Causal Additive Models with Unobserved Causal Paths and Backdoor Paths

Thong Pham
Shiga University
RIKEN AIP

Takashi Nicholas Maeda
Gakushuin University
RIKEN AIP

Shohei Shimizu
The University of Osaka
Shiga University
RIKEN AIP

Abstract

Causal additive models provide a tractable yet expressive framework for causal discovery in the presence of hidden variables. When unobserved backdoor or causal paths exist between two variables, their causal relationship is often unidentifiable under existing theories. We establish sufficient conditions under which causal directions can be identified in many such cases. These conditions rely on new characterizations of regression sets to determine independence among regression residuals and conditional independencies among observed variables. Building on these results, we introduce a search algorithm that incorporates these innovations and prove its soundness and completeness. Empirical evaluations demonstrate its competitive performance against state-of-the-art methods.

1 INTRODUCTION

Discovering causal structure from observational data is a fundamental challenge across the sciences, with implications for reliable prediction (Schölkopf, 2022), decision-making (Ge et al., 2025), and scientific understanding (Glymour et al., 2019). A central approach in this area is to model functional relationships between variables together with stochastic noise, which allows for distinguishing causal directions under suitable assumptions (Peters et al., 2011). Causal Additive Models (CAMs) (Bühlmann et al., 2014) form a particularly influential family of such models, as they capture nonlinear dependencies while remaining amenable to analysis. Due to their tractability, they have been studied considerably and have many practical applications in machine learning (Budhathoki et al., 2022; Yokoyama et al., 2025).

When there are hidden variables, the problem of identifying the causal graph in CAMs remains relatively underexplored, limiting its practical adoption in settings where hidden variables are almost always present. Although causal discovery with hidden variables can be handled in full generality using the Fast Causal Inference (FCI) framework (Spirtes et al., 2001), it is natural to expect that we can do better than FCI in certain aspects by exploiting specific properties of CAMs.

Maeda and Shimizu (2021) identify cases in which parent–child relationships in CAMs can be determined in the presence of unobserved variables by analyzing independencies and dependencies among certain regression residuals. This result is significant and relies on specific properties of CAMs, since parent–child relationships are unidentifiable in the FCI framework. Although their approach accommodates hidden variables, it requires the absence of unobserved backdoor or causal paths. When such paths exist, the parent–child relationship is deemed unidentifiable.

Given a target variable x_j, Schultheiss and Bühlmann (2024) provide sufficient conditions for identifying a causally well-specified set of observed variables, i.e., a set for which the causal effect of intervening on x_j is identifiable in CAMs with unobserved variables. Although they do not discuss causal search as an application, their framework can be used to identify causal directions in CAMs.

Focusing on CAMs with unobserved variables, we show that a) the parent–child relationship is identifiable in certain cases even in the presence of an unobserved backdoor or causal path, thereby extending Maeda and Shimizu (2021), and b) some causal directions beyond those identified by Schultheiss and Bühlmann (2024) are identifiable.

Using our theory, we show that in some cases, the parent–child relationship between x_i and x_j is identifiable even when they share a common hidden parent U. This configuration, known as a *bow*, is notoriously difficult for causal discovery because the hidden confounding introduced by U obscures the parent–child relationship between x_i and x_j (Wang and Drton, 2023; Ashman et al., 2023). To our knowledge, no prior work has established this kind of

---

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300.

identifiability without additional assumptions on the graphical structure.

A high-level summary of our main contributions is as follows.

- By characterizing the regression sets used in determining independence, we show that the causal direction or parent–child relationship between a pair of variables can sometimes be identified using independence between the residuals, even when there are unobserved backdoor paths. We provide examples where the parent–child relationship in a bow is fully identified using the regression approach alone.
- We introduce sufficient conditions, which combine conditional independence between the original variables with independence between regression residuals, to identify causal directions or parent–child relationships in pairs with unobserved backdoor or causal paths. We present examples showing that this hybrid approach can identify parent–child relationships in bow patterns and causal directions in pairs with unobserved causal paths, cases where neither regression alone nor conditional independence methods, such as the FCI framework, can.
- We introduce the CAM-UV-X algorithm that incorporates the above innovations and also addresses a previously overlooked limitation of the CAM-UV algorithm in identifying causal relationships when all backdoor and causal paths are observable. We prove the soundness and completeness of CAM-UV-X in identifying certain causal patterns. We have released the code for the algorithm[1].

The paper is organized as follows. We provide the background in Section 2. New identifiability results are presented in Section 3. Our proposed search method CAM-UV-X is described in Section 4. Numerical experiments are provided in Section 5. Conclusions are given in Section 6.

2 PRELIMINARIES

2.1 The causal model

We assume the following causal additive model with unobserved variables as in Maeda and Shimizu (2021). Let $X = \{x_i\}$ and $U = \{u_i\}$ be the sets of observable and unobservable variables, respectively. $G = (V, E)$ is the DAG with the vertex set $V = \{v_i\} = X \cup U$ and the edge set $E \subseteq V \times V$. The data generation model is

$$v_i = \sum_{j \in P_i} f_j^{(i)}(x_j) + \sum_{k \in R_i} f_k^{(i)}(u_k) + n_i, \qquad (1)$$

where $P_i = \{j \mid (i,j) \in E \wedge x_j \in X\}$ is the set of observable direct causes of v_i, $R_i = \{k \mid (i,k) \in E \wedge u_k \in U\}$ is the set of unobservable direct causes of v_i, $f_j^{(i)}$ is a non-linear function, and n_i is the external noise at v_i.

We assume the following Causal Faithfulness Condition (CFC):

Assumption 1. *Any conditional independence on V that is not entailed by the d-separation criterion on G does not hold.*

CFC is standard in causal discovery methods (Spirtes et al., 2001), and implies Assumption 1 of Maeda and Shimizu (2021) (see Appendix A).

Let $\mathcal{G}$ be the class of GAM regression functions (Hastie and Tibshirani, 1986). For a function $G_i \in \mathcal{G}$, $G_i(N) = \sum_{x_m \in N} g_{i,m}(x_m)$ where each $g_{i,m}(x_m)$ is a nonlinear function of x_m. As in Maeda and Shimizu (2021), we assume that the data-generating process (DGP) satisfies the following residual-faithfulness assumption with respect to $\mathcal{G}$:

Assumption 2. *For any $G_i, G_j \in \mathcal{G}$, any $x_i, x_j \in X$, any external noise n_k, and any $M, N \subseteq X$, $(n_k \not\perp\!\!\!\perp x_i - G_i(M)) \wedge (n_k \not\perp\!\!\!\perp x_j - G_j(N)) \Rightarrow x_i - G_i(M) \not\perp\!\!\!\perp x_j - G_j(N)$.*

The intuition is that, without such a restriction, a property identified using $\mathcal{G}$ may reflect either the underlying causal structure of the DGP or merely an artifact of the model class. Assumption 2 rules out this ambiguity by requiring that whenever two residuals obtained using functions in $\mathcal{G}$ both retain signal from the same exogenous noise variable, they must be statistically dependent. Analogous assumptions are common in functional causal discovery (Ashman et al., 2023).

2.2 Unobserved backdoor paths and unobserved causal paths

Unobserved backdoor paths (UBPs) and unobserved causal paths (UCPs) play central roles in the theory of causal additive models with unobserved variables. We introduce the concepts of UBPs/UCPs with respect to a set $X' \subseteq X$, extending the definitions of Maeda and Shimizu (2021), which were originally given for $X' = X$.

Definition 2.1 (Unobserved Causal Path)**.** *Consider a set of variables $X' \subseteq X$. A path in G is called an unobserved causal path from x_i to x_j with respect to X' if and only if it has the form $x_i \to \cdots \to v_k \to x_j$ with $x_i, x_j \in X'$ and $v_k \notin X'$. When omitted, X' is taken to be X, the full set of observable variables. We use the term UCP between x_i and x_j to denote the existence of a UCP in either direction, from x_i to x_j or from x_j to x_i.*

The path $x_4 \to U_3 \to x_5$ in Fig. 1c is a UCP from x_4 to x_5. The path $x_1 \to U_1 \to x_3 \to x_2$ in Fig. 1d is not a

---

[1] https://github.com/thongphamthe/CAM-UV-X

UCP between x_1 and x_2, but *is* a UCP with respect to the set $X' = X \setminus \{x_3\}$.

