Data-Driven and Cycle-Agnostic Covariate Selection for Nonparametric Causal Effect Estimation

Ana Letícia Garcez Vicente*
Leiden University
Netherlands
a.l.garcez.vicente@liacs.leidenuniv.nl

Gijs van Seeventer*
Leiden University
Netherlands
g.g.van.seeventer@liacs.leidenuniv.nl

Saber Salehkaleybar
Leiden University
Netherlands
s.salehkaleybar@liacs.leidenuniv.nl

Abstract

Estimating causal effects from observational data requires identifying valid adjustment sets. This task is especially challenging in realistic settings where latent confounding and feedback loops are present. Existing approaches typically assume acyclicity or rely on global causal structure learning, limiting applicability and computational efficiency. In this work, we study a local, data-driven method for covariate selection based on conditional independence information under a pretreatment assumption. While this method is known to be sound and complete in acyclic causal models, its validity in the presence of cycles has remained unclear. Our main contribution is to show that these guarantees extend to cyclic causal models. In particular, our result relies on the invariance of conditional independence assertions under σ-acyclification. These findings establish a unified, cycle-agnostic perspective on covariate selection, which allows for causal effect estimation using any consistent estimator. Therefore, showing that the method applies across cyclic and acyclic settings without modification. Empirically, we validate this on extensive synthetic data, showing reliable performance in cyclic causal models.

1 Introduction

Estimating causal effects between variables is a central objective across a wide range of scientific disciplines. In epidemiology and pharmacology, for example, understanding the effect of a treatment on an outcome is critical for public health and drug development [1]. In economics, causal questions arise when evaluating policy interventions or market mechanisms [2]. The standard for such tasks is the randomized controlled trial (RCT), where treatments are randomly assigned to eliminate confounding bias [3]. However, RCTs are often infeasible due to ethical constraints, cost, or even the impossibility of manipulating certain variables [4].

*Equal contribution

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

As a result, there has been extensive interest in estimating causal effects from observational (non-experimental) data. A central challenge in this setting is confounding, i.e., spurious associations induced by common causes of treatment and outcome. A common approach to address confounding is covariate adjustment, where one conditions on a suitable set of variables, an adjustment set, to identify the causal effect [5]. Many such approaches are formulated in a non-parametric framework, avoiding restrictive assumptions on functional forms and enabling broad applicability across data-generating processes. For instance, it allows the use of machine-learning-based estimators, which are popular in practice [6].

Despite substantial progress [7, 8], existing literature on adjustment set identification assumes that the underlying causal graph is acyclic, and does not extend to cyclic settings [9]. This assumption excludes a broad class of real-world systems that exhibit feedback loops and cyclic dependencies that are considered relevant in the causal inference literature [10, 11]. Such cycles naturally arise in many domains, including biological regulatory networks (e.g., gene–protein feedback), economic systems (e.g., supply–demand interactions), and ecological systems, where variables influence each other over time.

To provide a more detailed example, consider a health care setting. In cancer studies, baseline covariates often include socioeconomic status, lifestyle behaviors, and health indicators [12]. These variables are not independent and may form feedback structures before treatment assignment. For example, consider the causal graph in Figure 1. Socioeconomic status can influence lifestyle. Lifestyle can shape health indicators, which may in turn influence socioeconomic status, such as employment. This creates feedback among covariates. At the same time, these covariates may affect both treatment assignment, such as access to screening or choice of therapy, and cancer recovery outcomes. Thus, the covariate set itself may form a cyclic structure that must be accounted for when selecting valid adjustment sets.

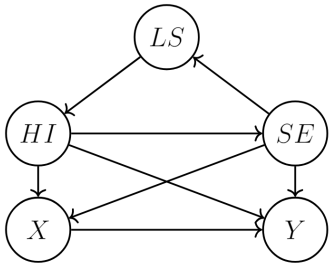


Figure 1: A causal graph of a health care setting. LS represents lifestyle, HI health indicators, SE socioeconomic status, X the treatment and Y the outcome.

This mismatch between modeling assumptions and real-world structure exposes a key limitation of existing methods: they do not extend to cyclic settings and therefore do not provide guarantees for covariate adjustment in the presence of feedback loops. In this paper, we address this gap by studying covariate selection in causal models that allow for cyclic dependencies. Our main contributions are: (i) we show that a local covariate selection method that allows for causal effect estimation [7], defined in a nonparametric setting under a pre-treatment assumption, remains valid in both acyclic and cyclic systems without requiring prior knowledge of the underlying structure; (ii) we extend theoretical guarantees of soundness and completeness for this method from acyclic to cyclic causal structures; (iii) we empirically validate these results through extensive simulation on synthetic data, comparing cyclic and acyclic settings under varying structural conditions.

2 Preliminaries and problem formulation

We begin by introducing structural causal models and their graphical representation. A structural causal model (SCM), defined as a tuple $\mathcal{M} = (\mathbf{V}, \mathbf{U}, F, P(\mathbf{U}))$, where $\mathbf{V}$ is the set of endogenous variables, $\mathbf{U}$ is the set of exogenous variables with joint distribution $P(\mathbf{U})$, and $F = \{f_i\}_{i=1}^{|\mathbf{V}|}$ is a collection of functions such that each endogenous variable satisfies a structural equation $X_i = f_i(Pa(X_i), U_i)$. We restrict attention to simple SCMs [13], meaning that for every subset of variables, the corresponding subsystem of structural equations admits a unique solution. This ensures well-defined observational and interventional distributions even in the presence of cycles, and guarantees that standard graphical and probabilistic semantics extend to cyclic settings. It is noteworthy to mention that all acyclic SCMs are simple.

An SCM induces a directed graph (DG), which may contain directed cycles. Let $G = (\mathbf{V}, \mathbf{E})$ be the directed graph associated with $\mathcal{M}$. We use the same symbol $\mathbf{V}$ to denote both the set of endogenous variables in the SCM and the corresponding node set in G, with each node representing one endogenous variable. The edge set $\mathbf{E} \subseteq \mathbf{V} \times \mathbf{V}$ contains a directed edge $X_i \to X_j$ whenever

X_i appears as an argument in the structural function for X_j. We partition the nodes as $\mathbf{V} = \mathbf{O} \cup \mathbf{L}$, where $\mathbf{O}$ denotes the set of observed variables, and $\mathbf{L}$ the set of latent (unobserved endogenous) variables. Moreover, we assume that the treatment and outcome are observable, i.e., $X, Y \in \mathbf{O}$. A DG is said to be acyclic if it contains no directed cycles, and it is called a directed acyclic graph (DAG); otherwise, it is cyclic.

A path between two nodes X and Y is a sequence $(X = Z_0, Z_1, \ldots, Z_k = Y)$ such that each consecutive pair of nodes is adjacent in G, and it is called directed if all edges are oriented consistently along the sequence. For any node $X \in \mathbf{V}$, we denote by $Pa(X) = \{Z \in \mathbf{V} \mid Z \to X \in \mathbf{E}\}$ the set of its parents, by $\mathrm{Ch}(X) = \{Z \in \mathbf{V} \mid X \to Z \in \mathbf{E}\}$ the set of its children, and define its ancestors $\mathrm{An}(X)$ and descendants $\mathrm{De}(X)$ as the sets of nodes that admit a directed path to X and from X, respectively. A strongly connected component (SCC) is a maximal subset of nodes $\mathbf{C} \subseteq \mathbf{V}$ such that for every $X, Y \in \mathbf{C}$ there exist directed paths from X to Y and from Y to X.

To read off conditional independence from a possibly cyclic graph, we adopt σ-separation [14].

Def. 2.1 (σ-separation)**.** *Suppose $G = (\mathbf{V}, \mathbf{E})$ is a DG, X and Y are two distinct variables in $\mathbf{V}$, and $\mathbf{S} \subseteq \mathbf{V} \setminus \{X, Y\}$. A path $\mathcal{P} = (X = Z_0, Z_1, \cdots, Z_k, Z_{k+1} = Y)$ between X and Y in G is σ-blocked by $\mathbf{S}$ if there exists $1 \leq i \leq k$ such that:*

1. *Z_i is a collider on $\mathcal{P}$ and $Z_i \notin An_G(\mathbf{S})$, or*
2. *(a) Z_i is not a collider on $\mathcal{P}$, $Z_i \in \mathbf{S}$, and*
   (b) either $Z_i \to Z_{i+1}$ and $Z_{i+1} \notin \mathrm{SCC}_G(Z_i)$, or $Z_{i-1} \leftarrow Z_i$ and $Z_{i-1} \notin \mathrm{SCC}_G(Z_i)$.

Since condition $2.(b)$ is always satisfied in DAGs, this definition reduces to standard d-separation [15] when G is acyclic. We write $X \perp\!\!\!\perp_{\sigma/d} Y \mid \mathbf{S}$ to denote σ/d-separation. We write $X \perp\!\!\!\perp Y \mid \mathbf{S}$ to denote conditional independence in the joint distribution $P(\mathbf{V})$, and assume the Markov and faithfulness properties in the following generalized form: for all disjoint sets $\mathbf{A}, \mathbf{B}, \mathbf{S} \subseteq \mathbf{V}$, $\mathbf{A} \perp\!\!\!\perp_\sigma \mathbf{B} \mid \mathbf{S}$ if and only if $\mathbf{A} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{S}$. In cyclic graphs this corresponds to σ-faithfulness, while in DAGs it reduces to standard faithfulness.

The notion of a Markov blanket depends on the underlying graph and separation criterion. Given a graph G, for a node $X \in \mathbf{O}$, the observed Markov Blanket, denoted by $\mathrm{MB}_\sigma^G(X)$, is the minimal set $\mathbf{S} \subseteq \mathbf{O} \setminus \{X\}$ such that $X \perp\!\!\!\perp_\sigma \mathbf{O} \setminus (\mathbf{S} \cup X) \mid \mathbf{S}$. When G is a DAG, this reduces to the standard (d-separation-based) Markov blanket, denoted by $\mathrm{MB}_d^G(X)$.

Throughout the paper, we impose the following assumptions: (i) We consider the pre-treatment assumption, which amounts to a treatment–outcome structure in which the outcome Y has no descendants, i.e., $\mathrm{De}(Y) = \emptyset$, and the treatment X has either no children or a single child equal to the outcome, i.e., $\mathrm{Ch}(X) = \emptyset$ or $\mathrm{Ch}(X) = \{Y\}$. (ii) We allow for latent confounding represented by unobserved endogenous variables $\mathbf{L}$, and assume that σ-faithfulness holds for the induced observational distribution restricted to $\mathbf{O}$.

Remark 1. *The pre-treatment assumption is common in the literature, see for instance Remark 1 in [7]. In fact, it is realistic in many application areas, such as economics and epidemiology, where covariates are often collected before treatment assignment and before the outcome is observed [16, 17, 18]. When not using the valid adjustment set, this can introduce bias [19, 20].*

Example: In a health care setting (Figure 1), variables such as socioeconomic status (SE) (e.g., pre-treatment employment), health indicators (HI) (e.g., prior hospitalizations), and lifestyle (LS) (e.g., pre-treatment physical activity) can be modeled as covariates measured solely prior to treatment assignment. Although these variables may influence one another dynamically over time, they are represented here as a pre-treatment snapshot and hence by time ordering not affected by the treatment X or outcome Y.

3 Related works

Adjustment sets and causal effect estimation Graphical criteria provide necessary and sufficient conditions for covariate adjustment in causal graphs. In DAGs, the back-door criterion [21, 15] characterizes when a set of variables $\mathbf{Z}$ is a valid adjustment, by requiring that all non-causal paths between treatment X and outcome Y are blocked and $\mathbf{Z}$ does not contain descendants of X. This

criterion has been generalized to characterize necessary and sufficient valid adjustment in equivalence classes of causal graphs with latent confounding, such as maximal ancestral graphs (MAGs) [22].

