. Each modification is sampled from the operations described in Appendix D.1, including directed-edge addition, deletion, or reversal, local mechanism changes, and changes in latent confounding. Thus, larger m corresponds to domains that differ in more aspects of their causal structure or mechanisms.

Table 3 reports the domain-averaged partial-graph SHD. When the domains are nearly shared, methods that exploit a common or strongly coupled representation can perform well. In particular, CD-NOD obtains the lowest SHD for $m \in \{1, 2, 4\}$. As the number of modifications increases, however, the performance of these approaches deteriorates substantially. For example, CD-NOD increases from 57.05 at $m = 1$ to 67.37 at $m = 16$, while pooled FCI increases from 58.98 to 111.02.

In contrast, Diverse-FCI remains comparatively stable across the entire sweep, with partial-graph SHD between 59.12 and 61.15, and becomes the best-performing method from $m = 8$ onward. It also increasingly improves over per-domain FCI as heterogeneity grows. These results illustrate the tradeoff between pooling across similar domains and explicitly allowing domain-specific causal graphs: stronger sharing can be beneficial when the domains are close, whereas pairwise cross-domain invariances remain useful when substantial structural differences make a single common representation less appropriate.

D.3 SERGIO Simulations

We additionally evaluate the method on SERGIO (Dibaeinia & Sinha, 2020), a simulator for single-cell gene regulatory networks. This experiment is intended as a stress test on noisy gene-expression data, rather than as a primary benchmark for full graph recovery. We use a two-domain setting with $n = 10$ observed genes. The first domain is generated from a sparse regulatory network, and the second domain is obtained by perturbing a small number of genes or regulatory relations. We generate 2000 samples from each domain.

Since SERGIO produces continuous count-like expression measurements, we preprocess the simulated data before applying discrete conditional independence tests. Specifically, expression values are library-size normalized, transformed by $\log(1 + x)$, standardized gene-wise, and discretized into quantile bins. Genes with near-zero variance after preprocessing are removed if necessary. For each run, we compare the learned augmented graphs with simulator-defined augmented ground-truth graphs, in which the F-node is connected to the explicitly perturbed genes. We report both standard graph-recovery metrics and the recall of these perturbation targets from the learned F-node adjacencies. This perturbation-target set serves as a simulation proxy for the mechanism-shift set I; under

our general definition, I may also contain variables whose causal mechanisms are unchanged but whose stability cannot be witnessed by a common observed conditioning set.

Table 4: SERGIO simulation results for $n = 10$ observed genes and two domains. Metrics are computed separately for each domain and reported as mean $\pm$ standard error across 50 repeats.

Domain	Partial graph SHD	Skeleton F1	Perturbation-target recall	Runtime (s)
1	40.2500 ± 1.2500	0.2108 ± 0.0238	1.0000 ± 0.0000	0.3450 ± 0.0046
2	41.5000 ± 0.0000	0.2606 ± 0.0453	1.0000 ± 0.0000	0.3660 ± 0.0042

Table 4 shows that the method attains a perturbation-target recall of 1.0, meaning that all explicitly perturbed genes are adjacent to the learned F-node in this pilot experiment. However, the learned graphs are much denser than the ground-truth regulatory networks, leading to low skeleton F1 and high partial-graph SHD. Thus, SERGIO should not be viewed as evidence of accurate full regulatory graph recovery. Instead, it illustrates a complementary behavior: even when noisy single-cell-like expression data make full constraint-based graph learning difficult, the F-node adjacencies can still capture the explicit perturbation targets. As noted above, these targets are only a proxy for the theoretical mechanism-shift set, since additional variables may be mechanism-shifted in our sense when their stability has no common observable witness. A stronger SERGIO benchmark would require more repeats, improved sparsity control, and preprocessing or conditional independence tests better suited to noisy gene-expression measurements.

D.4 SEER Breast Cancer

We next evaluate the method on a large real-world cancer-registry dataset derived from the SEER (Doll et al., 2018; National Cancer Institute, 2018) breast-cancer export. The purpose of this experiment is to test whether the learned inv-augmented graphs can separate a stable clinical-pathological treatment-selection backbone from domain-sensitive variables associated with demographic, access, coding, and socioeconomic heterogeneity. Unlike the synthetic experiments, the domains here are not interventional. We therefore interpret F-adjacencies as candidate mechanism shifts across patient subpopulations, not as causal effects of ethnicity.

We use a local SEER breast-cancer export restricted to harmonized 2010+ records, yielding 832,665 case rows. Each row corresponds to one breast-cancer case. The domain variable is the SEER Hispanic ethnicity recode which has two domains: Non-Spanish-Hispanic-Latino and Spanish-Hispanic-Latino. After harmonization and complete-case filtering, we draw a balanced sample of 15,000 cases from each domain, for a final analysis table of 30,000 rows. Balanced sampling is used so that conditional-independence and cross-domain-invariance tests have comparable power in the two domains.