Definition 2.2 (Unobserved Backdoor Path)**.** *Consider a set of variables $X' \subseteq X$. For $x_i, x_j \in X'$, a path in G is called an unobserved backdoor path between x_i and x_j with respect to X' if and only if it is of the form $x_i \leftarrow v_k \leftarrow \cdots \leftarrow v_c \rightarrow \cdots \rightarrow v_l \rightarrow x_j$ with $v_k, v_l \notin X'$. When omitted, X' is taken to be X.*

In other words, a UBP is a backdoor path between x_i and x_j where both parents of the endpoints on the path are not in X'.

If a path is a UBP/UCP between x_i and x_j with respect to some $S \subseteq X$, then it is also a UBP/UCP with respect to any $S' \subseteq S$ that contains x_i and x_j.

Consider Fig. 1a. The path $x_3 \leftarrow U_2 \rightarrow x_2$ is a UBP between x_3 and x_2 with respect to X, since both parents of x_3 and x_2 on the path, namely U_2, are not in X. The path $x_1 \leftarrow U_1 \rightarrow x_3 \rightarrow x_2$ is *not* a UBP between x_1 and x_2 with respect to X, since the parent of x_2 on the path, namely x_3, is in X. However, it *is* a UBP with respect to the set $X' = X \setminus \{x_3\}$.

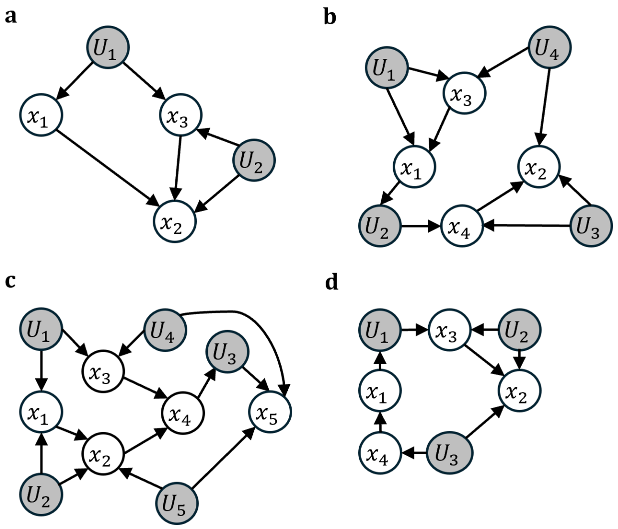

Figure 1: Examples of identifying causal relationships in the presence of UBPs/UCPs.

2.3 Visible and invisible pairs

Based on Lemmas 1-3 of Maeda and Shimizu (2021), we introduce the concepts of visibility/invisibility with respect to a set $X' \subseteq X$, which are fundamental to our theory and our proposed CAM-UV-X. The proofs of all lemmas in this section proceed by the same arguments as those of their corresponding lemmas, which were established for the case $X' = X$. We therefore omit the details.

Lemma 1 (visible parent)**.** *Consider $X' \subseteq X$ and $x_i, x_j \in X'$. x_j is a parent of x_i and there is no UBP or UCP between x_j and x_i with respect to X' if and only if*

$$\forall G_1, G_2 \in \mathcal{G}, M \subseteq X' \setminus \{x_i, x_j\}, N \subseteq X' \setminus \{x_j\}: \quad x_i - G_1(M) \not\perp\!\!\!\perp x_j - G_2(N), \tag{2}$$

$$\exists G_1, G_2 \in \mathcal{G}, M \subseteq X' \setminus \{x_i\}, N \subseteq X' \setminus \{x_i, x_j\}: \quad x_i - G_1(M) \perp\!\!\!\perp x_j - G_2(N). \tag{3}$$

In this case, we refer to x_j as a visible parent of x_i with respect to X'. When omitted, X' is taken to be X.

The regression functions G_1 and G_2 in $\mathcal{G}$ are used only as analytical tools; they need not correspond to the true causal function $f_j^{(i)}$ in Eq. (1). In practice, the function G_i is fitted from observed data by GAM regression (Hastie and Tibshirani, 1986).

Example 1. In Fig. 1a, $x_j = x_1$ is a visible parent of $x_i = x_2$, since there are no UBPs/UCPs between x_1 and x_2. In particular, due to the direct effect of x_1 on x_2, there is no set $M \subseteq X \setminus \{x_1, x_2\} = \{x_3\}$ and no set $N \subseteq X \setminus \{x_1\} = \{x_2, x_3\}$ that can make the residuals $x_2 - G_1(M)$ and $x_1 - G_2(N)$ independent, i.e., Eq. (2) is satisfied. Furthermore, there are some sets $M \subseteq X \setminus \{x_2\} = \{x_1, x_3\}$ and $N \subseteq X \setminus \{x_1, x_2\} = \{x_3\}$, in particular $M = \{x_1, x_3\}$ and $N = \varnothing$, that can make the residuals $x_2 - G_1(M)$ and $x_1 - G_2(N)$ independent, i.e., Eq. (3) is satisfied.

Visible non-edges are defined as follows.

Lemma 2 (visible non-edge)**.** *Consider $X' \subseteq X$ and $x_i, x_j \in X'$. There is no direct edge between x_j and x_i and there is no UBP or UCP between x_j and x_i with respect to X' if and only if*

$$\exists G_1, G_2 \in \mathcal{G}, M \subseteq X' \setminus \{x_i, x_j\}, N \subseteq X' \setminus \{x_i, x_j\}: \quad x_i - G_1(M) \perp\!\!\!\perp x_j - G_2(N). \tag{4}$$

In this case, we refer to (x_i, x_j) as a visible non-edge with respect to X'. When omitted, X' is taken to be X.

For a visible non-edge, independence can emerge even if x_i and x_j are excluded from the regression sets, e.g., the ranges of M and N in Eq. (4) are $X' \setminus \{x_i, x_j\}$. In contrast, for a visible edge $x_j \rightarrow x_i$, due to the direct effect of x_j on x_i, x_j must be included in the regression set of x_i for independence to emerge (see Proposition 2 in Appendix G).

Example 2. In Fig. 1b, (x_1, x_2) is a visible non-edge, since there are no UBPs/UCPs between x_1 and x_2. In this case, the non-edge (x_1, x_2) is identifiable from data by checking Lemma 2: there are some sets $M \subseteq X \setminus \{x_1, x_2\}$ and $N \subseteq X \setminus \{x_1, x_2\}$, in particular $M = \{x_4\}$ and $N = \{x_3\}$, that can make the residuals $x_2 - G_1(M)$ and $x_1 - G_2(N)$ independent, i.e., Eq. (4) is satisfied.

Finally, invisible pairs are defined as follows.

Lemma 3 (invisible pairs)**.** *Consider $X' \subseteq X$ and $x_i, x_j \in X'$. There is a UBP/UCP between x_j and x_i with respect to X' if and only if*

$$\forall M \subseteq X' \setminus \{x_i\}, N \subseteq X' \setminus \{x_j\}, \forall G_1, G_2 \in \mathcal{G} : \\ x_i - G_1(M) \not\perp\!\!\!\perp x_j - G_2(N). \quad (5)$$

In this case, we refer to (x_i, x_j) as an invisible pair with respect to X'. When omitted, X' is taken to be X.

Example 3. Consider Fig. 1a. (x_2, x_3) is invisible due to the presence of a UBP, namely $x_3 \leftarrow U_2 \rightarrow x_2$. In this case, the effects of U_2 on x_2 and x_3 cannot be removed through regressions, and thus there is no set $M \subseteq X \setminus \{x_2\}$ and no set $N \subseteq X \setminus \{x_3\}$ that can make $x_2 - G_1(M)$ and $x_3 - G_2(N)$ independent, i.e., Eq. (5) is satisfied. While (x_1, x_2) is visible with respect to X, it is invisible with respect to $X \setminus \{x_3\}$, because $x_1 \leftarrow U_1 \rightarrow x_3 \rightarrow x_2$ is a UBP with respect to $X \setminus \{x_3\}$.

The CAM-UV algorithm (Maeda and Shimizu, 2021) is designed to detect visible edges and non-edges with respect to X while marking invisible pairs as such. However, it cannot identify causal directions in invisible pairs.

Remark 1. *The visibility and invisibility results with respect to X extend beyond CAM-UV to the separable observed/unobserved-parent model studied by Ashman et al. (2023), but they do not extend to the general additive noise models (ANMs) of Hoyer et al. (2008a). The corresponding results with respect to a subset $X' \subset X$ hold only in CAM-UV. See Appendix B for further discussion.*

3 IDENTIFYING CAUSAL RELATIONSHIPS IN INVISIBLE PAIRS

We present conditions for identifying parent–child relationships or causal directions in invisible pairs. All omitted proofs can be found in Appendix G.