In practice, these criteria assume that the causal graph or the equivalence class is available. When the structure is unknown, a common approach is to first estimate an equivalence class (i.e., global structure) using constraint-based algorithms such as PC [23], and then perform adjustment. One such method is IDA [24], which enumerates all possible parent sets of treatment with the estimated graph and computes causal effect estimates. Building on this, LV-IDA [25] generalizes the approach to PAGs to account for latent confounding. In contrast, CE-SAT [26] avoids enumerating all graphs in the equivalence class by encoding the problem as a constraint task.

Complementary approaches aim to identify adjustment sets directly from conditional independence (CI) tests without recovering the equivalence graph first. Assuming that all variables are observed, [27] recently proposed LOAD, an algorithm that finds an optimal valid adjustment set. However, our main interest lies in the setting that allows for latent confounding. In that setting, the EHS algorithm [28] applies inference rules over CI to determine whether a causal effect is identifiable and to construct a valid adjustment set when it exists. EHS is sound and complete under the pre-treatment assumption, i.e., considering that covariates are not affected by the treatment. However, it requires an exhaustive search over subsets of variables. More recent methods aim to improve efficiency by exploiting locality. CEELS [8] uses structural patterns and COSO variables to guide a local search for adjustment sets. LSAS [7] shows that EHS can be restricted to search over the Markov blanket while remaining sound and complete. This reduction in the search space results in shorter run times and substantially fewer independence tests compared to EHS. On the other hand, LDP relaxes the pre-treatment assumption and introduces sufficient conditions, but is no longer complete [29]. Concurrent with our work, the LCS algorithm was proposed by [30]. LCS relaxes the pre-treatment assumption while being sound and complete. However, it assumes that the underlying causal graph is acyclic.

Overall, existing approaches either rely on global structure learning with high computational cost, or perform local search with some trade-offs between efficiency and completeness, but exclusively assume acyclicity.

Markov blankets Local causal discovery methods aim to identify subsets of variables relevant to a target using CI tests, without recovering the full causal graph. Algorithms such as (Fast)-IAMB [31, 32], HITON-MB [33] and TC [34] identify the Markov blanket set via CI tests. From the perspective of finding valid adjustment sets, the Markov blankets efficiently provide supersets of the relevant variables (treatment and outcome), but do not directly compute valid adjustment sets, since they may include colliders on paths between the treatment and the outcome or descendants of the treatment.

Cyclic causal models Results on covariate adjustment set assume acyclicity, where d-separation characterizes conditional independence. However, causal models with cycles require more general notions of separation. [14] introduced σ-separation, a generalization of d-separation for directed cyclic graphs, accounting for strongly connected components dependence. Building on this, [13] defined the class of simple SCMs and established equivalence notions for cyclic graphs. Furthermore, [35] introduces the notion of acyclicification, which maps a cyclic graph to an acyclic one while preserving relevant conditional independence relations . In addition, [35] shows that causal discovery algorithms like FCI work for simple cyclic SCMs. These developments provide a foundation for reasoning about conditional independence and equivalence in cyclic systems. However, to the best of our knowledge, they have not been utilized in designing covariate selection algorithms in cyclic models.

4 Acyclification

To extend graphical criteria from acyclic to cyclic models, we rely on the notion of σ-acyclification [35], which removes directed cycles while preserving separation properties.

Def. 4.1 (σ-acyclification)**.** *Let $G = (\mathbf{V}, \mathbf{E})$ be a directed graph, possibly cyclic. An acyclification of G is a directed acyclic graph $G^{\mathrm{acy}} = (\mathbf{V}, \mathbf{E}^{\mathrm{acy}})$ defined on the same node set $\mathbf{V}$ and constructed as follows. For any pair of distinct nodes $X, Y \in \mathbf{V}$:*

1. *if $X \notin \mathrm{SCC}(Y)$, then $X \to Y \in \mathbf{E}^{\mathrm{acy}}$ if and only if $\exists\, W \in \mathrm{SCC}(Y)$ such that $X \to W \in \mathbf{E}$;*

2. *if $X \in \mathrm{SCC}(Y)$, then exactly one of the edges $X \to Y$ or $Y \to X$ is included in $\mathbf{E}^{\mathrm{acy}}$ so as to break all directed cycles.*

Intuitively, all nodes in a SCC are made adjacent in an acyclic manner, while edges between different components are preserved by connecting every node in the source component to every node in the target component. By construction, G^{acy} is a DAG, and the acyclification is not necessarily unique.

A key property of this construction is that it preserves separation relations. In particular, σ-separation in G coincides with d-separation in any σ-acyclification G^{acy}, allowing graphical arguments in cyclic models to be reduced to the acyclic case.

Lemma 4.1 (Acyclification transfer of separation [35])**.** *Let $G = (\mathbf{V}, \mathbf{E})$ be a directed graph and let G^{acy} be any σ-acyclification of G. Then for all distinct $A, B \in \mathbf{V}$ and all $\mathbf{Z} \subseteq \mathbf{V} \setminus \{A, B\}$,*

$$(A \perp\!\!\!\perp_d B \mid \mathbf{Z})_{G^{\mathrm{acy}}} \quad \Longleftrightarrow \quad (A \perp\!\!\!\perp_\sigma B \mid \mathbf{Z})_G.$$

An immediate consequence of Lemma 4.1 is that local separation-based objects are invariant under acyclification. In particular, Markov blankets coincide in the cyclic and acyclic representations.

Lemma 4.2 (Markov blanket invariance)**.** *Let G be a directed graph and G^{acy} any σ-acyclification. Then for any node Y,*

$$\mathrm{MB}_\sigma^G(Y) = \mathrm{MB}_d^{G^{\mathrm{acy}}}(Y).$$

The proofs of all lemmas and theorems are given in the appendix. We next show that acyclification can be chosen so as to preserve specific local edge structures.

Lemma 4.3 (Preservation of the edge $X \to Y$ under σ-acyclification)**.** *Let $G = (\mathbf{V}, \mathbf{E})$ be a directed graph, possibly cyclic, and let $X, Y \in \mathbf{V}$ be distinct nodes. Assume that $\mathrm{De}(Y) = \emptyset$, and $\mathrm{De}(X) \subseteq \{Y\}$. Then all σ-acyclification $\hat{G}^{\mathrm{acy}} = (\mathbf{V}, \mathbf{E}^{\mathrm{acy}})$ such that Y remains a sink node (has no children), no node other than Y can be a child of X, i.e., $\mathrm{De}(X) \subseteq \{Y\}$, and the edge $X \to Y$ is present in $\mathbf{E}^{\mathrm{acy}}$ if and only if it is present in $\mathbf{E}$.*

In the proofs of the remaining theorems and lemmas, we repeatedly use the same strategy: map a possibly cyclic directed graph G to (all) acyclifications G^{acy} using results from this section, apply existing results for DAGs, and then transfer the conclusions back to G.

5 Adjustment set

A common way to identify the causal effect of X on Y using solely observational data is based on using a valid adjustment set [15, 9].

Def. 5.1 (Valid adjustment set)**.** *Let $G = (\mathbf{V}, \mathbf{E})$ be a directed graph, and $\{X\}, \{Y\}, \mathbf{Z} \subseteq \mathbf{O}$ be pairwise disjoint node sets. The set $\mathbf{Z}$ is called an* adjustment set *for estimating the causal effect of X on Y if, for all probability distributions P Markov with respect to G, the following holds:*

$$P(Y \mid \mathrm{do}(X)) \;=\; \sum_{z \in \mathbf{Z}} P(Y \mid X, \mathbf{Z} = z)\, P(\mathbf{Z} = z).$$

The invariance results in Section 4 allow us to reduce valid adjustment questions in cyclic graphs to the acyclic case. For DAGs, there exists a valid adjustment set known as the backdoor adjustment set formulated in terms of d-separation [15, 9]. Replacing d-separation by σ-separation, we get a backdoor adjustment set in possibly cyclic SCMs.

Def. 5.2 (Backdoor adjustment set in possibly cyclic SCMs)**.** *Let $G = (\mathbf{V}, \mathbf{E})$ be a (possibly cyclic) directed graph satisfying the assumptions of Section 2, and let $X, Y \in \mathbf{O}$ be treatment and outcome variables. A set $\mathbf{Z} \subseteq \mathbf{O} \setminus \{X, Y\}$ is a backdoor adjustment set for the effect of X on Y if the following conditions hold:*

1. $\mathbf{Z} \cap \mathrm{De}(X) = \emptyset$;

2. *$\mathbf{Z}$ σ-blocks every backdoor path from X to Y (i.e., every path between X and Y whose first edge on the path has an arrowhead into X).*

If the SCM is acyclic, this reduces to the standard definition. Therefore, in DAGs, Definition 5.2 of backdoor adjustment provides valid adjustment sets. To show that the definition gives a valid adjustment set in simple cyclic SCMs, we can use results from [36]. To use their results, we model interventions by augmenting the graph with an intervention node [15, 9, 36]. For a treatment variable X, we define G^{I_X} as the graph obtained by adding a new node I_X and a single directed edge $I_X \to X$. The node I_X encodes the external intervention $\mathrm{do}(X)$ and has no parents. With this notation, we can show the equivalence of their formulation and Definition 5.2.

Lemma 5.1 (Equivalence with the intervention-node criterion)**.** *Let G be a (possibly cyclic) SCM and $\mathbf{Z} \subseteq \mathbf{V} \setminus \{X, Y\}$. Let G^{I_X} denote the graph obtained by adding an intervention node I_X with the single directed edge $I_X \to X$. Then $\mathbf{Z}$ is a valid backdoor adjustment set in the sense of Definition 5.2 if and only if*

1. $\mathbf{Z} \perp\!\!\!\perp_\sigma I_X$ *in* G^{I_X}*, and*
2. $Y \perp\!\!\!\perp_\sigma I_X \mid X, \mathbf{Z}$ *in* G^{I_X}*.*

Using Lemma 5.1, we can claim that, by Corollary 8.4 from [36], Definition 5.2 is indeed a valid adjustment set for simple cyclic SCMs. Furthermore, with this result, we can show a useful property of backdoor adjustment sets with respect to Markov blankets, which was shown for acyclic SCMs by [7].

Lemma 5.2 (Existence of an adjustment set)**.** *Let $G = (\mathbf{V}, \mathbf{E})$ be a simple (possibly cyclic) directed graph satisfying the assumptions of Section 2. Then there exists a subset $\mathbf{Z} \subseteq \mathbf{O}$ that is a backdoor adjustment set for (X, Y) if and only if there exists a subset $\mathbf{Z} \subseteq \mathrm{MB}^G_\sigma(Y) \setminus \{X\}$ that is an adjustment set for (X, Y).*

6 Algorithm

We now introduce two rules, $R1$ and $R2$, that characterize the presence or absence of a direct causal effect from X to Y using only conditional independence relations restricted to the Markov blanket of Y, given the assumptions in Section 2. The rules are related to the causal estimation criteria proposed in [28, 7] for acyclic models. Our contribution extends these works to possibly cyclic SCMs.

Def. 6.1 (R1)**.** *Let $G = (\mathbf{V}, \mathbf{E})$ satisfy the assumptions of Section 2. Suppose there exists a variable $W \in \mathrm{MB}^G_\sigma(X) \setminus \{Y\}$ and a set $\mathbf{Z} \subseteq \mathrm{MB}^G_\sigma(Y) \setminus \{X, W\}$ such that*

$$W \not\perp\!\!\!\perp Y \mid \mathbf{Z} \quad \textit{and} \quad W \perp\!\!\!\perp Y \mid \mathbf{Z} \cup \{X\}.$$

Then it returns $\mathbf{Z}$ as a valid backdoor adjustment set for (X, Y), and that $X \to Y \in \mathbf{E}$.