We use a treatment-selection graph with 19 observed variables, grouped as follows.

- **Baseline and pathology:** `age_at_diagnosis`, `marital_group`, `race_domain`, `diagnosis_era`, `histology_group`, `grade_harmonized`, and `tumor_size_harmonized`.
- **Biomarkers and stage:** `er_status`, `pr_status`, `her2_status`, `T_harmonized`, and `N_harmonized`.
- **Treatment selection:** `surgery_primary_site`, `radiation_recode`, `chemotherapy_recode`, `reason_no_surgery`, and `treatment_delay_group`.
- **Socioeconomic context:** `rural_urban_group` and `income_group`.

Before running causal discovery, era-dependent SEER fields are harmonized into common variables for grade, tumor size, T stage, and N stage. We use ER/PR/HER2 separately rather than including breast subtype, since subtype is largely determined by these biomarkers. We also exclude outcome variables, such as survival and vital status, because this experiment is intended to represent a treatment-selection graph rather than a post-treatment outcome graph.

We run the discrete version of the method over a small parameter grid over $\alpha \in \{0.0015, 0.01, 0.05\}$ and maximum condition set size $\in \{1, 2\}$ and select the sparsest non-degenerate graph according to the density diagnostics. The selected run uses $\alpha = 0.01$, maximum conditioning set size $= 2$. The chosen graph has observed graph density 0.4649, mean observed edge count 79.5 per domain, and 63 stable observed edges across the two domains. Seven of the 19 observed variables are adjacent to the F node.

Figure 3 shows the learned augmented graphs for the two ethnicity domains. Dark edges are stable across both domains, blue edges appear in only one domain, and dashed red edges are F-node edges indicating candidate mechanism shifts.

The F-adjacent variables are:

- `age_at_diagnosis`,
- `grade_harmonized`,
- `income_group`,
- `race_domain`,
- `radiation_recode`,
- `rural_urban_group`,
- `treatment_delay_group`.

This set is clinically interpretable. Age at diagnosis, income, race, rural-urban context, radiation use, and treatment delay can plausibly vary across ethnicity SEER subcohorts due to cohort composition, screening patterns, insurance and transportation barriers, and coding practice. The F-adjacency of `grade_harmonized` suggests that the conditional distribution of tumor grade also differs across domains. However, many of its incident edges remain stable, so this should be interpreted as a domain-sensitive reporting pattern rather than as a claim of ethnicity-specific tumor biology.

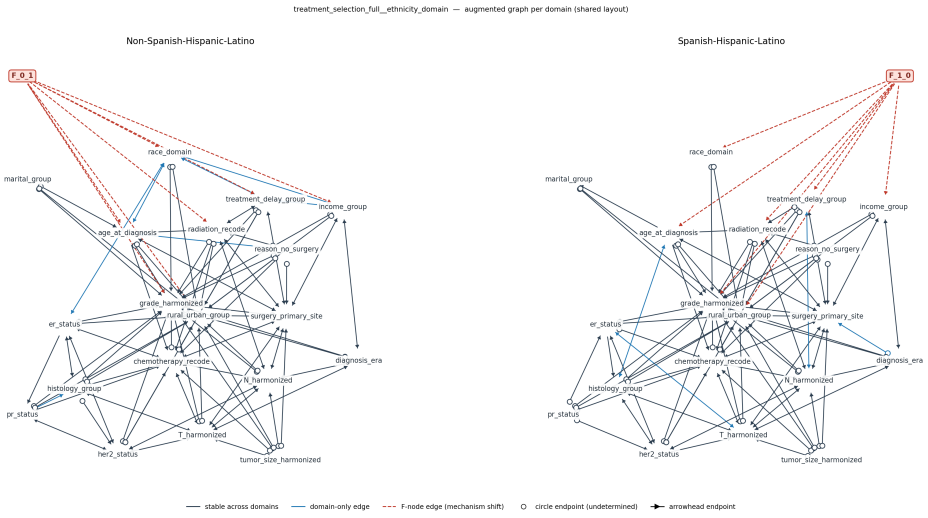


Figure 3: Learned inv-augmented graphs for the SEER ethnicity-domain treatment-selection experiment. Each panel shows the graph learned for one ethnicity domain using a shared layout. Dark edges are stable across both domains, blue edges appear only in one domain, and dashed red edges indicate F-node adjacencies, interpreted as candidate domain-sensitive mechanisms. Circle endpoints denote unresolved orientations.

At the same time, the graphs recover a large stable backbone. Many stable edges involve stage-treatment relations, such as connections among `T_harmonized`, `N_harmonized`, `surgery_primary_site`, `radiation_recode`, and `chemotherapy_recode`. The biomarker and treatment variables also form a stable structure, including relations among ER/PR/HER2 status and chemotherapy. Tumor size, T stage, N stage, grade, and treatment variables

remain highly connected in both domains, consistent with the expected clinical pathway from tumor burden and biomarkers to treatment selection.