3.1 Utilizing independence between regression residuals

In this section, new conditions are provided by characterizing the content of the regression sets M and N in Lemmas 1 and 2 with the following intuition.

Proposition 1. *If (x_i, x_j) is visible with respect to X and invisible with respect to $X \setminus \{x_k\}$, x_k must be a parent of x_i or x_j.*

Visibility with respect to X and invisibility with respect to $X \setminus \{x_{k_q}\}$ can be confirmed using Lemmas 1, 2 and 3. This allows us to conclude that x_{k_q} must be a parent of x_i or x_j, even when (x_{k_q}, x_i) and (x_{k_q}, x_j) are both invisible.

The following lemma characterizes the regression sets in Eq. (3).

Lemma 4. *Consider distinct x_i, x_j, and $x_{k_1}, \dots, x_{k_m}$. Let $K = \{x_{k_1}, \dots, x_{k_m}\}$. x_j is a visible parent of x_i, and for each x_{k_q}, (x_i, x_j) is invisible with respect to $X \setminus \{x_{k_q}\}$ if and only if*

$$\textit{For } q = 1, \dots, m : \forall M \subseteq X \setminus \{x_i, x_{k_q}\}, \\ N \subseteq X \setminus \{x_j, x_{k_q}\}, \forall G_i^1, G_j^1 \in \mathcal{G} : \\ x_i - G_i^1(M) \not\perp\!\!\!\perp x_j - G_j^1(N), \quad (6)$$

$$\forall M \subseteq X \setminus \{x_i, x_j\}, N \subseteq X \setminus \{x_j\}, \forall G_i^1, G_j^1 \in \mathcal{G} : \\ x_i - G_i^1(M) \not\perp\!\!\!\perp x_j - G_j^1(N), \quad (7)$$

$$\exists Q_1, Q_2 \subseteq K : Q_1 \cup Q_2 = K : \\ \exists G_i^2, G_j^2 \in \mathcal{G}, M, N \subseteq X \setminus \{x_i, x_j\} \setminus K : \\ x_i - G_i^2 \left(M \cup \{x_j\} \cup Q_1\right) \perp\!\!\!\perp x_j - G_j^2(N \cup Q_2) \quad (8)$$

are satisfied. Since x_j is a parent of x_i, each x_{k_q} is therefore an ancestor of x_i.

Eq. (6) means that for each node x_{k_q} in K, the residuals cannot be independent if one excludes x_{k_q} from X. Eq. (8) means that the residuals become independent when one includes each x_{k_q} into the regression sets. Eqs. (6) and (8) together imply that each x_{k_q} must be on a UBP/UCP between x_i and x_j with respect to $X \setminus \{x_{k_q}\}$, and x_{k_q} must be a parent of x_i or a parent of x_j. Eq. (7) means that the residuals cannot be independent if one does not add x_j to the regression set when regressing x_i. Eqs. (7) and (8) would imply that x_j is a parent of x_i. Eqs. (6), (7), and (8) together give the lemma.

Example 4. The pair (x_3, x_2) in Fig. 1a is invisible, due to the presence of the UBP $x_3 \leftarrow U_2 \rightarrow x_2$. Thus, existing theories do not provide sufficient conditions for certifying causal relationships in (x_3, x_2). If Eqs. (6), (7), and (8) hold with $x_i = x_2$, $x_j = x_1$, and $K = \{x_3\}$, Lemma 4 can certify that x_3 is an ancestor of x_2.

The full separability of causal effects in CAM-UV is crucial for our theory and may be the additional structural constraint that enables identification in cases not covered by Schultheiss and Bühlmann (2024). Additionally, they aim to identify variables whose causal effect on a target variable is identifiable, which is generally a harder problem than just identifying the ancestors of the target. For example, Hoyer et al. (2008b) showed examples where the causal direction is identifiable, even when the causal effect is not.

For visible non-edges, Lemma 5 characterizes the regression sets in Eq. (4).

Lemma 5. *Consider distinct x_i, x_j, and $x_{k_1}, \dots, x_{k_m}$. Let $K = \{x_{k_1}, \dots, x_{k_m}\}$. (x_j, x_i) is a visible non-edge, and for each x_{k_q}, (x_i, x_j) is invisible with respect to $X \setminus \{x_{k_q}\}$*

if and only if Eq. (6) and

$$\exists Q_1, Q_2 \subseteq K : Q_1 \cup Q_2 = K :$$
$$\exists G_i^2, G_j^2 \in \mathcal{G}, M, N \subseteq X \setminus \{x_i, x_j\} \setminus K :$$
$$x_i - G_i^2\,(M \cup Q_1) \perp\!\!\!\perp x_j - G_j^2(N \cup Q_2), \qquad (9)$$

are satisfied.

Example 5. Consider Fig. 1b. Since (x_3, x_1), (x_3, x_2), (x_4, x_1), and (x_4, x_2) are invisible, existing theories cannot identify causal relationships in these pairs. If Eqs. (6) and (9) hold in the observed data with $x_i = x_1$, $x_j = x_2$, and $K = \{x_3, x_4\}$, Lemma 5 can certify that x_3 is a parent of x_1 or x_2, and x_4 is a parent of x_1 or x_2.

See Appendix G.1 for Lemma 9, which provides an alternative condition to Lemma 5.

The presence of a bow between two variables poses a major challenge for causal discovery. For example, in non-Gaussian linear models, the parent–child relationship within a bow cannot be determined without additional assumptions, such as adopting a canonical model where every hidden variable is parentless and acts as a confounder of at least two observed variables (Hoyer et al., 2008b; Tramontano et al., 2024; Chen et al., 2024), or assuming that the number of hidden variables is known (Adams et al., 2021). Without such assumptions, only the causal direction, i.e., the ancestor relationship, is identifiable (Salehkaleybar et al., 2020). To our knowledge, no existing causal model can identify the parent–child relationship within a bow without imposing additional assumptions on the graph structure.

Lemmas 4, 5, or 9 imply that x_{k_q} must be a parent of x_i or x_j, without specifying which. If we can rule out x_{k_q} as a parent of x_i, for example, when (x_{k_q}, x_i) is a visible non-edge or $x_i \rightarrow x_{k_q}$ is a visible edge, then x_{k_q} must be a parent of x_j. Even when both (x_{k_q}, x_i) and (x_{k_q}, x_j) are invisible and form bows, the regression approach can sometimes still resolve the relationship (see Appendix C). When it cannot, conditional independence among the original variables may provide the needed information.

3.2 Utilizing conditional independence

We provide sufficient conditions utilizing independence between regression residuals and conditional independence between the original variables to identify causal relationships in invisible pairs. We show some examples where the causal direction is identifiable by this hybrid approach, but not by the regression approach alone or conditional independence alone.

We introduce the following lemma for identifying that some x_j is not an ancestor of x_i.

Lemma 6. *If x_k is an ancestor of x_i and $x_k \perp\!\!\!\perp x_j \mid x_i$, then 1) there is no backdoor path between x_i and x_j, and 2) x_j is not an ancestor of x_i.*

Lemma 6 is a direct consequence of standard d-separation reasoning (Spirtes et al., 2001) and not new. It is closely related to local causal-discovery arguments based on Y-structures (Mani et al., 2006). We therefore state it here only in the particular form needed for our setting.

Combining Lemma 6 with independence conditions between regression residuals can give sufficient conditions to identify the causal direction in certain invisible pairs in the following corollary.

Corollary 6.1. *Consider an invisible pair (x_i, x_j), i.e., Eq. (5) is satisfied. If x_k is an ancestor of x_i and $x_k \perp\!\!\!\perp x_j \mid x_i$, then x_i is an ancestor of x_j.*

Example 6. In Fig. 1c, consider the invisible pair (x_4, x_5). Since this pair is not on any backdoor path or causal path of any visible pair, Lemmas 4, 5, or 9 cannot identify the causal direction in this pair. Since (x_4, x_5) is invisible, Eq. (5) is satisfied for $x_i = x_4$ and $x_j = x_5$. From the visible pairs (x_2, x_4) and (x_3, x_4), x_2 and x_3 are identified as parents of x_4 by Lemma 1. If Lemma 5 is satisfied with $x_i = x_3$, $x_j = x_2$, and $K = \{x_1\}$, x_1 can be identified as a parent of x_3 or a parent of x_2, and thus is an ancestor of x_4. If $x_1 \perp\!\!\!\perp x_5 \mid x_4$, x_4 can be identified as an ancestor of x_5 by applying Corollary 6.1 with $x_k = x_1$.