To provide some insight regarding Definition 6.1, since Y has no descendants, the first condition ($W \not\perp\!\!\!\perp Y \mid \mathbf{Z}$) states that there is some open path from W to Y, despite the presence of $\mathbf{Z}$. At the same time, this path can only go through X to Y, since adding X to the conditioning set closes the path by the second condition ($W \perp\!\!\!\perp Y \mid \mathbf{Z} \cup \{X\}$). Therefore, $\mathbf{Z}$ closes the backdoor of X, and there has to be a causal effect from X to Y. Restricting $\mathbf{Z}$ and W to be inside the Markov blankets is mainly for reducing the search space.

Def. 6.2 (R2)**.** *Let $G = (\mathbf{V}, \mathbf{E})$ satisfy the assumptions of Section 2. If either of the following holds:*

1. $\exists\, \mathbf{Z} \subseteq \mathrm{MB}^G_\sigma(Y) \setminus \{X\}$ *such that* $X \perp\!\!\!\perp Y \mid \mathbf{Z}$*,*
2. $\exists\, W \in \mathrm{MB}^G_\sigma(X) \setminus \{Y\}$ *and* $\mathbf{Z} \subseteq \mathrm{MB}^G_\sigma(Y) \setminus \{X, W\}$ *such that* $W \not\perp\!\!\!\perp X \mid \mathbf{Z}$ *and* $W \perp\!\!\!\perp Y \mid \mathbf{Z}$*,*

then it returns that $X \to Y \notin \mathbf{E}$.

For Definition 6.2, (1) is trivial in the sense that if there exists a set that separates X and Y, then by σ-faithfulness, there is no causal effect. In (2), the first condition ($W \not\perp\!\!\!\perp X \mid \mathbf{Z}$) ensures that there is an active path from W to X despite the presence of $\mathbf{Z}$. The second condition ($W \perp\!\!\!\perp Y \mid \mathbf{Z}$) ensures there is no active path from W to Y given $\mathbf{Z}$. If there were an edge from X to Y, one of these statements would be false. The Markov blankets are again used for shrinking the search space for $\mathbf{Z}$ and W.

Note that the formulations and explanations for both $R1$ and $R2$ are already agnostic to the nature of the underlying paths; without further specification, this reasoning seems to allow both cyclic and acyclic causal structures. To prove this, we need the following lemma.

Lemma 6.1 (Invariance of witness variables under σ-acyclification)**.** *Let G be a (possibly cyclic) directed graph over $\mathbf{V}$ satisfying σ-faithfulness, and let G^{acy} be a σ-acyclification of G. Let $X, Y \in \mathbf{O}$ and $W \in \mathbf{O} \setminus \{X, Y\}$. Then, there exists $W \in \mathrm{MB}_\sigma^G(X) \setminus \{Y\}$ satisfying the R1/R2 conditions in G if and only if there exists a $W \in \mathrm{MB}_\sigma^{G^{\mathrm{acy}}}(X) \setminus \{Y\}$ satisfying the R1/R2 conditions in G^{acy}.*

With the help of Lemma 6.1, our main result is that the rules remain sound and complete in possibly cyclic SCMs.

Theorem 6.2 (Soundness and completeness in possibly cyclic SCMs)**.** *Under our assumptions in Section 2, Let G be a simple SCM with treatment-outcome pair (X, Y). Then R1 and R2 are sound. Moreover, if neither R1 nor R2 applies, the causal effect of X on Y is not identifiable from the conditional independence relations.*

Remark 2. *If neither R_1 nor R_2 applies, then the available CI relations are compatible with models that disagree on the interventional distribution of Y under $do(X = x)$. In one model, X may have a causal effect on Y, in this case the estimand may be an adjustment functional of the form $P(Y \mid do(X = x)) = \sum_z P(Y \mid X = x, Z = z)P(Z = z)$. In another compatible model, X may have no causal effect on Y, i.e. $P(Y \mid do(X = x)) = P(Y)$. Therefore, from CI information alone, one cannot decide which expression correctly represents the interventional distribution. This is the sense in which the rules are complete: failure of both rules means that no CI-based criterion can determine the causal effect.*

Unlike in the acyclic setting, where every SCC is a singleton, cyclic graphs can contain nontrivial SCCs. It is therefore natural to ask whether $R1$ and $R2$ remain applicable when the pre-treatment assumption (Assumption (i) in Section 2) is relaxed with respect to $\mathrm{SCC}(X)$ and $\mathrm{SCC}(Y)$. Specifically, suppose that $\mathrm{De}(\mathrm{SCC}(X)) \setminus \mathrm{SCC}(X) \subseteq \mathrm{SCC}(Y)$ and $\mathrm{De}(\mathrm{SCC}(Y)) \setminus \mathrm{SCC}(Y) = \emptyset$. We refer to this relaxed condition as Assumption $(i)'$. However, since nodes within an SCC are mutually reachable, if $|\mathrm{SCC}(X)| \geq 2$ or $|\mathrm{SCC}(Y)| \geq 2$ in a simple SCM G, then X or Y has descendants within its own SCC. Consequently, there may exist acyclifications of G for which the acyclic pre-treatment assumption is violated, and by extension no longer satisfy the conditions for the acyclic setting in which the rules were proposed. It is therefore not immediately clear whether the combined rules $R1$ and $R2$ remain sound and complete under this relaxation. In fact, the following lemma shows that $R1$ is no longer sound for simple cyclic SCMs under the specified condition.

Lemma 6.3. *Let G be a simple SCM under Assumption (ii) in Section 2 and the relaxed Assumption $(i)'$. If $Y \notin \mathrm{SCC}(X)$ and $|\mathrm{SCC(X)}| \geq 2$, then $R1$ is not sound.*

6.1 Local search of adjacent nodes in cyclic models

The local component of this method is the identification of Markov blankets using conditional independence tests. Existing Markov blanket discovery algorithms, are usually defined under the assumption of acyclic graphs. At first glance, this raises the question of whether these algorithms remain valid in cyclic settings. Our result addresses this concern by showing that Markov blanket identification depends only on the underlying conditional independence structure, which is invariant under σ-acyclification.

Theorem 6.4 (Validity of DAG Markov blanket algorithms in cyclic SCMs)**.** *Let G be a (possibly cyclic) directed graph over $\mathbf{V}$. Assuming σ-faithfulness, any sound Markov blanket discovery algorithm for DAGs that relies on observed conditional independence tests returns the observed Markov blanket $\mathrm{MB}_\sigma^G(X)$ when applied to a distribution induced by G.*

Remark 3. *The theorem does not assert that the underlying causal graph is acyclic. Rather, it shows that Markov blanket discovery depends only on the conditional independence assertions, which is preserved under σ-acyclification. Consequently, DAG-based algorithms recover the correct Markov blanket even when the true data-generating process contains cycles.*

7 Experimental results

In this section, we empirically evaluate the accuracy, robustness, and efficiency of the local adjustment method on simple (possibly cyclic) SCMs. Our source code is available via this link.

Data generation We generate synthetic datasets from randomly sampled SCMs with latent confounding and directed cycles, and parameterize using both linear and nonlinear SCMs. We consider: linear SCMs: $\mathbf{V} := W\mathbf{V} + \mathbf{U}$, and nonlinear SCMs: $\mathbf{V} := tanh(W\mathbf{V}) + \mathbf{U}$, where W is the weighted adjacency matrix and $\mathbf{U}$ is the vector of exogenous noises. All models are restricted to simple SCMs, ensuring well-defined solutions despite the presence of cycles. Additional details can be found in Appendix D.

We vary the structural properties of the SCMs. We generate six graph sizes with 8, 15, 25, 50, 100, and 250 nodes, and 12, 19, 40, 78, 150, and 360 edges, respectively. Moreover, the number of samples is $1k$, $5k$, $10k$, and $15k$ per graph. Latent confounding is introduced by randomly hiding a subset of variables (respectively, by the number of nodes, we make the following number of variables latent per graph size: 2, 3, 4, 10, 10, and 30), always excluding the treatment and outcome. Exogenous noises are sampled independently with standard deviations drawn from $\mathcal{U}(0.5, 1.0)$, under three settings: Gaussian noise, uniform noise (i.e., non-Gaussian noise), or mixed (node-wise random assignment to either Gaussian or uniform noise).

To probe different regimes, we generate datasets by combining four factors: graph structure (acyclic vs. cyclic), functional form (linear vs. nonlinear), edge condition (presence or absence of $X \to Y$), and noise type (Gaussian, non-Gaussian, or mixed).

For each configuration, we sample 25 independent graphs per noise type, resulting in 75 graphs per (graph structure, functional form, edge condition) setting. Overall, this yields 600 graphs per graph size. This total decomposes as follows: 300 graphs per functional form; within each functional form, 150 graphs per graph structure; and within each graph structure, 75 graphs per edge condition. We repeat this procedure for each graph size.

Markov blankets The Markov blankets are estimated using the TC [34] for graphs with up to 50 nodes due to the limited scalability, and Fast-IAMB [32] for larger graphs, which we select based on its favorable trade-off between CI tests and computational time. We also study the impact of using other Markov blanket discovery algorithms, such as IAMB [31] and HITON-MB [33] (see Appendix A for a detailed comparison).

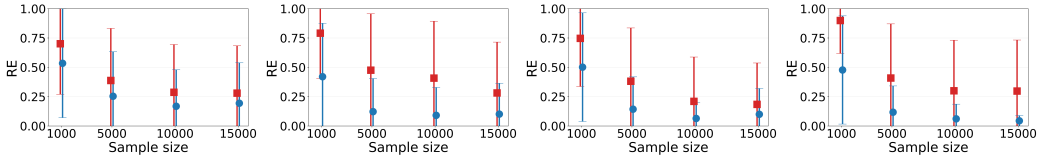
Evaluation metrics Let $\mathbf{Z}$ be the set of all valid adjustment sets (ground truth), and $\hat{\mathbf{Z}}$ the set of all adjustment sets returned by the method. Let $e \in \{0, 1\}$ indicate the presence of an edge $X \to Y$ in the true graph, and $\hat{e}$ the output of the method. Let n be the number of graphs where $R1$ or $R2$ applies. We evaluate performance using: Relative Error (RE), defined as $RE = \frac{|\hat{CE} - CE|}{|CE|}$; Precision, defined as $Pre = \frac{|\hat{\mathbf{Z}} \cap \mathbf{Z}|}{|\hat{\mathbf{Z}}|}$; and Edge Fraction accuracy (EF), defined as $EF = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}[e_i = \hat{e}_i]$.

Relative error quantifies the accuracy of the estimated causal effect ($\hat{CE}$) and is reported only when $R1$ applies in linear SCMs (see more details in Appendix D.1.2). Precision, i.e., the proportion of valid adjustment sets returned, is also computed under $R1$. The edge fraction accuracy evaluates performance across both $R1$ and $R2$ by measuring agreement on the presence of a causal edge. We emphasize that precision and edge fraction accuracy are not standard metrics in this context. However, we include them to enable evaluations in nonlinear settings, where the RE does not apply. The non-linear causal effect depends on the specific treatment value and, therefore, cannot be quantified by a single scalar. Finally, we report the fraction of empty instances (neither $R1$ nor $R2$ applies), which reflects cases where the method is unable to return a definitive output (see Appendix F.3).