The two graphs also contain domain-only edges. In the non-Hispanic domain, several race-related adjacencies appear, including edges involving `race_domain`, `er_status`, `treatment_delay_group`, and `income_group`. In the Hispanic domain, some treatment-delay and diagnosis-era relations differ, including edges involving `N_harmonized`, `treatment_delay_group`, `radiation_recode`, and `surgery_primary_site`. These differences suggest that the fine-grained treatment-selection graph is not identical across the two subcohorts, even though the main pathology-stage-treatment mechanisms are largely shared.

Overall, this experiment shows that the method identifies a small and interpretable set of domain-sensitive variables while preserving a large stable clinical backbone. The F-adjacent variables are concentrated in age, race, income, rural-urban context, treatment delay, radiation, and grade, which are precisely the kinds of variables expected to reflect access, screening, registry, and population-composition heterogeneity. In contrast, the stable edges are concentrated among biomarkers, tumor size, stage, surgery, radiation, and chemotherapy, matching the clinical structure of breast-cancer treatment selection. We emphasize that ethnicity-defined domains are observational and non-interventional. The F-adjacent variables should therefore be read as candidates for domain-sensitive mechanisms induced by social, clinical, and registry factors, not as innate biological differences between patient groups.

D.5 MIMIC-IV Insurance-Domain ICU Experiment

This experiment evaluates whether the proposed method can separate payer-sensitive demographic and care-pathway variables from more stable physiology and outcome relations in a real ICU cohort. We use MIMIC-IV (Johnson et al., 2024) and define domains by insurance group: Medicaid, Medicare, and private insurance. These domains are observational and should not be interpreted as interventions. Instead, they represent heterogeneous patient populations and care-access pathways.

We construct an adult ICU-stay cohort from MIMIC-IV after excluding stays with missing or other insurance labels. The resulting cohort contains 81,707 ICU stays, with 46,613 Medicare stays, 22,282 private-insurance stays, and 12,812 Medicaid stays. To avoid the full-cohort sample-size asymmetry from dominating the invariance tests, we use balanced subsampling and draw 12,000 stays from each insurance domain. All processing of row-level MIMIC-IV data is performed in the credentialed data environment.

We use a compact 12-variable feature set designed to capture demographics, care pathway, early ICU physiology, and outcomes:

- **Demographics and care pathway:** `age_bin`, `sex`, `first_icu_stay`, `unit_group`, and `admission_type_group`.
- **Outcomes:** `hospital_mortality` and `icu_los_bin`.
- **First-24h physiology:** `heart_rate_bin`, `map_bin`, `resp_rate_bin`, `spo2_bin`, and `temperature_bin`.

Continuous or high-cardinality variables are discretized into clinically interpretable bins before running the discrete conditional-independence and cross-domain invariance tests. We run five balanced subsampling seeds using $\alpha = 0.0015$ and maximum conditioning set size 3.

Figure 4 shows the learned inv-augmented graphs for one representative balanced run. Across five seeds, the selected setting has mean observed graph density 0.442 ± 0.008, mean F-adjacent fraction 0.467 ± 0.046, and 5.6 ± 0.55 F-adjacent variables out of 12. The F-adjacent set is highly reproducible and is concentrated in demographic and care-pathway variables. In all five seeds, the variables `admission_type_group`, `age_bin`, `first_icu_stay`, `sex`, and `unit_group` are F-adjacent. The variable `map_bin` is F-adjacent in three of five seeds. In contrast, `hospital_mortality`, `icu_los_bin`, `heart_rate_bin`, `resp_rate_bin`, `spo2_bin`, and `temperature_bin` are never F-adjacent in these balanced runs.

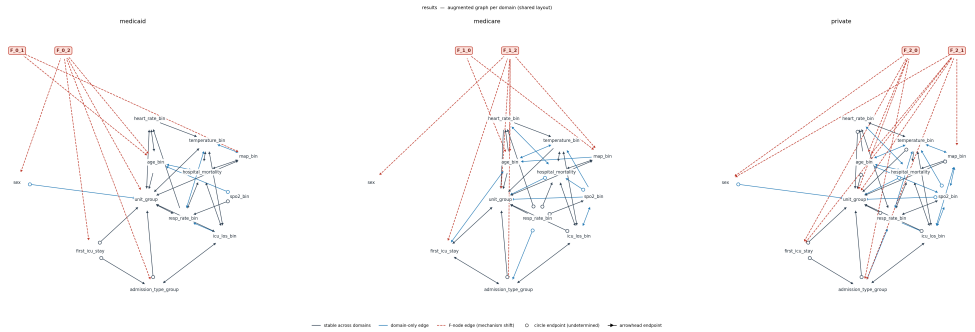


Figure 4: Learned inv-augmented graphs for the MIMIC-IV insurance-domain experiment. Panels correspond to Medicaid, Medicare, and private-insurance domains using a shared layout. Dark edges are stable across domains, blue edges appear only in one domain, and dashed red edges indicate F-node adjacencies, interpreted as candidate payer-sensitive mechanisms. Circle endpoints denote unresolved orientations.