Remark 2. *See Appendix D for an example in which Corollary 6.1 can identify the causal direction in an invisible pair, whereas neither the regression approach alone nor conditional independence alone, e.g., FCI, can.*

Lemma 6 can also identify parent–child relationships in invisible pairs, as in the following corollary.

Corollary 6.2. *Suppose that x_k is a parent of x_i or a parent of x_j (e.g., by Lemmas 4, 5, or 9). Furthermore, for a fourth variable x_u, if x_u is an ancestor of x_i and $x_u \perp\!\!\!\perp x_k \mid x_i$, then x_k is a parent of x_j.*

Example 7. Consider the bow (x_2, x_3) in Fig. 1d. Corollary 6.1 cannot identify the causal direction in this pair. If Lemma 5 is satisfied with $x_i = x_1$, $x_j = x_2$, and $K = \{x_3\}$, one can identify that x_3 is a parent of x_1 or a parent of x_2. Note that this is a hard case where the regression approach alone cannot resolve. The edge $x_4 \rightarrow x_1$ is visible and thus can be identified. If $x_4 \perp\!\!\!\perp x_3 \mid x_1$ holds, applying Corollary 6.2 with $x_k = x_3$, $x_i = x_1$, $x_j = x_2$, and $x_u = x_4$ identifies x_3 as a parent of x_2.

Remark 3. *This is an example in which regression and conditional independence together can identify the parent–child relationship in a bow that remains unidentifiable by regression alone.*

Remark 4. *Corollary 6.2 relies on the pair (x_i, x_j) being visible with respect to X but invisible with respect to $X \setminus \{x_k\}$. Consequently, it does not extend beyond CAM-UV.*

By contrast, Corollary 6.1 extends beyond CAM-UV to a slightly larger class of models considered in Ashman et al. (2023); see Appendix B for further discussion.

4 SEARCH METHODS

In Section 4.1, we discuss previously overlooked limitations of the CAM-UV algorithm, which can lead to incorrect identification of some visible pairs as invisible. Our proposed method is presented in Section 4.2.

4.1 Limitations of the CAM-UV algorithm

The CAM-UV algorithm is incomplete in identifying visible pairs and unsound in identifying invisible pairs. Soundness and completeness (Spirtes et al., 2001) can be defined in our setting as follows. An algorithm is complete for visible edges if and only if every visible edge in the ground truth also appears as a visible edge in the algorithm's output. Conversely, an algorithm is sound for visible edges if and only if every visible edge in the output also exists in the ground truth. Analogous definitions apply to visible non-edges and invisible pairs.

Fig. 1a illustrates that CAM-UV might misclassify a visible edge as invisible, making it incomplete for visible edges and unsound for invisible pairs. Although the visible edge $x_1 \to x_2$ is identifiable by Lemma 1, the presence of the invisible parent x_3 of x_2 leads CAM-UV to wrongly label it as invisible.

Fig. 1b illustrates that CAM-UV might misclassify some visible non-edge as invisible, making it also incomplete for visible non-edges. The visible non-edge (x_1, x_2) is identifiable by Lemma 2. However, to block all backdoor and causal paths between x_1 and x_2, one must add x_3 and x_4 to the regression sets in Eq. (4). Furthermore, the pairs (x_1, x_3), (x_1, x_4), (x_2, x_3), and (x_2, x_4) are all invisible. In this case, CAM-UV will wrongly label the visible non-edge as invisible.

See Appendix H for step-by-step executions of CAM-UV in these examples.

4.2 Proposed search method

Our proposed method CAM-UV-X, described in Algorithm 1, addresses the limitations of CAM-UV in identifying visible pairs, and leverages the new theory in this paper to infer parent–child relationships and causal directions in invisible pairs.

CAM-UV-X additionally uses Lemmas 8 and 9 in Appendix G.1. Lemma 8 provides sufficient conditions to certify that (x_i, x_j) is visible and x_i is not a parent of x_j. If one can certify non-parent relations in both directions, then one can conclude that (x_i, x_j) is a visible non-edge, i.e., a different route than using Lemma 2. Lemma 9 utilizes this new visible non-edge condition in the same way as Lemma 5.

The output of CAM-UV-X is A, $M_1, \dots, M_p$, $H_1, \dots, H_p$, and $C_1, \dots, C_p$. A is the adjacency matrix over the observed variables. $A(i,j) = 1$ if x_j is inferred to be a parent of x_i, 0 if there is no directed edge from x_j to x_i, and NaN (Not a Number) if (x_i, x_j) is inferred to be invisible. M_i is the set of ancestors of x_i identified by Lemma 4 and Corollary 6.1. H_i is the set of nodes guaranteed not to be an ancestor of x_i, identified by, for example, Lemma 6. C_k contains ordered pairs $[i, j]$ such that x_k is a parent of x_i or x_k is a parent of x_j. See Appendix F for the running time of CAM-UV-X.

Algorithm 1 CAM-UV-X

```
Input: $n \times p$ data matrix $X$ for $p$ observed variables, maximum number of parents $d$, significance level $\alpha$
Output: $A, \{M_1, \dots, M_p\}, \{H_1, \dots, H_p\}, \{C_1, \dots, C_p\}$
 1: $A \leftarrow$ CAM-UV$(X, d, \alpha)$
 2: Initialize $M_i \leftarrow \varnothing, H_i \leftarrow \varnothing, C_i \leftarrow \varnothing$ for $i = 1, \dots, p$
 3: Find the set $S = \{(i,j) \mid A(i,j) = A(j,i) = \text{NaN}\}$
 4: for each $(i,j) \in S, i < j$ do
 5:     checkVisible(i,j)
 6: end for
 7: Find the set $I = \{(i,j) \mid A(i,j) \neq \text{NaN}; A(j,i) \neq \text{NaN}\}$
 8: for each $(i,j) \in I, i < j$ do
 9:     checkOnPath(i,j)
10: end for
11: Find the set $S = \{(i,j) \mid A(i,j) = A(j,i) = \text{NaN}\}$
12: for each $(i,j) \in S$ do
13:     checkCI(i,j)
14: end for
15: checkParentInvi()
```

In line 1, CAM-UV is executed to obtain an initial estimate of A. The procedure `checkVisible`, described in Algorithm 2, tests whether each NaN element in the current matrix A can be converted to 1, that is, a visible edge, or to 0, i.e., a visible non-edge.

On line 6 of `checkVisible`, Lemma 2 is checked. To find the functions G_i, we use GAM regression implemented in pyGAM (Servén and Brummitt, 2018). Eq. (4) is satisfied if the value e, calculated on line 5, is greater than α. Here, $\widehat{\text{p-HSIC}}$ is the p-value of the gamma independence test based on Hilbert–Schmidt Independence Criterion (Gretton et al., 2007). If so, (x_i, x_j) is concluded to be a visible non-edge.

If the condition on line 6 is not satisfied, the p-HSIC values a_1 and a_2 are calculated. If $a_1 > \alpha$ in line 11, Eq. (11) is satisfied, and we conclude that x_i is not a parent of x_j in line 12, due to Lemma 8 (see Appendix G.1). We proceed

Algorithm 2 `checkVisible`

```
Input: indices $i$ and $j$
 1: $P_i \leftarrow \{v \mid A(i,v) = 1\}$; $P_j \leftarrow \{v \mid A(j,v) = 1\}$
 2: $Q \leftarrow \{k \mid k \notin \{i,j\},\ A(j,k) = \text{NaN or } A(i,k) = \text{NaN}\} \cup P_i \cup P_j$
 3: $iNotParent \leftarrow False$; $jNotParent \leftarrow False$
 4: for each $M \subseteq Q$ and $N \subseteq Q$ do
 5:     $e \leftarrow \widehat{\text{p-HSIC}}\big(x_i - G_1(M), x_j - G_2(N)\big)$
 6:     if $e > \alpha$ then
 7:         $A(i,j) \leftarrow 0$; $A(j,i) \leftarrow 0$; return
 8:     else
 9:         $a_1 \leftarrow \widehat{\text{p-HSIC}}\big(x_i - G_1(M \cup \{x_j\}),\ x_j - G_2(N)\big)$
10:         $a_2 \leftarrow \widehat{\text{p-HSIC}}\big(x_i - G_1(M),\ x_j - G_2(N \cup \{x_i\})\big)$
11:         if $a_1 > \alpha$ then
12:             $iNotParent \leftarrow True$
13:         end if
14:         if $a_2 > \alpha$ then
15:             $jNotParent \leftarrow True$
16:         end if
17:         if $iNotParent \wedge jNotParent$ then
18:             $A(i,j) \leftarrow 0$; $A(j,i) \leftarrow 0$; return
19:         end if
20:     end if
21: end for
22: if $iNotParent$ then
23:     $A(i,j) \leftarrow 1$; $A(j,i) \leftarrow 0$
24: end if
25: if $jNotParent$ then
26:     $A(j,i) \leftarrow 1$; $A(i,j) \leftarrow 0$
27: end if
```

similarly to a_2. If x_i is not a parent of x_j and x_j is not a parent of x_i, i.e., line 17 is executed, we conclude that (x_i, x_j) is a visible non-edge, due to Lemma 8.