Results Since existing local adjustment methods are derived under acyclicity assumptions, they do not provide guarantees in cyclic SCMs. We therefore study how the method behaves when extended to cyclic settings by comparing to the acyclic case as a baseline.

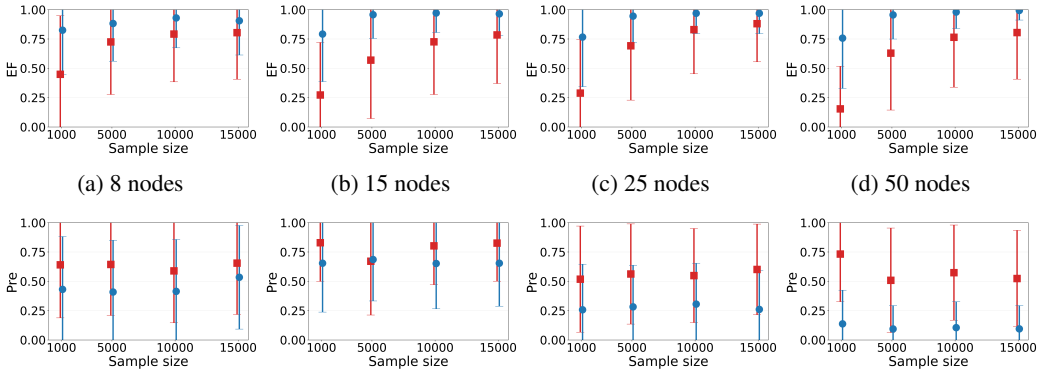
All the graphs discussed in this section are constructed such that they satisfy the pre-treatment assumption (see Appendix F.2 for the setting where this assumption is violated).

Figure 2 reports RE for linear SCMs across cyclic and acyclic settings. Increasing the sample size improves estimation accuracy in both settings. More interestingly, RE is lower in cyclic graphs compared to acyclic ones, often with smaller variance. This is notable because the original local adjustment method was developed for acyclic models. Moreover, the results empirically support our theoretical claim that the method extends naturally to cyclic SCMs.



(a) 8 nodes (b) 15 nodes (c) 25 nodes (d) 50 nodes

Figure 2: Relative error of linear models in acyclic (red square) and cyclic (blue circle) settings across sample sizes. Markers show the mean, with error bars indicating one standard deviation. Subfigures correspond to graph sets with different numbers of nodes.



(a) 8 nodes (b) 15 nodes (c) 25 nodes (d) 50 nodes

(e) 8 nodes (f) 15 nodes (g) 25 nodes (h) 50 nodes

Figure 3: Edge fraction accuracy (first row) and precision (second row) of linear and non-linear models in acyclic (red square) and cyclic (blue circles) settings across sample sizes. Markers show the mean, with error bars indicating one standard deviation. Subfigures correspond to graph sets with different numbers of nodes.

Figure 3 reports the edge fraction accuracy across both linear and nonlinear SCMs. Similar to the RE results, performance improves monotonically with increasing sample size. Cyclic graphs consistently achieve higher EF compared to acyclic graphs, often approaching near-perfect recovery in larger sample regimes.

The precision results reveal a different behavior. While cyclic settings achieve lower RE and EF, acyclic graphs in general obtain higher precision. Notable, this reduction in precision does not translate into worse causal estimation or edge detection; cyclic configurations still achieve better RE and EF overall. Furthermore, the sample size seems to have little effect on the precision.

Finally, the empty fraction, the proportion of graphs for which the algorithm does not apply either rule, remains in general below 20% for the cyclic setting, while in the acyclic case, it usually (with an exception for the 15 nodes configuration) stays around 25% (see more details in Appendix F.3).

Additional experiments on larger graphs reported in Appendix F.1 show the same qualitative trends, suggesting that the observed behavior is stable across graph scales.

Discussion Our experiments reveal two important phenomena. First, the proposed local adjustment method remains effective in cyclic SCMs in terms of the relative error, the standard metric in the literature. This empirically supports the theoretical results established in this work. Second, cyclic settings often exhibit lower relative error and improved edge recovery compared to acyclic settings, while simultaneously having lower precision. The mechanism underlying this trade-off is unclear. Since our goal is to show that existing methods can provide cycle-agnostic causal effect estimation,

we leave a theoretical explanation of this phenomenon to future work. However, we do note that precision is a very sensitive metric, where the addition or absence of a single variable relative to a true valid adjustment set is already penalized as heavily as a set without any overlap. Hence, while this metric is easy to interpret, it seems too sensitive to align well with causal effect estimation performance.

Regarding the consistency of our implementation's performance in acyclic settings relative to prior work, we make two observations. To begin, the experiments exhibit relatively large standard deviations. While this variability may initially appear concerning, it is consistent with prior observations reported by [28], who similarly reported substantial variance. This suggests that the instability is not unique to our implementation, but may instead reflect inherent variance in the problem. To continue, for acyclic graphs, the average relative error aligns with the results reported by [7], supporting the validity of our implementation and evaluation protocol. In contrast, for cyclic graphs, direct comparisons with other works and methods are not possible due to the absence of prior work.

8 Conclusion

We addressed the problem of covariate selection for nonparametric causal effect estimation in the presence of cycles and latent variables. While existing approaches rely on acyclicity assumptions, many real-world scenarios exhibit feedback loops that violate the requirement. In this work, we show that a local, data-driven covariate selection method that allows causal effect estimation under a pre-treatment assumption [7] remains valid in cyclic settings.

Our main theoretical contribution establishes that the soundness and completeness guarantees of this method continue to hold in simple structural causal models. Notably, this result holds in a nonparametric setting, requiring no assumptions about the presence or absence of cycles. As a consequence, it is cycle-agnostic and applies uniformly to both acyclic and cyclic systems.

Empirically, we validated our approach on a wide range of synthetic data, including linear and nonlinear models with varying graph sizes, noise levels, and cyclic and acyclic cases. The results demonstrate consistent performance across settings, with notably strong accuracy and stability in cyclic regimes.

An important direction for future work is to better understand the empirical and theoretical differences between cyclic and acyclic settings, particularly the apparent trade-off between estimation accuracy and edge recovery versus adjustment-set precision. In addition, we could explore finding the optimal set among the recovered valid adjustment sets. Finally, extending local adjustment methods beyond the pre-treatment assumption while preserving soundness and completeness also remains an important open problem in the cyclic setting.

References

[1] Miguel A Hernán and James M Robins. Estimating causal effects from epidemiological data. *Journal of Epidemiology & Community Health*, 60(7):578–586, 2006.

[2] Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press, 2009.

[3] Al K Akobeng. Understanding randomised controlled trials. *Archives of disease in childhood*, 90(8):840–844, 2005.

[4] Paul Rosenbaum. *Observation and experiment: An introduction to causal inference*. Harvard University Press, 2017.

[5] Elias Bareinboim and Judea Pearl. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352, 2016.

[6] Denis Talbot, Awa Diop, Miceline Mésidor, Yohann Chiu, Caroline Sirois, Andrew J. Spieker, Antoine Pariente, Pernelle Noize, Marc Simard, Miguel Angel Luque Fernandez, Michael Schomaker, Kenji Fujita, Danijela Gnjidic, and Mireille E. Schnitzer. Guidelines and best practices for the use of targeted maximum likelihood and machine learning when estimating causal effects of exposures on time-to-event outcomes. *Statistics in Medicine*, 44(6), March 2025.

[7] Zheng Li, Xichen Guo, Feng Xie, Yan Zeng, Hao Zhang, and Zhi Geng. Local learning for covariate selection in nonparametric causal effect estimation with latent variables. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen, editors, *Advances in Neural Information Processing Systems*, volume 38, pages 124423–124455. Curran Associates, Inc., 2025.

[8] Debo Cheng, Jiuyong Li, Lin Liu, Jiji Zhang, Jixue Liu, and Thuc Duy Le. Local search for efficient causal effect estimation. *IEEE Transactions on Knowledge and Data Engineering*, 35(9):8823–8837, 2022.

[9] Jonas Peters, Dominik Janzing, and Bernhard Scholkopf. *Elements of causal inference: foundations and learning algorithms*. MIT press, 2017.

[10] Judea Pearl. A causal calculus for statistical research. In Doug Fisher and Hans-Joachim Lenz, editors, *Pre-proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics*, volume R0 of *Proceedings of Machine Learning Research*, pages 430–449. PMLR, 04–07 Jan 1995. Reissued by PMLR on 01 May 2022.

[11] Haoyue Dai, Immanuel Albrecht, Peter Spirtes, and Kun Zhang. Distributional equivalence in linear non-gaussian latent-variable cyclic causal models: Characterization and learning. In *The Fourteenth International Conference on Learning Representations*, 2026.

[12] Christopher P Wild, Elisabete Weiderpass, and Bernard W Stewart. World cancer report. 2020.

[13] Stephan Bongers, Patrick Forré, Jonas Peters, and Joris M Mooij. Foundations of structural causal models with cycles and latent variables. *The Annals of Statistics*, 49(5):2885–2915, 2021.

[14] Patrick Forré and Joris M Mooij. Markov properties for graphical models with cycles and latent variables. *arXiv preprint arXiv:1710.08775*, 2017.

[15] Judea Pearl. *Causality*. Cambridge university press, 2009.

[16] Jennifer L Hill. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240, 2011.

[17] Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge university press, 2015.

[18] Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.

[19] Jouni Helske, Santtu Tikka, and Juha Karvanen. Estimation of causal effects with small data in the presence of trapdoor variables. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 184(3):1030–1051, May 2021.

[20] Felix Thoemmes. M-bias, butterfly bias, and butterfly bias with correlated causes – a comment on ding and miratrix (2015). *Journal of Causal Inference*, 3(2):253–258, 2015.

[21] Judea Pearl. [bayesian analysis in expert systems]: Comment: Graphical models, causality and intervention. *Statistical Science*, 8(3):266–269, 1993.

[22] Emilija Perković, Johannes Textor, Markus Kalisch, and Marloes H Maathuis. Complete graphical characterization and construction of adjustment sets in markov equivalence classes of ancestral graphs. *Journal of Machine Learning Research*, 18(220):1–62, 2018.

[23] Peter Spirtes and Clark Glymour. An algorithm for fast recovery of sparse causal graphs. *Social science computer review*, 9(1):62–72, 1991.

[24] Marloes H Maathuis, Markus Kalisch, and Peter Bühlmann. Estimating high-dimensional intervention effects from observational data. 2009.

[25] Daniel Malinsky and Peter Spirtes. Estimating bounds on causal effects in high-dimensional and possibly confounded systems. *International Journal of Approximate Reasoning*, 88:371–384, 2017.

[26] Antti Hyttinen, Frederick Eberhardt, and Matti Järvisalo. Do-calculus when the true graph is unknown. In *UAI*, volume 15, pages 395–404, 2015.

[27] Mátyás Schubert, Tom Claassen, and Sara Magliacane. Local causal discovery for statistically efficient causal inference. In Emtiyaz Khan, Yingzhen Li, Arno Solin, and Aaditya Ramdas, editors, *Proceedings of The 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pages 1990–1998. PMLR, 02–05 May 2026.

[28] Doris Entner, Patrik Hoyer, and Peter Spirtes. Data-driven covariate selection for nonparametric estimation of causal effects. In *Artificial intelligence and statistics*, pages 256–264. PMLR, 2013.

[29] Jacqueline Maasch, Weishen Pan, Shantanu Gupta, Volodymyr Kuleshov, Kyra Gan, and Fei Wang. Local discovery by partitioning: Polynomial-time causal discovery around exposure-outcome pairs. In Negar Kiyavash and Joris M. Mooij, editors, *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence*, volume 244 of *Proceedings of Machine Learning Research*, pages 2350–2382. PMLR, 15–19 Jul 2024.