The stable edges also form a clinically coherent graph backbone. The most reproducible stable relations include `admission_type_group` with `icu_los_bin`, `icu_los_bin` with `temperature_bin`, and relations between `hospital_mortality` and `icu_los_bin`, `map_bin`, or `temperature_bin`. Thus, the learned F-adjacencies are concentrated in variables related to age, sex, ICU unit, admission pathway, and first ICU stay, while mortality, length of stay, and early vital-sign relations are comparatively stable across the learned graphs.

The learned graphs support a care-access and cohort-composition interpretation. Insurance status is strongly associated with demographic and care-pathway differences: Medicare patients are older on average, private insurance is associated with different admission and unit patterns, and Medicaid comparisons may reflect different access pathways. The algorithm localizes these differences through F-node adjacencies to demographic and admission variables. At the same time, the stable backbone links admission type, ICU length of stay, mortality, blood pressure, and temperature. These stable relations are consistent with common ICU physiology and severity-of-illness pathways that persist across payer groups. As in the SEER experiment, we do not interpret the domains as interventions. The F-adjacent variables are candidates for domain-sensitive conditional mechanisms induced by patient mix, access, and workflow differences.

D.6 Computing Resources

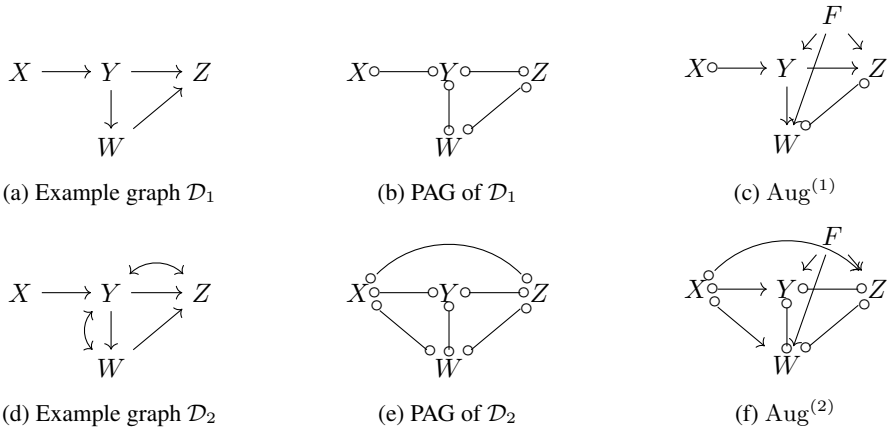
All experiments were run on a shared academic high-performance computing cluster. Each run was allocated 2 CPU cores and 16 GB of memory.

E Further Discussion

E.1 Comparison with Observational MEC

Figure 1 illustrates how cross-domain invariances can identify causal orientations that remain unresolved within either domain's observational equivalence class. Here, we give a complementary latent-confounded example in which several additional orientations can be recovered simultaneously.

Figure 5 illustrates the setting with latent confounding. For simplicity, we assume in this example that nodes with the same parent set have the same mechanisms. This assumption is used only for illustration; the general framework also permits mechanisms to change without a change in their parent sets. Figure 5a and Figure 5d show the causal graphs $\mathcal{D}_1$ and $\mathcal{D}_2$ for the two domains, respectively. The graph $\mathcal{D}_2$ contains two additional latent confounders, causing the mechanisms of Y, Z, and W to differ across domains. Figure 5b and Figure 5e show the corresponding observational PAGs. Within each domain, the observational CI statements identify the skeleton but do not provide any additional orientations. Figure 5c and Figure 5f show the corresponding inv-augmented PAGs. The node F points to Y, Z, and W because their mechanisms differ across domains. In both graphs, $\langle X, Y, F\rangle$ forms an unshielded collider, allowing us to orient $Xo\!\!\rightarrow Y$. In $\mathrm{Aug}^{(1)}$, FCI Rule 1 further orients $Y \rightarrow W$ and $Y \rightarrow Z$. In $\mathrm{Aug}^{(2)}$, Rule 0 orients $Xo\!\!\rightarrow W$ and $Xo\!\!\rightarrow Z$. These orientations



(a) Example graph $\mathcal{D}_1$ (b) PAG of $\mathcal{D}_1$ (c) $\mathrm{Aug}^{(1)}$

(d) Example graph $\mathcal{D}_2$ (e) PAG of $\mathcal{D}_2$ (f) $\mathrm{Aug}^{(2)}$

Figure 5: An example showing how cross-domain invariances provide orientations beyond the separate observational PAGs in the presence of latent confounding. Figure 5a and Figure 5d show the causal graphs $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively. Figure 5b and Figure 5e show their observational PAGs, while Figure 5c and Figure 5f show their inv-augmented PAGs.

remain unresolved in either observational PAG considered separately. Thus, the example shows how cross-domain invariances can refine observational equivalence under latent confounding. It does not imply uniform dominance over intervention-based equivalence classes, which rely on different targets and assumptions.