If line 22 is reached, Eq. (2) is satisfied. Strictly speaking, we only check subsets in Q, not all subsets of X as in Eq. (2). Reassuringly, this is enough; see Appendix G.10. Then, if x_i is not a parent of x_j, i.e., the condition on line 22 is satisfied, or if x_j is not a parent of x_i, i.e., the condition on line 25 is satisfied, we conclude that (x_i, x_j) is a visible edge due to Lemma 1. If the conditions in both line 22 and line 25 are not satisfied, we conclude that (x_i, x_j) is invisible.

The algorithm `CAM-UV-X` then invokes the procedure `checkOnPath`, described in Algorithm 3, to check Eq. (6) for every visible pair (x_i, x_j) and every x_{k_q} whose parent–child relationship to x_i or x_j is invisible. If the equation is satisfied, i.e., $isOnPath$ is $True$ at line 10, x_{k_q} is concluded to be a parent of x_i or a parent of x_j, due to Lemmas 4, 5, and 9.

Algorithm 3 `checkOnPath`

```
Input: indices $i, j$
Output: Boolean value $isOnPath$
 1: for each $x_k \in X \setminus \{x_i, x_j\}$ that satisfies $A(i,k) = \text{NaN}$ or $A(j,k) = \text{NaN}$ do
 2:     $isOnPath \leftarrow True$
 3:     for each $M \subseteq X \setminus \{x_i, x_k\}$ and $N \subseteq X \setminus \{x_j, x_k\}$ do
 4:         $a \leftarrow \widehat{\text{p-HSIC}}\big(x_i - G_1(M), x_j - G_2(N)\big)$
 5:         if $a > \alpha$ then
 6:             $isOnPath \leftarrow False$;
 7:             break
 8:         end if
 9:     end for
10:     if $isOnPath$ then
11:         $C_k \leftarrow C_k \cup \{[i,j]\} \cup \{[j,i]\}$
12:         if A(i,j) = 1 then
13:             $M_i \leftarrow M_i \cup \{x_k\}$; $H_k \leftarrow H_k \cup \{x_i\}$; $A(k,i) \leftarrow 0$
14:         else if A(j,i) = 1 then
15:             $M_j \leftarrow M_j \cup \{x_k\}$; $H_k \leftarrow H_k \cup \{x_j\}$; $A(k,j) \leftarrow 0$
16:         end if
17:     end if
18: end for
```

The procedure `checkCI`, described in Algorithm 4, checks conditional independence of the form $x_k \perp\!\!\!\perp x_j \mid x_i$ with (x_i, x_j) being invisible, and x_k being an ancestor of x_i. $\widehat{\text{p-CI}}$ is the p-value of some conditional independence test. Some examples are the conditional mutual information test based on nearest-neighbor estimator (CMIknn from Runge (2018)) and the conditional independence test based on Gaussian process regression and distance correlations (GPDC from Székely et al. (2007)). If conditional independence is satisfied (line 3), x_i is an ancestor of x_j due to Lemma 6 and Corollary 6.1.

Algorithm 4 `checkCI`

```
Input: indices $i$ and $j$
 1: for each currently identified ancestor $x_k$ of $x_i$ do
 2:     $e \leftarrow \widehat{\text{p-CI}}(x = x_k, y = x_j, z = x_i)$
 3:     if $e > \alpha$ then
 4:         $H_i \leftarrow H_i \cup \{x_j\}$; $A(i,j) \leftarrow 0$; $M_j \leftarrow M_j \cup \{x_i\}$
 5:     end if
 6: end for
```

In `checkParentInvi`, described in Algorithm 5, for each ordered pair $[i,j] \in C_k$ such that $A(j,k) = \text{NaN}$, we check a) whether x_k is not an ancestor of x_i by checking whether x_k is in H_i and b) whether x_k is not a parent of x_i by checking whether $A(i,k) = 0$. If either case is true, we conclude that x_k is a parent of x_j. The loop repeats until there is no new change.

Algorithm 5 `checkParentInvi`

```
repeat
    $isChange \leftarrow False$
    for $k = 1, \ldots, p$ do
        for each ordered pair $[i, j] \in C_k$ do
            if $(A(j,k) = \text{NaN}) \wedge (x_k \in H_i \vee A(i,k) = 0)$ then
                $A(j,k) \leftarrow 1; A(k,j) \leftarrow 0$
                $isChange \leftarrow True$
            end if
        end for
    end for
until $isChange = False$
```

To prove the correctness of CAM-UV-X in identifying visibilities, we need the following assumption.

Assumption 3. *For any $x_i, x_j \in X$, $M \subseteq X \setminus \{x_i\}$, $N \subseteq X \setminus \{x_j\}$, let $e = \widehat{p\text{-}HSIC}\big(x_i - G_i(M), x_j - G_j(N)\big)$, where G_i, G_j are GAM regression functions fitted by pyGAM. For a given α, the following equation holds:*

$$\exists G_1, G_2 \in \mathcal{G} : x_i - G_1(M) \perp\!\!\!\perp x_j - G_2(N) \iff e > \alpha.$$

This assumption means that the pyGAM regression and the HSIC independence test can identify the regression functions in $\mathcal{G}$. Note that CAM-UV is not complete in identifying visible pairs, and is not sound in identifying invisible pairs, even with this assumption. The following theorem, whose proof can be found in Appendix G, holds.

Theorem 7. *With Assumptions 1, 2, and 3, CAM-UV-X is sound and complete in identifying visible edges, visible non-edges, and invisible pairs.*

Note that Theorem 7 does not assert soundness or completeness for identifying the causal direction within an invisible pair. Section 3 provides sufficient conditions for such identification, and therefore CAM-UV-X is sound whenever it orients an invisible pair. Completeness, however, remains open, even when restricting the class of graphs G to exclude obvious non-identifiable cases such as a standalone bow. Indeed, there may exist bows that are resolvable in principle but not by the current theory. We are unaware of any such example, but we also cannot rule out this possibility, since that would require a characterization of the necessary conditions under which such graphical structures arise, which is currently unknown.

See Appendix E for a discussion of related works.

5 EXPERIMENTS

5.1 Illustrative examples

We demonstrate that CAM-UV-X can address the limitations of CAM-UV discussed in the previous section by applying it to Figs. 1a and 1b. The DGP in Eq. (1) is defined as follows. The function $f_j^{(i)}(x_j)$ is set to $(x_j + a)^c + b$ with random coefficients a, b, and c. Note that this function belongs to GAMs. We use the same setting for $f_k^{(i)}$. The noise terms n_i are Gaussian. For each graph, 100 datasets, each of 500 samples, were generated. We ran CAM-UV and CAM-UV-X with the significance level $\alpha = 0.1$. In CAM-UV-X, we used CMIknn from Runge (2018) as the conditional independence estimator.

We measured the success rate of identifying the visible edge $x_1 \rightarrow x_2$ and x_3 as an ancestor of x_2 in Fig. 1a, and identifying the visible non-edge (x_1, x_2) in Fig. 1b. Figure 2 shows the results. CAM-UV-X successfully addresses the limitations of CAM-UV discussed in Section 4.1.

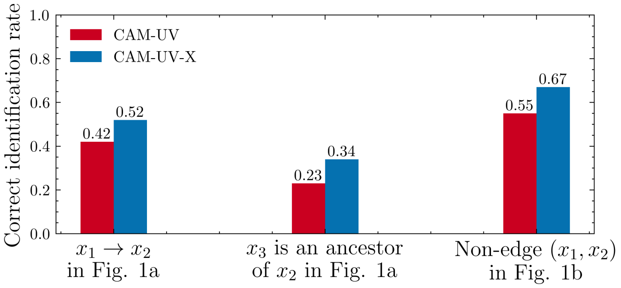

Figure 2: Performance on the graphs in Figs. 1a and 1b.