[30] Zeyu Liu, Zheng Li, Feng Xie, Yan Zeng, Hao Zhang, and Kun Zhang. Local covariate selection for average causal effect estimation without pretreatment and causal sufficiency assumptions. In *Forty-third International Conference on Machine Learning*, 2026.

[31] Ioannis Tsamardinos, Constantin F Aliferis, Alexander R Statnikov, and Er Statnikov. Algorithms for large scale markov blanket discovery. In *FLAIRS*, volume 2, pages 376–81, 2003.

[32] Sandeep Yaramakala and Dimitris Margaritis. Speculative markov blanket discovery for optimal feature selection. In *Fifth IEEE International Conference on Data Mining (ICDM'05)*, pages 4–pp. IEEE, 2005.

[33] Constantin F Aliferis, Ioannis Tsamardinos, and Alexander Statnikov. Hiton: a novel markov blanket algorithm for optimal variable selection. In *AMIA annual symposium proceedings*, volume 2003, page 21, 2003.

[34] Jean-Philippe Pellet and André Elisseeff. Using markov blankets for causal structure learning. *Journal of Machine Learning Research*, 9(7), 2008.

[35] Joris M Mooij and Tom Claassen. Constraint-based causal discovery using partial ancestral graphs in the presence of cycles. In *Conference on Uncertainty in Artificial Intelligence*, pages 1159–1168. Pmlr, 2020.

[36] Patrick Forré and Joris M Mooij. Causal calculus in the presence of cycles, latent confounders and selection bias. In *Uncertainty in Artificial Intelligence*, pages 71–80. PMLR, 2020.

[37] Markus Kalisch and Peter Bühlmann. Estimating high-dimensional directed acyclic graphs with the pc-algorithm. *Journal of Machine Learning Research*, 8(22):613–636, 2007.

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

A Markov blanket discovery algorithms

We considered four standard Markov blankets discovery algorithms: IAMB [31], Fast-IAMB [32], HITON-MB [33] and TC [34]. All methods aim to recover the Markov blanket $\mathrm{MB}(X)$ of a target variable X using conditional independence (CI) tests.

The TC algorithm relies on an exhaustive characterization: a variable Y belongs to $\mathrm{MB}(X)$ iff X and Y are dependent conditional on all remaining variables. This characterization is sound and complete under faithfulness, however it requires CI tests with conditioning sets of size $|\mathbf{V}| - 2$, which can be computationally expensive and statistically unreliable under large graphs regimes.

In contrast, the other methods adopt a grow-shrink strategy that incrementally constructs the Markov blanket. The IAMB algorithm alternates between the forward phase, where variables associated with X are added, and a backward phase, where it removes the false positives via CI tests. HITON-MB refines the strategy by enforcing stricter inclusion criteria, performing more tests before including a new candidate. Lastly, Fast-IAMB prioritizes variables with the strongest conditional association in the forward phase and only considers conditional independence tests if there is a significant number of samples. Since we are using Fisher's Z test, we require $N > |\boldsymbol{Z}| + 3$ where $\boldsymbol{Z}$ is the valid adjustment set [37]. We note that for our sample sizes starting at 1000 samples, while we have at most 250 nodes in total, this condition is always satisfied.

Empirical results in Figure 4 show that for small graphs ($|\mathbf{V}| \leq 50$), all algorithms exhibit comparable runtimes, with HITON-MB occasionally being slower due to its conservative inclusion strategy. However, as the number of nodes increases, TC becomes significantly slower than the other methods due to the large conditioning sets involved. Fast-IAMB shows better performance in the larger settings, combining a moderate number of CI tests with consistently small conditioning sets. Based on these observations, we use TC for small graphs, where its exhaustive characterization remains tractable, and Fast-IAMB for larger graphs.

Remark 4. *All conditional independence statements in the algorithms are evaluated using a statistical test at significance level α. The CI test $\mathrm{CI}(X, Y \mid \mathbf{S}; \alpha)$ returns true if X and Y are judged conditionally independent given $\mathbf{S}$, i.e. $X \perp\!\!\!\perp Y \mid \mathbf{S}$.*

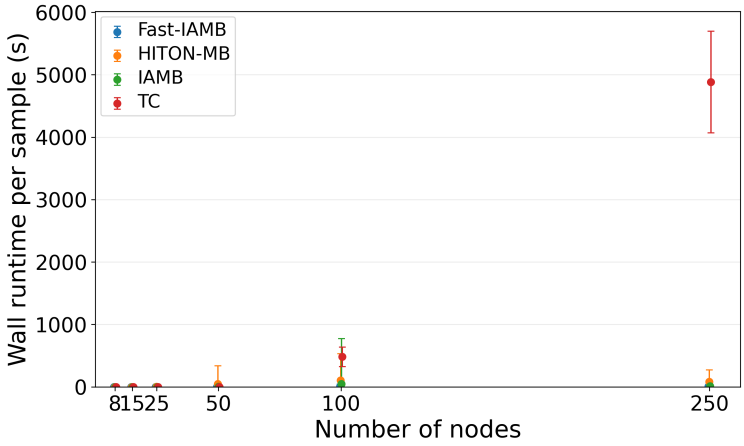
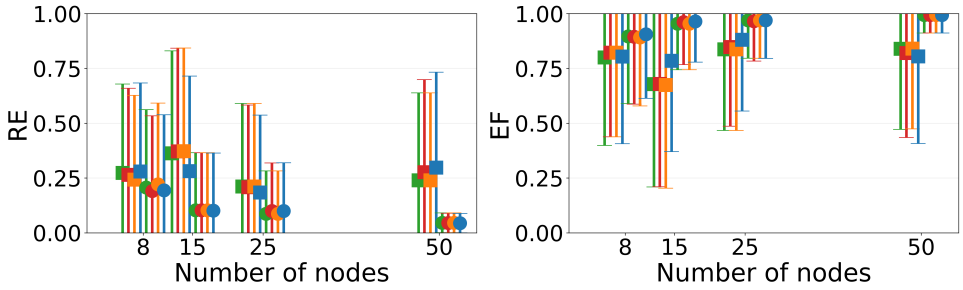


Figure 4: The y-axis shows the wall run time in seconds, while the x-axis is the number of nodes. The figure shows the mean runtime with one standard deviation for Fast-IAMB (blue), HITON-MB (orange), IAMB (green) and TC (red). For the smaller number of nodes, up to 50, the algorithms run in similar times, with HITON-MB sometimes taking the longest. However, as the number of nodes increases, TC starts to become very slow compared to the other algorithms.

Furthermore, Figure 5 shows a comparison of the Markov blanket discovery algorithms on the relative error and the edge fraction accuracy. The experimental setup is the same as for the smaller graphs discussed in Section 7. The figure shows that all Markov blanket discovery algorithms have approximately the same performance per (a)cyclic configuration. Based on this, we assume that all algorithms seem to perform equally well on our relevant metrics, making it possible to pick whichever is preferred based on other considerations.



(a) Relative error for linear models

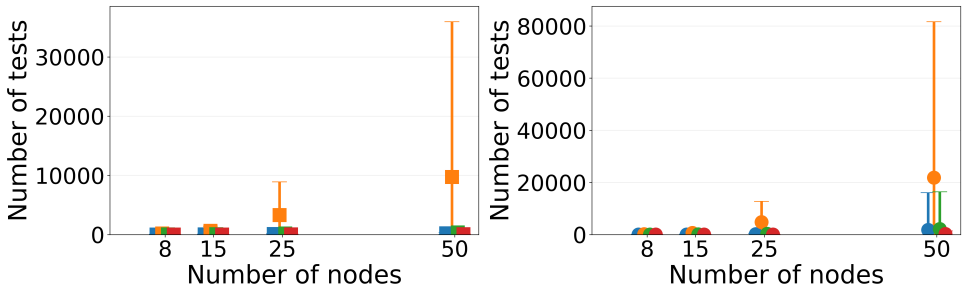
(b) Edge fraction acc. for (non-)linear models

Figure 5: Performance across varying graph sizes of the different Markov blanket discovery algorithms. The x-axis shows the number of nodes, while the y-axis shows the evaluated metric. Results are reported for acyclic (squares) and cyclic (circles) configurations, where the offset from left to right corresponds to Fast-IAMB (green), HITON-MB (red), IAMB (orange) and TC (blue). Markers indicate mean values, and error bars denote one standard deviation. Each subfigure corresponds to a different performance metric.

Finally, Figure 6 shows the number of independence tests across varying graph sizes of the different Markov blanket discovery algorithms for a sample size of $15K$. In both graph structures, HITON-MB (second in the offset) is considerably worse than Fast-IAMB, IAMB and TC. The latter are quite similar. Comparing the cyclic and acyclic cases, the cyclic configuration has a larger number of independence tests and variance compared to the acyclic configuration across the board.

B Local adjustment set search

In this section, we outline the procedure of the local causal adjustment set search. The algorithm LSAS [7] is shown in Algorithm 3 and uses Θ to store the estimated causal effect of X on Y. If $\Theta = \emptyset$, the procedure is unable to identify a valid adjustment set. If $\Theta = 0$, the algorithm concludes that there is no causal effect. Otherwise, Θ contains an estimate of the causal effect.



(a) Acyclic (b) Cyclic

Figure 6: Number of independence tests across varying graph sizes of the different Markov blanket discovery algorithms for a sample size of $15K$. The x-axis shows the number of nodes, while the y-axis shows the number of independence tests. Results are reported for acyclic (squares) and cyclic (circles) configurations, where the offset from left to right corresponds to Fast-IAMB (blue), HITON-MB (orange), IAMB (green) and TC (red). Markers indicate mean values, and error bars denote one standard deviation. Each subfigure corresponds to a different graph structure.

Algorithm 1: Fast-IAMB Markov Blanket Discovery [32]

Input: Data D, variables $\mathbf{V}$, target X, significance level k
Output: $\widehat{MB}(X)$
$\widehat{MB}(X) \leftarrow \emptyset$;
$\mathbf{A} \leftarrow \{Y \in \mathbf{V} \setminus \{X\} : X \not\perp\!\!\!\perp Y \mid \emptyset\}$;
while $\mathbf{A} \neq \emptyset$ **do**
  $\mathrm{MB}_{\text{old}} \leftarrow \widehat{MB}(X)$;
  insufficient_data $\leftarrow$ False;
  Sort $\mathbf{A}$ by decreasing association with X given $\widehat{MB}(X)$;
  foreach $Y \in \mathbf{A}$ **do**
    if *Significant*$\big(X, Y, \widehat{MB}(X), k\big)$ = True **then**
      $\widehat{MB}(X) \leftarrow \widehat{MB}(X) \cup \{Y\}$;
    else
      insufficient_data $\leftarrow$ True;
  foreach $Y \in \widehat{MB}(X)$ **do**
    if $X \perp\!\!\!\perp Y \mid \widehat{MB}(X) \setminus \{Y\}$ **then**
      $\widehat{MB}(X) \leftarrow \widehat{MB}(X) \setminus \{Y\}$;
  $\mathbf{A} \leftarrow \{Y \in \mathbf{V} \setminus (\{X\} \cup \widehat{MB}(X)) : X \not\perp\!\!\!\perp Y \mid \widehat{MB}(X)\}$;
  if *insufficient_data* = True **and** $\widehat{MB}(X) = MB_{old}$ **then**
    break;
return $\widehat{MB}(X)$;

C Proofs

We first introduce maximal ancestral graphs (MAGs), which generalize DAGs to settings with latent variables.