E.2 Comparison with CD-NOD

CD-NOD (Huang et al., 2020) is closely related to our work because both approaches exploit distributional heterogeneity to identify changing mechanisms and improve causal discovery. There is genuine overlap between the two frameworks. CD-NOD introduces an auxiliary context or time variable C and learns a single causal graph over $\mathbf{V} \cup C$. Adjacencies between C and observed variables indicate changing local mechanisms. It also permits causal links to become active or inactive across domains when the domain-specific models can be represented within a common acyclic causal template. In contrast, our method targets a tuple of domain-specific augmented PAGs with F-nodes,

$$\mathcal{L}(\mathbf{G}) = \big(\mathcal{P}^{(1)}, \ldots, \mathcal{P}^{(K)}\big),$$

where each $\mathcal{P}^{(i)}$ represents the causal structure in domain i, and F-node adjacencies encode pairwise mechanism-shift information.

The distinction is small when all domains share the same causal graph and only local mechanisms change. For example, suppose both domains have the graph $X \to Y \to Z$, but the mechanism of Y changes:

$$P^{(0)}(Y \mid X) \neq P^{(1)}(Y \mid X), \qquad P^{(0)}(Z \mid Y) = P^{(1)}(Z \mid Y).$$

CD-NOD may represent this by an adjacency between C and Y. Our method represents the same mechanism shift by an edge $F^{(0,1)} \to Y$ in the corresponding augmented graph. Thus, both approaches can identify that Y is domain-sensitive in this shared-graph setting, although their graphical targets differ.

Edge addition or deletion alone also does not separate the two frameworks. Consider Domain 1 with $X \to Y \to Z$ and Domain 2 with $X \to Y \to Z$ and $X \to Z$. CD-NOD can represent this variation through a common acyclic template containing $X \to Z$, with this link inactive in Domain 1 and active in Domain 2. Our method instead returns two domain-specific graphs, one without $X \to Z$ and one with $X \to Z$, together with the F-node adjacency $F^{(1,2)} \to Z$. The difference in this example is therefore primarily the learning target: CD-NOD summarizes the variation through one common template, whereas our representation records which graph applies in each domain.

The distinction becomes essential when the domain-specific directions are incompatible with a common acyclic template. Consider the edge reversal

$$\text{Domain 1: } X \to Y \to Z \leftarrow W, \qquad \text{Domain 2: } X \to Y \leftarrow Z \leftarrow W.$$

Our target can represent the two domain-specific structures as

$$\mathcal{P}^{(1)} : Xo\!\!\rightarrow Y \rightarrow Z \leftarrow\!\!oW, Y \leftarrow F \rightarrow Z, \qquad \mathcal{P}^{(2)} : Xo\!\!\rightarrow Y \leftarrow Z \leftarrow\!\!oW, Y \leftarrow F \rightarrow Z.$$

No common acyclic directed template contains both $Y \rightarrow Z$ and $Z \rightarrow Y$. Their unlabeled union contains a directed two-cycle and, moreover, does not record which direction occurs in which domain. A CD-NOD context graph may indicate that the mechanisms of Y and Z vary with C, but its existing single-DAG equivalence and recovery theory does not represent the statement that $Y \rightarrow Z$ holds in one domain while $Z \rightarrow Y$ holds in another.

There are at least two possible ways to extend CD-NOD to this case, and neither is immediate. One approach would introduce context-indexed, domain-specific DAGs, requiring a new tuple-level representation and corresponding equivalence and recovery theory. Alternatively, one could treat the union as a single cyclic causal template. Under this interpretation, however, the existing DAG-based separation and equivalence machinery would not apply directly. For general nonlinear cyclic SCMs, σ-separation provides the relevant graphical separation criterion (Forré & Mooij, 2018). With latent variables, further mixed-graph machinery is required; σ-MAGs provide a distinct representation and Markov-equivalence characterization for possibly cyclic systems (Yao & Mooij, 2025). Thus, permitting cycles in the common graph would require new causal semantics, separation conditions, equivalence results, and recovery guarantees rather than a purely syntactic modification.