5.2 Random graph experiment

We investigate CAM-UV-X using simulated data generated from Eq. (1). We report the results for the Barabási–Albert (BA) model (Barabási and Albert, 1999) here; results for the Erdös-Rényi (ER) model (Erdös and Rényi, 1959) are deferred to Appendix I.

We generate 50 BA graphs with 40 nodes, where each node has five children. For each graph, we create data using the same process as in Section 5.1. We then randomly select 10 variables and form the final dataset using only these 10 variables. Each dataset contains 500 samples. In addition to CAM-UV, we also include an adaptation of Schultheiss and Bühlmann (2024), called S-B, as another baseline. See Appendix I for details.

We ran the algorithms with significance levels $\alpha = 0.05$, 0.1, and 0.2. We show results of two tasks: identifying the adjacency matrix and identifying ancestor relations between the observed variables, in Figs. 3a and 3b, respectively. See Appendix I for definitions of evaluation metrics used in each task.

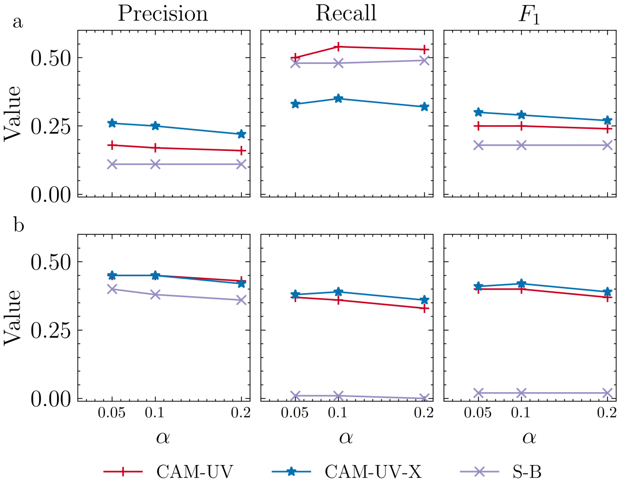

Figure 3: Performance in BA random graphs with Gaussian noise. a: identifying the adjacency matrix, b: identifying ancestral relationships.

For the identification of the adjacency matrix, CAM-UV-X achieved higher precision and F1 scores than CAM-UV. For the identification of ancestor relationships, CAM-UV-X achieved higher recall and F1 scores than CAM-UV. The remaining results in Appendix I.4 also show that CAM-UV-X compares favorably with the baselines.

5.3 Real-world sociological data

We present in Fig. 4 the results of applying CAM-UV and CAM-UV-X to a sociological dataset (Shimizu et al., 2011). The data consisted of six observed variables: father's occupation (x_1), son's income (x_2), father's education (x_3), son's occupation (x_4), son's education (x_5), and number of siblings (x_6).

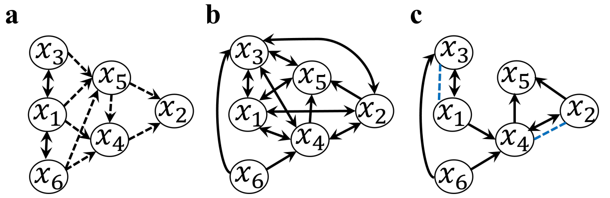

Figure 4: Results on sociological data. **a**: ground truth based on domain knowledge (Duncan et al., 1972). A bidirected edge indicates that the relationship is not modeled. A dashed directed edge represents an ancestor relationship. **b**: result by CAM-UV. A solid directed edge denotes a visible parent–child relationship (adjacency). An empty edge denotes a visible non-edge (non-adjacency). A bidirected edge denotes an invisible pair. **c**: result of CAM-UV-X. Solid directed/bidirected edges and empty edges have the same meaning as those of CAM-UV. A dashed undirected edge in blue, unique to CAM-UV-X and appearing only in an invisible pair connected by a bidirected edge, indicates non-adjacency.

Compared to CAM-UV, CAM-UV-X pruned more spurious bidirected edges, namely edges that CAM-UV should have pruned if it were sound in identifying invisible pairs. For example, CAM-UV-X replaced the bidirected edge between x_1 and x_4 in the CAM-UV result by a directed edge from x_1 to x_4, which is consistent with the ground truth.

Furthermore, CAM-UV-X can provide more information for invisible pairs than CAM-UV. For example, for the pairs (x_1, x_3) and (x_2, x_4), while there is a UBP/UCP between these pairs, CAM-UV-X asserts that each pair is not adjacent. See Appendix I.5 for details of the experiment settings.

6 CONCLUSIONS

We provided new identifiability results for causal additive models with hidden variables, including cases where bows are fully identified. Furthermore, we introduced the CAM-UV-X algorithm, proved its soundness and completeness, and demonstrated that it performs comparably to state-of-the-art methods.

Our current theory does not fully utilize all constraints implied by d-separation among the observed variables. A more refined characterization of the regression sets that are sufficient or necessary to certify visibility or invisibility could lead to more efficient algorithms. Furthermore, systematically investigating the gap between sufficient and necessary conditions in CAMs with hidden variables remains an important open problem. Finally, extending the use of conditional independence between observed variables and regression residuals to other classes of causal models offers a promising direction for future research.

Acknowledgements

This work was partially supported by the Japan Science and Technology Agency (JST) under CREST Grant Number JPMJCR22D2 and by the Japan Society for the Promotion of Science (JSPS) under KAKENHI Grant Numbers JP24K20741 and JP25K03084.

References

J. Adams, N. Hansen, and K. Zhang (2021). Identification of partially observed linear causal models: Graphical conditions for the non-Gaussian and heterogeneous cases. *Advances in Neural Information Processing Systems*.

M. Ashman, C. Ma, A. Hilmkil, J. Jennings, and C. Zhang (2023). Causal reasoning in the presence of latent confounders via neural ADMG learning. *International Conference on Learning Representations*.

A.-L. Barabási and R. Albert (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

R. Bhattacharya, T. Nagarajan, D. Malinsky, and I. Shpitser (2021). Differentiable causal discovery under unmeasured confounding. *International Conference on Artificial Intelligence and Statistics*.

K. Budhathoki, L. Minorics, P. Blöbaum, and D. Janzing (2022). Causal structure-based root cause analysis of outliers. *International Conference on Machine Learning*.

P. Bühlmann, J. Peters, and J. Ernest (2014). CAM: Causal additive models, high-dimensional order search and penalized regression. *Annals of Statistics*, 42(6):2526–2556.

T. Chen and C. Guestrin (2016). XGBoost: A scalable tree boosting system. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

W. Chen, Z. Huang, R. Cai, Z. Hao, and K. Zhang (2024). Identification of causal structure with latent variables based on higher order cumulants. *AAAI Conference on Artificial Intelligence*.

D. M. Chickering (2002). Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3(Nov):507–554.

O. D. Duncan, D. L. Featherman, and B. Duncan (1972). *Socioeconomic Background and Achievement*. Seminar Press, New York.

P. Erdös and A. Rényi (1959). On random graphs I. *Publicationes Mathematicae Debrecen*, 6:290.

L. Ge, H. Cai, R. Wan, Y. Xu, and R. Song (2025). A review of causal decision making. *arXiv*, arXiv:2502.16156.

C. Glymour, K. Zhang, and P. Spirtes (2019). Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, 10:524.

A. Gretton, K. Fukumizu, C. Teo, L. Song, B. Schölkopf, and A. Smola (2007). A kernel statistical test of independence. *Advances in Neural Information Processing Systems*.

T. Hastie and R. Tibshirani (1986). Generalized additive models. *Statistical Science*, 1(3):297 – 310.

P. O. Hoyer, D. Janzing, J. Mooij, J. Peters, and B. Schölkopf (2008a). Nonlinear causal discovery with additive noise models. *Advances in Neural Information Processing Systems*.

P. O. Hoyer, S. Shimizu, A. Kerminen, and M. Palviainen (2008b). Estimation of causal effects using linear non-Gaussian causal models with hidden variables. *International Journal of Approximate Reasoning*, 49(2):362–378.

T. N. Maeda and S. Shimizu (2020). RCD: Repetitive causal discovery of linear non-Gaussian acyclic models with latent confounders. *International Conference on Artificial Intelligence and Statistics*.