Def. C.1. *A maximal ancestral graph (MAG) $\mathcal{M}_{\mathbf{O}} = (\mathbf{O}, \mathbf{E})$ over observed variables $\mathbf{O}$ is a mixed graph containing directed edges ($\rightarrow$), representing direct causal relations between observed variables, and bidirected edges ($\leftrightarrow$), encoding the presence of latent common causes. A MAG satisfies that for every pair of non-adjacent nodes, there exists a conditioning set that m-separates them.*

Conditional independence in MAGs is characterized through m-separation, which generalizes d-separation in DAGs.

Algorithm 2: TC Markov Blanket Discovery [34]

Input: Data D, variables $\mathbf{V}$, target X, significance level α
Output: $\widehat{MB}(X)$
$\widehat{MB}(X) \leftarrow \emptyset$;
$n \leftarrow |D|$;
$q \leftarrow \lfloor n/10 \rfloor$;
if $q > 0$ **then**
  $\alpha' \leftarrow \alpha/(10q)$;
else
  $\alpha' \leftarrow \alpha$;
foreach $Y \in \mathbf{V} \setminus \{X\}$ **do**
  $\mathbf{S} \leftarrow \mathbf{V} \setminus \{X, Y\}$;
  if $X \not\perp\!\!\!\perp Y \mid \mathbf{S}$ *at significance level* α' **then**
    $\widehat{MB}(X) \leftarrow \widehat{MB}(X) \cup \{Y\}$;
return $\widehat{MB}(X)$;

Algorithm 3: Local Search Adjustment Sets (LSAS) [7]

Input: Observational data D, treatment X, outcome Y
Output: Estimated causal effect Θ
$\widehat{MB}(X), \widehat{MB}(Y) \leftarrow$ Markov Blanket Discovery(X, Y, D);
$\Theta \leftarrow \emptyset$;
foreach $W \in \widehat{MB}(X) \setminus \{Y\}$ **do**
  foreach $\mathbf{Z} \subseteq \widehat{MB}(Y) \setminus \{X\}$ **do**
    if (W, Z) *satisfies R1 6.1* **then**
      Estimate $\theta = \mathbb{E}[Y \mid X, \mathbf{Z}]$;
      $\Theta \leftarrow \theta$;
    if $(W, \mathbf{Z})$ *satisfies R2 6.2* **then**
      return $\Theta \leftarrow 0$;
return Θ;

Lemma C.1 (Invariance of Markov blanket between DAGs and MAGs)**.** *Let G be a DAG possibly containing latent variable, and let $\mathcal{M}_{\mathbf{O}}$ be the MAG over $\mathbf{O}$ obtained by latent projection of G. Therefore:*

$$\mathrm{MB}_m^{\mathcal{M}_{\mathbf{O}}}(Y) = \mathrm{MB}_d^G(Y)$$

Proof. By [38], the observed conditional independence assertions induced by G^{acy} is exactly the m-separation model of $\mathcal{M}_{\mathbf{O}}$. By the latent projection [38], for all disjoint $A, B, S \subseteq \mathbf{O}$,

$$A \perp\!\!\!\perp_d B \mid \mathbf{S} \text{ in } G^{\mathrm{acy}} \quad \Longleftrightarrow \quad A \perp\!\!\!\perp_m B \mid \mathbf{S} \text{ in } \mathcal{M}_{\mathbf{O}}.$$

Therefore, a set $\mathbf{S} \subseteq \mathbf{O} \setminus \{X\}$ separates X from all observed variables outside $\mathbf{S} \cup \{X\}$ in G^{acy} if and only if it does so in $\mathcal{M}_{\mathbf{O}}$. Since the collection of separating sets is identical in both graphs, their minimal sets coincide. □

C.1 Lemma 4.1

Proof. The independence models induced by σ-separation in G and d-separation in G^{acy} coincide, i.e., $\mathrm{IM}_\sigma(G) = \mathrm{IM}_d(G^{\mathrm{acy}})$ [35]. The result follows immediately from the definition of these models.

□

C.2 Lemma 4.2

Proof. Let $\mathbf{S} := \mathrm{MB}_{\sigma}^{G}(Y)$. By definition, $Y \perp\!\!\!\perp_{\sigma} \mathbf{O} \setminus (\mathbf{S} \cup \{Y\}) \mid \mathbf{S}$ in G. By Lemma 4.1, this is equivalent to $Y \perp\!\!\!\perp_{d} \mathbf{O} \setminus (\mathbf{S} \cup \{Y\}) \mid \mathbf{S}$ in G^{acy}, so $\mathbf{S}$ is a separating set in G^{acy}, implying $\mathbf{S} \supseteq MB_{d}^{G^{\mathrm{acy}}}(Y)$ by minimality. The reverse inclusion follows symmetrically, yielding equality. □

C.3 Lemma 4.3

Proof. Since $\mathrm{De}(Y) = \emptyset$, node Y has no outgoing edges in G. Therefore, no directed cycles can involve Y. Hence $\mathrm{SCC}(Y) = \{Y\}$ and no acyclification step introduces edges out of Y. This implies $\mathrm{Ch}(Y) = \emptyset$ in any acyclification.

Now consider X. The assumption $\mathrm{De}(X) \subseteq \{Y\}$ implies that X cannot lie on a directed cycle (as $\mathrm{Ch}(Y) = \emptyset$). Hence, $\mathrm{SCC}_{G}(X) = \{X\}$. It follows that no acyclification step modifies edges incident to X inside a SCC. Therefore, the only possible child of X in $\hat{G}^{\mathrm{acy}}$ is Y, and we have $\mathrm{Ch}_{\hat{G}^{\mathrm{acy}}}(X) \subseteq \{Y\}$. Finally, consider the edge $X \to Y$. Since both x and y are singleton SSCs, the σ-acyclification rule for edges between distinct components preserves the edge $X \to Y$ if and only if it is present in G. □

C.4 Lemma 5.1

Proof. ("if" part) Assume that $\mathbf{Z}$ satisfies Definition 5.2. Hence: (i) $\mathbf{Z} \cap \mathrm{De}(X) = \emptyset$, and (ii) $\mathbf{Z}$ σ-blocks every backdoor path from X to Y.

We first show condition (1) from Lemma 5.1: since the intervention node I_X has no incoming edges and a single outgoing edge $I_X \to X$, every path from I_X to an observed variable must first pass through X, hence, any σ-open path from I_X to a node $Z_i \in \mathbf{Z}$ would imply a σ-open path from X to Z_i, which in particular requires that $\mathbf{Z}$ lies in the descendant region of X. This contradicts (i). Therefore, $\mathbf{Z} \perp\!\!\!\perp_{\sigma} I_X$.

Condition (2) from Lemma 5.1: Any path from I_X to Y must begin with $I_X \to X$. After reaching X, either the path leaves X via an outgoing edge, in which case it is blocked by conditioning on X (Note $\mathrm{SCC}(X) = \{X\}$, since for X the only possible descendant Y has no descendants), or it reaches X via an incoming edge when viewed as a path from X to Y, corresponding to a backdoor path. All such backdoor paths are σ-blocked by $\mathbf{Z}$ by (ii). Hence, all paths from I_X to Y are σ-blocked given X,$\mathbf{Z}$, and thus $Y \perp\!\!\!\perp_{\sigma} I_X \mid X, \mathbf{Z}$.

("only if" part) Assume that $\mathbf{Z} \perp\!\!\!\perp_{\sigma} I_X$ and $Y \perp\!\!\!\perp_{\sigma} I_X \mid X, \mathbf{Z}$. We first show that $\mathbf{Z}$ contains no descendants of X. Suppose, for contradiction, that there exists $Z_i \in \mathbf{Z} \cap \mathrm{De}(X)$. Since I_X has directed edges into X, and Z_i is a descendant of X, there exists a directed path from I_X to Z_i. Such a directed path is σ-open without conditioning, contradicting $\mathbf{Z} \perp\!\!\!\perp_{\sigma} I_X$. Hence $\mathbf{Z} \cap \mathrm{De}(X) = \emptyset$.

It remains to show that $\mathbf{Z}$ σ-blocks every backdoor path from X to Y. Suppose that it is not the case. Then there exists a backdoor path from X to Y that is σ-open given $\mathbf{Z}$. Prepending the edge $I_X \to X$ to this path gives a path from I_X to Y. Since the path enters X through a backdoor edge, conditioning on X does not block this path; rather, the path remains σ-open given $X, \mathbf{Z}$. Therefore, $Y \not\perp\!\!\!\perp_{\sigma} I_X \mid X, \mathbf{Z}$, contradicting the second assumed condition. Thus $\mathbf{Z}$ contains no descendants of X and blocks every backdoor path from X to Y. Therefore, $\mathbf{Z}$ satisfies Definition 5.2. This proves the equivalence. □

C.5 Lemma 5.2

Proof. Let G^{acy} be a σ-acyclification of G. By Lemma 5.1 and the equivalence of σ-separation and d-separation under σ-acyclification, a set $\mathbf{Z} \subseteq \mathbf{O}$ is a backdoor adjustment set for (X, Y) in G if and only if it is a backdoor adjustment set for (X, Y) in G^{acy}. Hence

$$\exists\, \mathbf{Z} \subseteq \mathbf{O} \text{ backdoor adjustment in } G \iff \exists\, \mathbf{Z} \subseteq \mathbf{O} \text{ backdoor adjustment in } G^{\mathrm{acy}}.$$

Also, using the fact that X is a singleton and G^{acy} is a DAG, we have by [22] that

$$\mathbf{Z} \text{ backdoor adjustment in } G^{\mathrm{acy}} \iff \mathbf{Z} \text{ adjustment in } G^{\mathrm{acy}}.$$

In addition, since G^{acy} is a DAG, let $\mathcal{M}_{\mathbf{O}}$ be the MAG induced by G^{acy} over the observed variables $\mathbf{O}$. By [38], $\mathcal{M}_{\mathbf{O}}$ preserves all conditional independence relations among observed variables from G^{acy}: for any $\mathbf{Z} \subseteq \mathbf{O}$, X, Y are m-separated given $\mathbf{Z}$ in $\mathcal{M}_{\mathbf{O}}$ if and only if they are d-separated given $\mathbf{Z}$ in G^{acy}. Therefore, for observed adjustment sets, validity in the MAG is equivalent to validity in the DAG.

$$\exists\, \mathbf{Z} \subseteq \mathbf{O} \text{ adjustment in } G^{\mathrm{acy}} \iff \exists\, \mathbf{Z} \subseteq \mathbf{O} \text{ adjustment in } \mathcal{M}_{\mathbf{O}}.$$

Previous results for MAG $\mathcal{M}_{\mathbf{O}}$ [7] state that a valid set exists if and only if there exists one contained in the Markov blanket of Y (excluding X) in $\mathcal{M}_{\mathbf{O}}$

$$\exists\, \mathbf{Z} \subseteq \mathbf{O} \text{ adjustment in } \mathcal{M}_{\mathbf{O}} \iff \exists\, \mathbf{Z} \subseteq \mathrm{MB}_{m}^{\mathcal{M}_{\mathbf{O}}}(Y) \setminus \{X\}.$$

By the Markov blanket equivalence between DAGs and MAGs from Lemma C.1:

$$\exists\, \mathbf{Z} \subseteq \mathrm{MB}_m^{\mathcal{M}_\mathbf{O}}(Y) \setminus \{X\} \iff \exists\, \mathbf{Z} \subseteq \mathrm{MB}_d^{G^{\mathrm{acy}}}(Y) \setminus \{X\}.$$

By the equivalence of Markov blankets from Lemma 4.2, we have $\mathrm{MB}_\sigma^G(Y) = \mathrm{MB}_d^{G^{\mathrm{acy}}}(Y)$. Substituting yields

$$\exists\, \mathbf{Z} \subseteq \mathrm{MB}_d^{G^{\mathrm{acy}}}(Y) \setminus \{X\} \iff \exists\, \mathbf{Z} \subseteq \mathrm{MB}_\sigma^G(Y) \setminus \{X\}.$$

Combining the above equivalences completes the proof. □

C.6 Lemma 6.1

Proof. We show that the following two statements hold:

1. $W \in \mathrm{MB}_\sigma^G(X)$ in G if and only if $W \in MB_d^{G^{\mathrm{acy}}}(X)$ in G^{acy}.
2. For every $\mathbf{Z} \subseteq \mathbf{O} \setminus \{X, Y\}$,
$$W \not\perp\!\!\!\perp_\sigma Y \mid \mathbf{Z} \iff W \not\perp\!\!\!\perp_d Y \mid \mathbf{Z},$$
$$W \perp\!\!\!\perp_\sigma Y \mid \mathbf{Z} \cup \{X\} \iff W \perp\!\!\!\perp_d Y \mid \mathbf{Z} \cup \{X\},$$
where the right-hand side is evaluated in G^{acy}.