The treatment of latent confounding also differs. Our framework permits general, potentially domain-specific latent confounding and characterizes its observable implications through latent projections, MAGs, and PAGs. CD-NOD's principal recovery guarantees instead use pseudo-confounders that are functions of the context or time index and are therefore fixed within each domain. Its discussion of stationary confounders considers particular configurations, but it does not provide a general MAG/PAG equivalence characterization or a recovery algorithm for tuples of domain-specific latent-variable graphs.

The heterogeneity experiment in Table 3 illustrates the practical consequences of these different targets. CD-NOD performs better when the domains are nearly shared, obtaining lower partial-graph SHD for one, two, and four modifications. As structural heterogeneity increases, Diverse-FCI becomes more effective: at 12 modifications, its partial-graph SHD is 59.12, compared with 65.52 for CD-NOD, and at 16 modifications the corresponding values are 59.80 and 67.37. These results do not establish a representational impossibility or uniform superiority. Rather, they are consistent with the intended scope of the methods: a common-template approach can benefit from stronger pooling when domains are nearly shared, whereas explicitly learning domain-specific graphs becomes advantageous as genuine structural heterogeneity increases.

In summary, CD-NOD and our method both exploit cross-domain distributional changes, but they have different graphical targets and assumptions. CD-NOD learns changing mechanisms relative to a common acyclic causal template, while our framework learns a reference-free tuple of domain-specific latent-variable graphs coupled through pairwise invariances.

E.3 Comparison with Unknown General Interventions

The unknown-general-intervention framework of Mascaro & Castelletti (2023) is closely related to ours in the causally sufficient setting. They consider DAGs and allow a general intervention to replace both the parent set and conditional mechanism of an unknown target. Thus, unlike standard hard or soft interventions on a fixed parent set, their framework can represent structural changes including edge additions, deletions, and reversals. They characterize the corresponding interventional Markov equivalence classes and develop a Bayesian procedure for jointly inferring the causal DAG, intervention targets, and induced parent sets.

To compare the frameworks directly, first restrict our setting to causally sufficient DAGs. Both approaches then allow the causal DAG to differ across domains and can exploit stability of unchanged mechanisms. The main distinction is how the domains are related. They assume a designated observational baseline DAG and model every other domain as the result of a valid general intervention on that baseline. Our formulation does not designate a reference domain or require intervention semantics; instead, all domains are treated symmetrically and related through pairwise mechanism invariances. The resulting learning targets also differ: their target includes the baseline DAG and

unknown intervention descriptions, whereas ours is a reference-free tuple of domain-specific graphs together with their pairwise invariance structure.

The distinction becomes clearer with three or more domains. Consider a variable Y whose causal mechanism is q_0 in domain 1 and the same alternative mechanism q_1 in domains 2 and 3, where $q_0 \neq q_1$. Suppose further that the shared mechanism between domains 2 and 3 admits a common observed conditioning set witnessing its stability. Then our pairwise representation records

$$Y \in I^{(1,2)}, \qquad Y \in I^{(1,3)}, \qquad Y \notin I^{(2,3)}. \tag{32}$$

Thus, our algorithm can directly test and exploit the invariance between domains 2 and 3.

If domain 1 is chosen as the observational baseline in the framework of Mascaro & Castelletti (2023), Y is treated as an intervention target in both domains 2 and 3. However, the two replacement mechanisms are modeled separately, and the intervention descriptions do not impose that they are both equal to q_1. Consequently, this equality between two non-reference mechanisms is not part of their baseline-relative intervention specification unless an additional sharing constraint is imposed, whereas our pairwise formulation represents it directly through $Y \notin I^{(2,3)}$. This does not preclude their method from recovering either post-intervention DAG using other information; the distinction is that the corresponding pairwise invariance between domains 2 and 3 is available to our algorithm but is not encoded by their model without an additional sharing constraint.

Beyond this causally sufficient comparison, our framework additionally permits general, potentially domain-specific latent confounding. Its observable targets are therefore represented through latent projections, MAGs, and PAGs. Mascaro & Castelletti (2023) study Markovian DAG models and do not provide a corresponding MAG/PAG equivalence characterization for latent-variable graph tuples.

These differences also explain why we do not include their Bayesian estimator as an additional baseline in our current experiments. Their accompanying implementation and all reported experiments use linear-Gaussian, causally sufficient DAG models with a designated observational baseline. Our synthetic evaluation instead uses binary domain-specific ADMGs with latent confounding and evaluates recovery of the corresponding domain-specific partial graphs. Applying the released estimator directly would therefore violate both its parametric likelihood and its causal-sufficiency assumption.