T. N. Maeda and S. Shimizu (2021). Causal additive models with unobserved variables. *Conference on Uncertainty in Artificial Intelligence*.

S. Mani, P. Spirtes, and G. F. Cooper (2006). A theoretical study of Y structures for causal discovery. *Conference on Uncertainty in Artificial Intelligence*.

J. M. Ogarrio, P. Spirtes, and J. Ramsey (2016). A hybrid causal search algorithm for latent variable models. *Conference on Probabilistic Graphical Models*.

J. Peters, J. M. Mooij, D. Janzing, and B. Schölkopf (2011). Identifiability of causal graphs using functional models. *Conference on Uncertainty in Artificial Intelligence*.

J. Runge (2018). Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information. *International Conference on Artificial Intelligence and Statistics*.

S. Salehkaleybar, A. Ghassami, N. Kiyavash, and K. Zhang (2020). Learning linear non-Gaussian causal models in the presence of latent variables. *Journal of Machine Learning Research*, 21(39):1–24.

B. Schölkopf (2022). *Causality for Machine Learning*. Association for Computing Machinery, New York, NY, USA, 1st edition.

C. Schultheiss and P. Bühlmann (2024). Assessing the overall and partial causal well-specification of nonlinear additive noise models. *Journal of Machine Learning Research*, 25(159):1–41.

D. Servén and C. Brummitt (2018). pyGAM: Generalized additive models in Python. URL https://doi.org/10.5281/zenodo.1208723.

S. Shimizu, P. O. Hoyer, A. Hyvärinen, and A. Kerminen (2006). A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7: 2003–2030.

S. Shimizu, T. Inazumi, Y. Sogawa, A. Hyvärinen, Y. Kawahara, T. Washio, P. O. Hoyer, and K. Bollen (2011). DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. *Journal of Machine Learning Research*, 12:1225–1248.

P. Spirtes and C. Glymour (1991). An algorithm for fast recovery of sparse causal graphs. *Social Science Computer Review*, 9:67–72.

P. Spirtes, C. Meek, and T. Richardson (1995). Causal inference in the presence of latent variables and selection bias. *Conference on Uncertainty in Artificial Intelligence*.

P. Spirtes, C. Glymour, and R. Scheines (2001). *Causation, Prediction, and Search*. MIT Press, 2nd edition.

H. Suzuki, K. Kanamori, T. Takagi, T. Pham, T.N. Maeda, and S. Shimizu (2026). I-CAM-UV: Integrating causal graphs over non-identical variable sets using causal additive models with unobserved variables. *AAAI Conference on Artificial Intelligence*.

G. J. Székely, M. L. Rizzo, and N. K. Bakirov (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769 – 2794.

T. Tashiro, S. Shimizu, A. Hyvärinen, and T. Washio (2014). ParceLiNGAM: A causal ordering method robust against latent confounders. *Neural Computation*, 26(1):57–83.

D. Tramontano, Y. Kivva, S. Salehkaleybar, M. Drton, and N. Kiyavash (2024). Causal effect identification in LiNGAM models with latent confounders. *International Conference on Machine Learning*.

Y. S. Wang and M. Drton (2023). Causal discovery with unobserved confounding and non-Gaussian data. *Journal of Machine Learning Research*, 24(271):1-61.

H. Yokoyama, R. Shingaki, K. Nishino, S. Shimizu, and T. Pham (2025). Causal-discovery-based root-cause analysis and its application in time-series prediction error diagnosis. *International Joint Conference on Neural Networks*.

K. Zhang, B. Schölkopf, and D. Janzing (2010). Invariant Gaussian process latent variable models and application in causal discovery. *Conference on Uncertainty in Artificial Intelligence*.

X. Zheng, B. Aragam, P. K. Ravikumar, and E. P. Xing (2018). DAGs with NO TEARS: Continuous optimization for structure learning. *Advances in Neural Information Processing Systems*.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   (d) A description of the computing infrastructure used (e.g., type of GPUs, internal cluster, or cloud provider). [Yes]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator if your work uses existing assets. [Not Applicable]
   (b) The license information of the assets, if applicable. [Yes]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Yes]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Causal Additive Models with Unobserved Causal Paths and Backdoor Paths: Supplementary Materials

A ASSUMPTIONS OF THE CAM-UV MODEL

A.1 CFC implies Assumption 1 of Maeda and Shimizu (2021)

Assumption 1 of Maeda and Shimizu (2021) is stated as the following Assumption 1b.

Assumption 1b. If variables v_i and v_j have terms involving functions of the same external effect n_k, then v_i and v_j are mutually dependent, i.e., $(n_k \not\perp\!\!\!\perp v_i) \wedge (n_k \not\perp\!\!\!\perp v_j) \Rightarrow v_i \not\perp\!\!\!\perp v_j$.

In our paper, we assume the standard CFC and Assumption 2. Since the proofs of Lemmas 1, 2, and 3 in Maeda and Shimizu (2021) rely crucially on Assumption 1b, we prove that CFC implies Assumption 1b as follows.

Proof. Assume CFC. If variables v_i and v_j have terms involving functions of the same external effect n_k, there is a confounder between v_i and v_j, or a directed path from v_i to v_j, or a directed path from v_j to v_i. This implies that v_i and v_j are not d-separated by the empty set. By CFC, this implies that v_i is not independent of v_j. Thus, Assumption 1b is true under CFC. This means CFC implies Assumption 1b. ☐

B SCOPE OF THE VISIBILITY AND INVISIBILITY RESULTS

Ashman et al. (2023) extended Lemmas 1-3 of Maeda and Shimizu (2021), i.e., visibility/invisibility with respect to X, to a broader class of models where the causal effects between the observed and unobserved parents are separable, while the effects among the observed parents might not be separable:

$$v_i = f_i(P_i) + g_i(R_i) + n_i, \tag{10}$$

where P_i and R_i are the sets of observed and unobserved parents of v_i, respectively. The functions f_i and g_i are non-linear functions. Equation (10) is equivalent to Eq. (5) of Ashman et al. (2023).

Beyond the model in Eq. (10), visibility/invisibility with respect to X cannot be extended further to more general ANMs, as shown in the following example. Consider the model: $x_2 \to x_3$, $x_1 \to x_3$, $U \to x_3$, $U \to x_1$. x_2 and U are external. The equation for x_3 is $x_3 = \sin(x_1 + x_2 + U) + n_3$. This is a more general ANM than the model in Eq. (10), since the unobserved parent U is not separable from the observed parents x_1 and x_2 in the equation of x_3.

In this example, there is no UBP/UCP between x_2 and x_3, and x_2 is a parent of x_3. However, there is no regression function G of two variables x_1 and x_2 such that $x_3 - G(x_1, x_2)$ can be independent of x_2, due to the unobserved factor U inside the non-separable function $sin(x_1 + x_2 + U)$. This means the lemma about visible parents does not hold.

Lemmas 1, 2, and 3 in this paper, however, cannot be extended to the model in Eq. (10). These lemmas introduce the new concept of visibility/invisibility with respect to a subset $X' \subset X$, and thus require that the model class must be closed under reclassifying an observed variable as unobserved. CAM-UV satisfies this closure property, while the model in Eq. (10) of Ashman et al. (2023) does not.

Therefore, all results in this paper that rely on invisibility with respect to a proper subset $X' \subset X$ can only hold in CAM-UV, not in the model of Eq. (10).

On the other hand, since Corollary 6.1 only requires invisibility with respect to X, it holds in the larger model of Eq. (10).

C EXAMPLES WHERE THE PARENT–CHILD RELATIONSHIP IN AN INVISIBLE PAIR IS IDENTIFIED BY THE REGRESSION APPROACH

Fig. C.1a: There are two visible pairs in this example: (x_2, x_3) is a visible non-edge and $x_1 \to x_2$ is a visible edge. Applying Lemma 4 to the visible edge $x_1 \to x_2$ with $K = \{x_3\}$, we get that x_3 is a parent of either x_1 or x_2. However, since (x_2, x_3) is a visible non-edge, x_3 cannot be a parent of x_2. Therefore, x_3 can be identified as a parent of x_1 by the regression approach.