Statement (1) is exactly the Markov blanket equivalence under σ-acyclification (Lemma 4.2).

For (2), let $W \in \mathbf{O} \setminus \{X, Y\}$ and $\mathbf{Z} \subseteq \mathbf{O} \setminus \{X, Y\}$. By the equivalence of σ-separation in G and d-separation in G^{acy}, we have

$$W \perp\!\!\!\perp_\sigma Y \mid \mathbf{Z} \iff W \perp\!\!\!\perp_d Y \mid \mathbf{Z}, \qquad W \perp\!\!\!\perp_\sigma Y \mid \mathbf{Z} \cup \{X\} \iff W \perp\!\!\!\perp_d Y \mid \mathbf{Z} \cup \{X\}.$$

Taking negations yields the corresponding dependence statements, proving (2).

Then the claim follows immediately: by (1), the candidate witnesses W lie in the same Markov blanket in both graphs, and by (2), the R1/R2 conditions are preserved. Hence existence is invariant under σ-acyclification. □

C.7 Lemma 6.2

Proof. Soundness: Assume that the conditions of R1 or R2 hold in G. By Lemma 6.1, the same conditions hold in any G^{acy}. Since G^{acy} is acyclic, and a candidate adjustment set $\mathbf{Z}$ is contained in the observed $\mathrm{MB}_d^{G^{\mathrm{acy}}}(Y) \setminus \{X\}$, Theorem 2 in [28] implies that R1 and R2 are sound in G^{acy}.

If R1 applies in G^{acy}, it concludes that: (i) there is a direct edge $X \to Y$ in G^{acy}; (ii) $\mathbf{Z}$ is a valid adjustment set for estimating the effect of X on Y in G^{acy}. By Lemma 4.3, the edge $X \to Y$ in G^{acy} implies the presence in G. By Lemma 5.1 and 4.1, the set Z is also a valid adjustment set in G. Therefore, both conclusions of R1 are sound in the original cyclic graph G.

If R2 applies in G^{acy}, it concludes that there is no direct edge in G^{acy}. By Lemma 4.3, this absence is preserved in G. Hence, R2 is also sound in G.

Therefore, R1 and R2 are sound in simple cyclic SCMs.

Completeness: Suppose that neither R1 nor R2 applies in G. By Lemma 6.1, the same conditions over $\mathbf{O}$ fail in G^{acy}. Let $\mathcal{M}_\mathbf{O}$ be the MAG over $\mathbf{O}$ obtained by latent projection of G^{acy}. By Lemma C.1, these observed conditional independence relations are exactly the m-separation relations in $\mathcal{M}_\mathbf{O}$.

Therefore, neither R1 nor R2 applies in $\mathcal{M}_\mathbf{O}$. By the completeness result from [7], no criterion based only on the observed conditional independence relations can determine whether the effect of X on Y is identifiable. Since G^{acy} and $\mathcal{M}_\mathbf{O}$ induce the same observed conditional independence assertions by [38], the same holds for G^{acy} and G by Lemma 4.1. Therefore, the same non-identifiabilityfrom conditional independence information holds for the original possibly cyclic SCM G. In conclusion, any ambiguity in determining the presence or absence of a direct causal effect from X to Y in G^{acy} based solely on conditional independencies also holds in G. □

C.8 Lemma 6.3

Proof. We show $R1$ is no longer sound by providing an example. Consider the simple cyclic SCM G in Figure 7 such that $Y \notin \mathrm{SCC}(X)$ and $|\mathrm{SCC(X)}| = 2$. Then since $Z \in \mathrm{De}(X)$, we have that $Z \not\perp\!\!\!\perp_\sigma I_X$ in G^{I_X} and thus that Z cannot be in a valid backdoor adjustment set by Corollary 8.4 from [36]. However, for G, we have that $W \in \mathrm{MB}_\sigma^G(X) \setminus \{Y\}$ and $\{Z\} \subseteq \mathrm{MB}_\sigma^G(Y) \setminus \{X, W\}$, while

$$W \not\perp\!\!\!\perp_\sigma Y \mid Z \quad \text{and} \quad W \perp\!\!\!\perp_\sigma Y \mid Z \cup \{X\}.$$

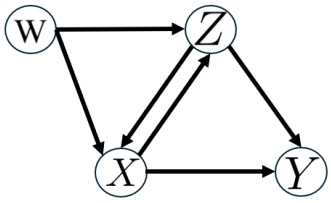


Figure 7: Example graph G used in the proof of Lemma 6.3.

and thus by σ-faithfulness (Assumption (ii))

$$W \not\perp\!\!\!\perp Y \mid Z \quad \text{and} \quad W \perp\!\!\!\perp Y \mid Z \cup \{X\}.$$

Hence, $R1$ will return $\{Z\}$ as a valid backdoor adjustment set. Therefore, $R1$ is not sound.

Finally, we note that adding any node to the $\mathrm{SCC}(X)$ that is not connected to either W or Y will not change these conditional independence relations. Hence, we can always construct a simple SCM G' such that $Y \notin \mathrm{SCC}(X)$ and $|\mathrm{SCC(X)}| \geq 2$ for which $R1$ is not sound. □

C.9 Lemma 6.4

Proof. Let $\mathcal{A}$ be any sound Markov blanket algorithm. Since $\mathcal{A}$ only applies conditional tests under observed variables, the behavior is determined entirely by the observed independence assertions over $\mathbf{O}$. By Lemma 4.1, every conditional independence query has the same truth value under the observational distribution generated by G as implied by the d-separation in G^{acy}. Since G^{acy} may contain latent variables, let $\mathcal{M}_{\mathbf{O}}$ be its latent projection MAG over $\mathbf{O}$. Thus by soundness and completeness of $\mathcal{A}$ for the observed CI relations, $\mathcal{A}$ returns $\mathrm{MB}_m^{\mathcal{M}_{\mathbf{O}}}(X)$. By Lemma C.1, the Markov blanket in G^{acy}, i.e., $\mathrm{MB}_d^{G^{\mathrm{acy}}}(X)$ coincides with the m-Markov blanket in $\mathcal{M}_{\mathbf{O}}$, i.e., $\mathrm{MB}_m^{\mathcal{M}_{\mathbf{O}}}(X)$. By Lemma 4.2, $\mathrm{MB}_d^{G^{\mathrm{acy}}}(X) = \mathrm{MB}_\sigma^{G}(X)$, which completes the proof. □

D Data generation

Let's consider simple SCMs where $\mathbf{V}$ are random variables, W is the weighted adjacency matrix consistent with the underlying graph structure and $\mathbf{U}$ is the vector of independent noise variables. Interventions are modeled using the do-operator.

D.1 Linear SCMs

We define a linear SCM:

$$\mathbf{V} := W\mathbf{V} + \mathbf{U}.$$

In acyclic graphs, W is strictly triangular, and the system admits a unique solution. In the presence of cycles, we need to enforce that $(I - W)$ is invertible during data generation, which guarantees a unique solution (simple SCMs), s.t. $\mathbf{V} = (I - W)^{-1}\mathbf{U}$. In the code, we verify that $I - W$ is invertible to ensure the existence of the matrix $(I - W)^{-1}$.

D.1.1 Interventions

Let X be the treatment variable and Y the outcome variable. Under an intervention $do(X = x)$, the structural equation for X is replaced by a constant:

$$\begin{aligned} \mathbf{V}_K &= W_{KK}\mathbf{V}_K + xW_{KX} + \mathbf{U}_K \\ (I - W_{KK})\mathbf{V}_K &= xW_{KX} + \mathbf{U}_K \\ \mathbf{V}_K &= (I - W_{KK})^{-1}(xW_{KX} + \mathbf{U}_K) \end{aligned}$$

where $\mathbf{K} = \mathbf{V} \setminus \{X\}$.

D.1.2 Causal effect

The total causal effect (CE) from X on Y can be defined in both a binary and a continuous setting. In the setting where the treatment is continuous:

$$CE = \frac{\partial}{\partial x} E^{\mathrm{do}(X:=x)}[Y],$$

$E^{\mathrm{do}(\mathbf{X}:=x)}$ indicates the distribution where the do action assignment has been performed.

Acyclic case In the acyclicity setting, when considering the pre-treatment assumption, the causal effect of X on Y is:

$$\begin{aligned}
CE &:= \frac{\partial}{\partial x} E^{\mathrm{do}(X:=x)}[Y] \\
&= \frac{\partial}{\partial x} E[(W_{KK}\mathbf{V}_K + xW_{KX} + \mathbf{U}_K)e_Y] \\
&= \frac{\partial}{\partial x} E[xW_{KX}e_Y] + \frac{\partial}{\partial x} E[(W_{KK}\mathbf{V}_K + \mathbf{U}_K)e_Y] \\
&= \frac{\partial}{\partial x} xW_{KX}e_Y \\
&= W_{KX}e_Y = W_{YX},
\end{aligned}$$

Here we use linearity of expectation and the fact that, in DAGs, $\mathbf{V}_K$ does not depend on x through feedback and in the pre-treatment setting

Cyclic case In the cyclic setting, feedback induces dependence on x through $\mathbf{V}_k$. Using the same form above, we obtain:

$$\begin{aligned}
CE &= \frac{\partial}{\partial x} E^{\mathrm{do}(X:=x)}[Y] \\
&= \frac{\partial}{\partial x} E[(I - W_{KK})^{-1}(xW_{KX} + \mathbf{U})e_Y] \\
&= \frac{\partial}{\partial x} E[x(I - W_{KK})^{-1}W_{KX}e_y] + \frac{\partial}{\partial x} E[(I - W_{KK})^{-1}\mathbf{U}_Y] \\
&= \frac{\partial}{\partial x} x(I - W_{KK})^{-1}W_{KX}e_y \\
&= (I - W_{KK})^{-1}W_{KX}e_y,
\end{aligned}$$

D.2 Nonlinear SCMs

We define a nonlinear SCM as:

$$\mathbf{V} := tanh(W\mathbf{V}) + \mathbf{U}$$

where $tanh$ is applied componentwise. In acyclic graphs, the system can be solved by forward substitution. In cyclic graphs, we enforce a condition on the operator norm of W.

Lemma D.1. *If $\|W\| < 1$, where $\|\cdot\|$ is an induced operator norm, then the equation*

$$\mathbf{V} = \tanh(W\mathbf{V}) + \mathbf{U} \tag{D.2.1}$$

admits a unique solution.