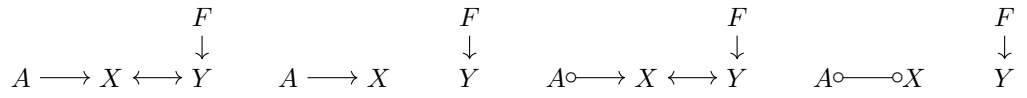
The graphical framework of Mascaro & Castelletti (2023) is not intrinsically restricted to continuous variables. As the authors note, a categorical version could in principle be constructed using a multinomial–Dirichlet model and an intervention-aware analogue of the BDeu score. However, this extension is left for future work and is not part of the released estimator. Developing and validating the required categorical score would therefore constitute a new implementation, and it would still not address the latent confounding present in our primary experiments.

Under its correctly specified parametric assumptions, the Bayesian approach may benefit from efficient pooling of unchanged mechanisms and may perform particularly well in the baseline-intervention setting. Our contribution addresses a different regime: reference-free causal discovery for tuples of domain-specific latent-variable graphs using pairwise invariances.

E.4 Why Algorithm 1 Is Incomplete under Latent Variables

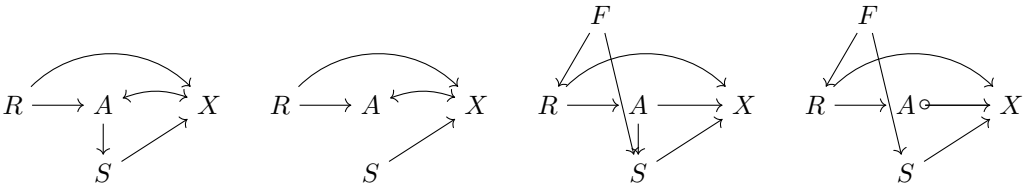
The completeness result for ordinary FCI is component-wise: for a fixed MAG, the output PAG contains exactly the invariant marks in that MAG Markov equivalence class (Zhang, 2008b). In our setting, however, the target is a tuple of domain-specific augmented PAGs. This introduces an additional joint constraint: the domain-specific augmented MAGs must arise from a single tuple of causal graphs whose unchanged mechanisms are shared across domains. Under latent confounding, this joint constraint can make some marks invariant in the tuple even though they are not invariant in one component considered alone.

We illustrate this with a two-domain example in Figure 6. Let the observed variables be A, X, Y, and suppose Y is the only variable in the mechanism-shift set, so $I^{(0,1)} = \{Y\}$. Thus, the causal mechanisms of A and X are shared across the two domains, and each admits a common observed conditioning witness. In domain 0, let the MAG over observed variables be $A \to X \leftrightarrow Y$, where the bidirected edge represents a latent common cause of X and Y. In domain 1, let the graph contain only $A \to X$, with no latent confounding between X and Y while the latent variable still points to X



(a) Domain-0 augmented MAG (b) Domain-1 augmented MAG (c) Learning of domain 0 (d) Learning of domain 1

Figure 6: Illustration of the incompleteness phenomenon under latent variables. Figures (a)–(b) show the two domain-specific augmented MAGs for the example in Appendix E.4, where only Y is adjacent to the F-node. In domain 0, the unshielded collider $A \to X \leftrightarrow Y$ implies an invariant arrowhead at X on the edge $A \ast\!\!-\!\!\ast X$, so the component-wise learned graph in (c) contains $Ao\!\!\to X$. In domain 1, the same edge is left unoriented as $Ao\!\!-\!\!oX$ when learned component-wise, as shown in (d). This illustrates that the current procedure is sound but may fail to propagate tuple-level invariant marks across domains in the presence of latent confounding.



(a) Domain 0 ADMG $\mathcal{G}^{(0)}$ (b) Domain 1 ADMG $\mathcal{G}^{(1)}$ (c) Learning of domain 0 (d) Learning of domain 1

Figure 7: A counterexample showing that one cannot simply copy a learned edge across domains even when both endpoints are non-adjacent to F. In both domains, there is a latent confounder between A and X. For this domain pair, the mechanism-shift set is $I^{(0,1)} = \{R, S\}$, so the mechanisms of A and X are stable. Nevertheless, changes involving the shifted variables alter the ancestral and latent-path structure associated with the observed A–X adjacency. Component-wise learning can orient $A \to X$ in domain 0, while in domain 1 it may identify only $Ao\!\!\to X$, with $A \leftrightarrow X$ remaining compatible with the domain-1 structure. Thus, copying the entire edge $A \to X$ from domain 0 to domain 1 is not valid.

invisibly. After hard intervention on Y, the changed incoming mechanism of Y is removed, and both domains share the same post-intervention core containing $A \to X$.

The corresponding domain-0 augmented graph contains $F^{(0,1)} \to Y, A \to X \leftrightarrow Y$. Since A and Y are non-adjacent, the triple $A \to X \leftrightarrow Y$ is an unshielded collider at X. Hence the arrowhead at X on the edge $A \ast\!\!-\!\!\ast X$ is invariant in the domain-0 augmented MAG's equivalence class.

By contrast, the domain-1 augmented graph contains $F^{(0,1)} \to Y, A \to X$, and Y is disconnected from A and X. If this domain-1 augmented MAG is considered in isolation, the edge between A and X cannot be oriented by any FCI rule. Thus the component-wise PAG for domain 1 leaves this edge unoriented, for example as $Ao\!\!-\!\!oX$. However, the endpoint mark is invariant in the full two-domain equivalence class. Since $I^{(0,1)} = \{Y\}$ and $A \to X$ is part of the common post-intervention core in this example, the corresponding relation between A and X is shared across the two domains. Therefore any valid graph pair in the inv-Markov equivalence class must use the same core relation between A and X in both domains. Domain 0 forces an arrowhead at X on the edge $A \ast\!\!-\!\!\ast X$, and the shared-core constraint transfers this invariant mark to domain 1. Hence, tuple-wise, the domain-1 edge should be oriented as $Ao\!\!\to X$, or more informatively if additional rules orient the other endpoint. The current algorithm, which applies FCI orientation rules separately to each domain-centric augmented graph, does not propagate this mark across domains. Therefore it is sound but not complete for the full inv-augmented PAG tuple in the presence of latent confounding.

In this example, the additional orientation information arises from the latent-confounding structure in domain 0. The bidirected edge $X \leftrightarrow Y$ changes the collider structure around the $A \ast\!\!-\!\!\ast X$ edge and reveals an endpoint mark that is not identified from the domain-1 augmented graph considered in isolation. Thus, latent confounding can create cross-domain orientation information that is invisible to a component-wise application of the standard FCI rules.

A tempting response to the preceding example is to copy learned edges between variables that are non-adjacent to F. This is not sound in general. Figure 7 gives a second example in which $A, X \notin I^{(0,1)}$ while $R, S \in I^{(0,1)}$. Changes involving R and S can alter the ancestral relations and inducing paths associated with the observed A–X adjacency without changing the local causal mechanisms of A or X.

As shown in Figure 7, the domain-0 inv-augmented graph contains enough orientation information for the component-wise procedure to recover $A \to X$. In domain 1, however, the learned graph may only recover $A \circ\!\!\to X$, as shown in Figure 7d. In particular, the domain-1 structure remains compatible with an arrowhead at A, i.e., $A \leftrightarrow X$. Therefore, copying the entire learned edge $A \to X$ from domain 0 would incorrectly impose a tail at A in domain 1.

This counterexample rules out naive copying of the entire edge, but it does not rule out more selective cross-domain propagation of endpoint marks. For example, the arrowhead at X is identified in both domains in Figure 7. It may therefore be possible to derive sound rules that propagate particular endpoint marks under additional graphical conditions.

A complete procedure would consequently require a graphical certificate specifying which orientation information is preserved across domains. In the earlier example of Figure 6, the common post-intervention structure provides additional tuple-level information that component-wise FCI fails to exploit. Figure 7 shows, however, that F-nonadjacency alone is insufficient to justify copying an entire edge. Characterizing the additional conditions under which individual endpoint marks can be propagated, and deriving a complete set of cross-domain orientation rules, are left for future work. These issues do not affect the soundness guarantee of Theorem 6.3.

F Limitations, Future Work, and Broader Impact

Limitations: Our characterization is given by a tuple of augmented graphs rather than a single unified graph, which makes the representation less compact and sometimes harder to interpret. In addition, although the proposed learning algorithm is sound, it need not be complete in the presence of latent confounding. The current framework also relies on cross-domain invariance tests, whose reliability may degrade in finite samples or when the mechanism shifts are weak.

Future Work: Several directions remain open. First, it would be valuable to further analyze the identifiability gains brought by diverse data and to understand under what conditions cross-domain invariances yield substantial improvements over purely observational discovery. Second, developing a complete learning algorithm, particularly in the presence of latent confounding, is an important theoretical direction. Finally, relaxing the modeling assumptions on mechanism stability and developing statistical tests tailored to this setting are important steps toward broader practical applicability.

Broader Impact: This work advances the foundations of causal discovery in settings with heterogeneous data, with potential applications in healthcare, biology, economics, and the social sciences, where data often come from multiple related but non-identical sources. By moving beyond intervention-based formulations, the proposed framework may broaden the applicability of causal discovery to more realistic settings in which controlled interventions are unavailable or difficult to justify.

At the same time, causal discovery methods can be misused if their assumptions are ignored or if estimated structures are treated as definitive causal conclusions. In sensitive applications, errors caused by limited data, hidden confounding, or model misspecification may lead to harmful downstream decisions. These risks highlight the importance of transparency, careful validation, and the use of domain expertise when applying such methods. Future work on uncertainty quantification, robustness, and responsible deployment will be essential for reducing these risks.