Proof. Define $F : \mathbb{R}^d \to \mathbb{R}^d$ as

$$F(\mathbf{V}) := \tanh(W\mathbf{V}) + \mathbf{U}.$$

For any $\mathbf{V}, Y \in \mathbb{R}^d$,

$$\|F(\mathbf{V}) - F(Y)\| = \|\tanh(W\mathbf{V}) - \tanh(WY)\|.$$

Since $\tanh$ is componentwise 1-Lipschitz,

$$\|\tanh(W\mathbf{V}) - \tanh(WY)\| \le \|W\mathbf{V} - WY\|.$$

Using the definition of the induced operator norm,

$$\|W\mathbf{V} - WY\| \le \|W\|\,\|\mathbf{V} - Y\|.$$

Therefore,

$$\|F(\mathbf{V}) - F(Y)\| \leq \|W\| \, \|\mathbf{V} - Y\|.$$

If $\|W\| < 1$, then F is a contraction. Since $\mathbb{R}^d$ is complete, the Banach fixed-point theorem guarantees the existence and uniqueness of a fixed point $\mathbf{V}^*$ such that

$$\mathbf{V}^* = \tanh(W\mathbf{V}^*) + \mathbf{U}.$$

□

Implementation details. In practice, we enforce the condition $\|W\| < 1$ using the Euclidean operator norm, i.e., the largest singular value of W.

D.2.1 Causal Effect

In contrast to the linear case, the causal effect in non-linear SCMs does not admit a simple closed-form expression. In particular, the quantity $\frac{\partial}{\partial x}\mathbb{E}^{\mathrm{do}(X:=x)}[Y]$ generally depends on the intervention value x, and therefore does not reduce to a constant effect.

D.3 Generation and assumptions translations

Latent and observed variables To simulate partial observability, we randomly select a subset of nodes as observed variables. The final dataset contains only samples corresponding to these observed nodes, while always including the designated treatment and outcome variables. This ensures that causal queries remain well-defined.

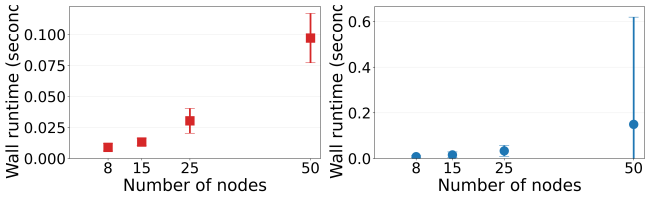
Treatment and outcome For each generated SCM, we fix a treatment variable X and an outcome variable Y. We enforce the presence (or absence) of a direct edge from X to Y.

E Experimental setup

Conditional independence tests In all experiments, conditional independence tests were performed using Fisher's Z-test with significance level $\alpha = 0.01$.

Hardware We ran all experiments on a machine with 16 Intel Xeon E5-2630 v3 cores at 2.40 GHz, supporting 32 hardware threads, and 512 GB of RAM available. No GPU was used.

E.1 Compute time

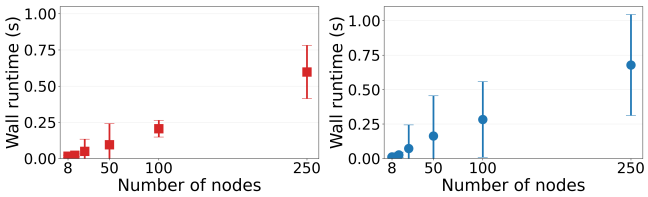


(a) Acyclic graphs (b) Cyclic graphs

Figure 8: Wall runtime in seconds (s) of the adjustment set discovery method under the pre-treatment assumption for increasing graph sizes when computing the precision and using TC for the Markov blanket discovery. The x-axis represents the number of nodes in the graph, while the y-axis shows the measured wall runtime. Subfigures 8a and 8b are for acyclic and cyclic settings, respectively.

Figure 8 shows that the wall runtime of the adjustment set discovery method under the pre-treatment assumption for graphs of varying sizes, considering both acyclic and cyclic settings when computing the precision and using TC for the Markov blanket discovery. As expected, the runtime generally increases as the graph size grows. Larger graphs exhibit higher runtimes due to the increased computational complexity. Comparing the two configurations, cyclic graphs consistently show slightly higher runtimes than acyclic graphs.

Figure 9 shows the same setup, but without computing the precision and using Fast-IAMB instead of TC for the Markov blanket discovery. As expected, the runtime generally increases as the graph size grows. Larger graphs exhibit higher runtimes due to the increased computational complexity. Comparing the two settings, cyclic graphs consistently show slightly higher runtimes than acyclic graphs.



(a) Acyclic graphs (b) Cyclic graphs

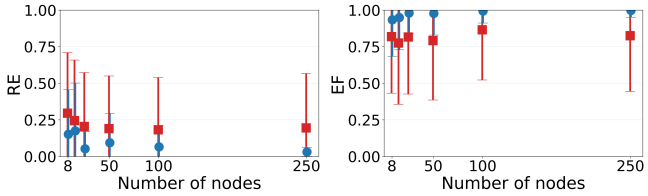
Figure 9: Wall runtime in seconds (s) of the adjustment set discovery method under the pre-treatment assumption for increasing graph sizes without computing the precision and using Fast-IAMB for the Markov blanket discovery. The x-axis represents the number of nodes in the graph, while the y-axis shows the measured wall runtime. Subfigures 9a and 9b are for acyclic and cyclic settings, respectively.

Comparing the smaller graph sizes in Figures 9 and 8 shows that omitting the precision computation yields a substantial speedup, especially for the cyclic graph with 50 nodes. This is expected, since computing precision requires checking whether a set is a valid adjustment set, which entails verifying the σ-backdoor criterion over a combinatorially growing number of paths.

F Extension of the results

F.1 Large graphs

Figure 10 shows the behavior of the evaluated metrics as a function of the number of nodes. The graphs were created under the same experimental setups as described in Section 7 with the important change that the possible size of the set $|\mathbf{Z}| \leq 10$ in checking rule $R1$ and $R2$ (Definitions 6.1, 6.2). Overall, there is a clear trend of improving performance with increasing graph size. Consistent with the results in Section 7, cyclic models show better performance than acyclic ones across both metrics and graph sizes. This seems to suggest that even when enforcing a maximum size on the set $\mathbf{Z}$, the algorithm works well. This could be due to the algorithm correctly rejecting scenarios where it cannot decide based on the smaller set, and/or the valid adjustment sets being small in general in our graphs.

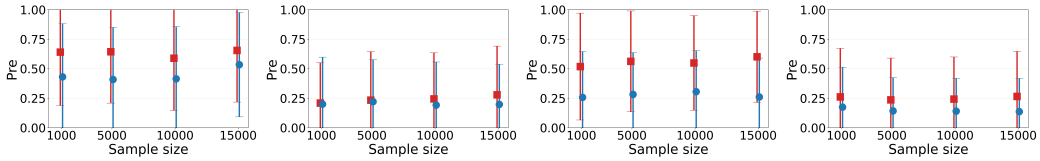


(a) Relative error for linear models (b) Edge fraction acc. for (non-)linear models

Figure 10: Performance across varying graph sizes. The x-axis represents the number of nodes, while the y-axis shows the evaluated metric. Results are reported for acyclic (red squares) and cyclic (blue circles) configurations. Markers indicate mean values, and error bars denote one standard deviation. Each subfigure corresponds to a different performance metric.

F.2 Violation of the pre-treatment assumption

Figure 11 shows the precision of linear models in acyclic (red square) and cyclic (blue circle) settings across sample sizes, comparing scenarios that violate the pre-treatment assumption (non-pre.) with those that satisfy it (pre.). Markers show the mean, with error bars indicating one standard deviation. When looking at the effect of the violation of the pre-treatment assumption for 8 nodes by comparing Figures 11a and 11b, we see that the acyclic SCM reduces in precision by the violation more than the cyclic configuration. Moreover, in the pre-treatment violation, both configuration types are performing equally. This is in contrast with the better performance of the acyclic configuration when looking at Figure 3. The same pattern repeats itself for the other graphs, see for example, the 25 nodes in Figures 11c and 11d.

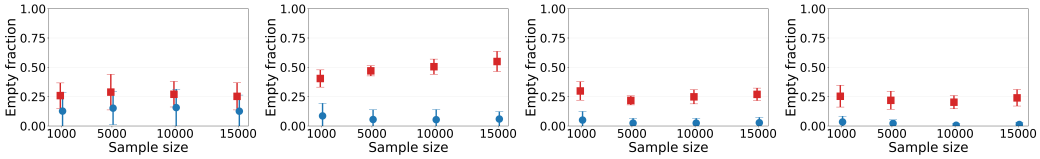


(a) 8 nodes pre. (b) 8 nodes non-pre. (c) 25 nodes pre. (d) 25 nodes non-pre.

Figure 11: Precision of linear models in acyclic (red square) and cyclic (blue circle) settings across sample sizes, comparing scenarios that violate the pre-treatment assumption (non-pre.) with those that satisfy it (pre.). Markers show the mean, with error bars indicating one standard deviation. Subfigures correspond to graph sets with different numbers of nodes and pre-treatment assumptions.

F.3 Empty fraction

Figure 12 shows the empty fraction, i.e. the percentage where the algorithm returns an empty set, for linear and non-linear models together across acyclic (red square) and cyclic (blue circle) settings across sample sizes. Markers show the mean, with error bars indicating one standard deviation. Subfigures correspond to graph sets with different numbers of nodes. Overall, cyclic settings consistently exhibit lower empty fraction, remaing bellow 20%. Acyclic cases show slightly higher empty fraction, typically around 25%, with the exception in the 15 nodes settings, where it increases substantially, being around 50%.



(a) 8 nodes (b) 15 nodes (c) 25 nodes (d) 50 nodes

Figure 12: Empty fraction of linear and non-linear models in acyclic (red square) and cyclic (blue circle) settings across sample sizes. Markers show the mean, with error bars indicating one standard deviation. Subfigures correspond to graph sets with different numbers of nodes.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes], [No], or [N/A].
- [N/A] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for [N/A]).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will also be asked to include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While [Yes] is generally preferable to [No], it is perfectly acceptable to answer [No] provided a proper justification is given (e.g., error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering [No] or [N/A] is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS Paper Checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction state the main theoretical and empirical contributions: extending local covariate selection to cyclic SCMs, proving soundness/completeness, and validating on synthetic cyclic and acyclic data.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper discusses limitations throughout Sections 7–9: the pre-treatment assumption, the relatively high variance in some experimental regimes, and the different behavior between cyclic and acyclic setting under the precision metric.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.

- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The paper states the main assumptions in Section 2 and provides formal lemmas/theorems with proofs in Appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper describes the synthetic data generation process, as well the algorithms in Appendix. The code containing the data generation and the algorithms in the paper discussed is also available.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to

the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.

- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The paper provides the source code.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Section 7 and Appendix specify the synthetic SCM setup, graph sizes, number of samples, latent-variable generation, functional forms, noise distributions, and evaluation metrics. The paper also describes which Markov blanket algorithms are used for different graph sizes.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: The experimental figures report means with error bars indicating standard deviation.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The paper specifies the compute resources used, including the use of CPU, the available memory, as well the wall runtime in Appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The work uses synthetic data and focuses on theoretical and methodological causal inference. The research conforms to the Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

Justification: This paper is a theoretical work, and we do not currently identify direct societal impacts.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper does not release high-risk models, scraped datasets, pre-trained generative models, or assets with obvious misuse risk. Experiments are based on synthetic data.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Proper citations have been provided.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The code is available and well documented.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing or research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing or human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLM was used only for writing, editing, or formatting purpose.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.