Fig. C.1b: This is an example where both (x_{k_q}, x_i) and (x_{k_q}, x_j) are invisible, and are bows. There are two visible pairs in this example: $x_4 \to x_3$ is a visible edge, and (x_1, x_2) is a visible non-edge. Applying Lemma 5 to (x_1, x_2) with $K = \{x_3\}$, we get that x_3 is a parent of either x_1 or x_2. Note that both (x_1, x_3) and (x_2, x_3) are invisible, and are also bows. Applying Lemma 4 to $x_4 \to x_3$ with $K = \{x_1\}$, we get that x_1 is an ancestor of x_3. Therefore, x_3 must be a parent of x_2. Thus, the parent–child relationship in the bow (x_3, x_2) is fully resolved.

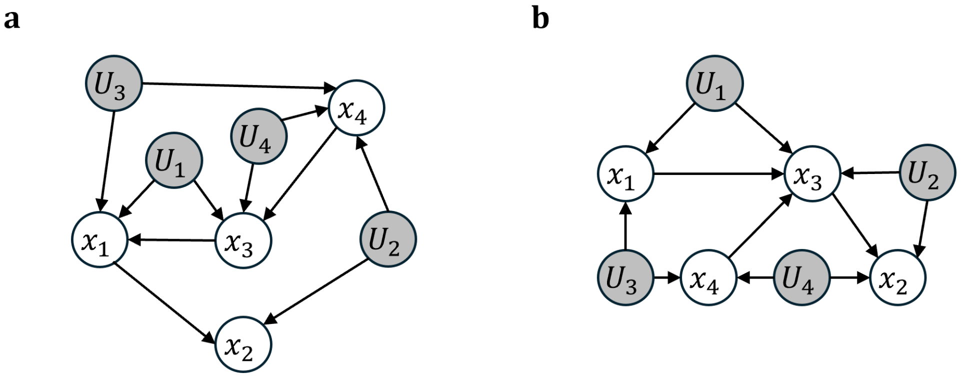

Figure C.1: Examples where the regression approach alone is sufficient to resolve parent–child relationships in an invisible pair.

D AN EXAMPLE WHERE COROLLARY 6.1 CAN IDENTIFY THE CAUSAL DIRECTION IN AN INVISIBLE PAIR

Consider Fig. D.2. The pair (x_2, x_3) is invisible, due to the UCP $x_2 \to U_1 \to x_3$.

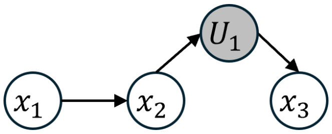

Figure D.2: Corollary 6.1 can identify the ancestor relationship, i.e., causal direction, in the invisible pair (x_2, x_3), whereas neither the regression approach alone nor conditional independence alone, e.g., the FCI framework, can.

Using Corollary 6.1. $x_1 \to x_2$ is a visible edge, and is identifiable by Lemma 1. Thus, x_1 can be identified as a parent of x_2, and therefore an ancestor of x_2. (x_2, x_3) can be identified as an invisible pair by Lemma 3. Furthermore, $x_1 \perp\!\!\!\perp x_3 \mid x_2$ if we assume the causal Markov condition. Therefore, the condition of Corollary 6.1 is satisfied, and thus x_2 can be identified as an ancestor of x_3 by Corollary 6.1.

Using the regression approach alone. $x_1 \to x_2$ can be identified as a visible edge. (x_2, x_3) is identified as invisible, due to Lemma 3. The pair (x_1, x_3) is invisible, due to the UCP $x_1 \to x_2 \to U_1 \to x_3$. Thus, (x_1, x_3) is identified as invisible by Lemma 3. Lemmas 4 and 5 cannot be applied in this example. Therefore, the causal direction in (x_2, x_3) is unidentifiable by the regression approach alone.

Using conditional independence alone. With the causal Markov condition and CFC, step 1 of FCI outputs the skeleton $x_1 \circ - \circ x_2 \circ - \circ x_3$, where there is one unshielded triple $x_1 \circ - \circ x_2 \circ - \circ x_3$, with $x_1 \perp\!\!\!\perp x_3 \mid \{x_2\}$. Since the middle node x_2 of this unshielded triple belongs to the d-separation set $\{x_2\}$, one cannot identify more causal directions. Therefore, the final output of FCI is $x_1 \circ - \circ x_2 \circ - \circ x_3$, where the causal direction between x_2 and x_3 is unidentifiable.

E RELATED WORKS

CAMs (Bühlmann et al., 2014) belong to a subclass of the Additive Noise Models (ANMs) (Hoyer et al., 2008a), which are causal models that assume nonlinear causal functions with additive noise terms, but the causal effects can be non-additive. Another important causal model is the Linear Non-Gaussian Acyclic Model (LiNGAM) (Shimizu et al., 2006), which assumes linear causal relationships with non-Gaussian noise.

A key extension of these causal models involves cases with hidden common causes (Hoyer et al., 2008b; Zhang et al., 2010; Tashiro et al., 2014; Salehkaleybar et al., 2020). To address such scenarios, methods like the Repetitive Causal Discovery (RCD) algorithm (Maeda and Shimizu, 2020) and the CAM-UV (Causal Additive Models with Unobserved Variables) algorithm (Maeda and Shimizu, 2021) have been developed. CAM-UV has been applied to integrate causal graphs estimated from non-overlapping variable sets (Suzuki et al., 2026).

Estimation approaches for causal discovery can be generally categorized into three groups: constraint-based methods (Spirtes and Glymour, 1991; Spirtes et al., 1995), score-based methods (Chickering, 2002), and continuous-optimization-based methods (Zheng et al., 2018; Bhattacharya et al., 2021). Additionally, a hybrid approach that combines the ideas of constraint-based and score-based methods has been proposed for non-parametric cases (Ogarrio et al., 2016).

F RUNNING TIME OF CAM-UV-X

Assume that each regression, each calculation of p-HSIC, and each conditional independence check costs $O(n^2)$, where n is the number of samples. After executing CAM-UV, the remaining running time of CAM-UV-X is bounded by the time of executing lines 4-6, where `checkVisible` is executed for every NaN entry of the adjacency matrix, and by the time of executing lines 8-9, where `checkOnPath` is executed for every non-NaN entry of the adjacency matrix. Let the two times be t_1 and t_2, respectively.

- For t_1: The number of NaN entries in the adjacency matrix is $O(p^2)$. In each execution of `checkVisible`, in the worst case, one needs to check all sets $M \subseteq Q$ and $N \subseteq Q$, where the size of Q is at most $p-2$. Therefore, t_1 is $O(p^2 \times 2^{p-2} \times 2^{p-2} \times n^2)$.
- For t_2: The number of non-NaN entries of the adjacency matrix is $O(p^2)$. For each execution of `checkOnPath`, a third variable x_k is considered. For each x_k, in the worst case, the procedure checks all sets $M \subseteq X \setminus \{x_i, x_k\}$ and $N \subseteq X \setminus \{x_j, x_k\}$. The size of $X \setminus \{x_i, x_k\}$ is $p-2$. Therefore, t_2 is $O(p^2 \times p \times 2^{p-2} \times 2^{p-2} \times n^2)$.

Given that the running time of CAM-UV is $O(p2^pn^2)$ (Maeda and Shimizu, 2021), the running time of CAM-UV-X is $O(p2^pn^2 + p^2 2^{2p-4}n^2 + p^3 2^{2p-4}n^2) = O(p^3 2^{2p-4}n^2)$.

G ADDITIONAL THEORETICAL RESULTS

G.1 Some additional identifiability results

We provide another sufficient condition for certifying that some variable x_k must be a parent of x_i or x_j, even if (x_k, x_i) and (x_k, x_j) are both invisible, by utilizing visible non-edges. First, we have the following sufficient condition for non-parent relations and non-edges.

Lemma 8. *If the following equation holds:*

$$\begin{aligned}&\exists G_1, G_2 \in \mathcal{G}, M \subseteq X \setminus \{x_i, x_j\}, N \subseteq X \setminus \{x_i, x_j\} :\\ &x_i - G_1(M \cup \{x_j\}) \perp\!\!\!\perp x_j - G_2(N),\end{aligned} \tag{11}$$

then (x_i, x_j) is not invisible, and x_i is not a parent of x_j.

Similarly, if the following equation holds:

$$\begin{aligned}&\exists G_1, G_2 \in \mathcal{G}, M \subseteq X \setminus \{x_i, x_j\}, N \subseteq X \setminus \{x_i, x_j\} :\\ &x_i - G_1(M) \perp\!\!\!\perp x_j - G_2(N \cup \{x_i\}),\end{aligned} \tag{12}$$

then (x_i, x_j) is not invisible, and x_j is not a parent of x_i. If Eqs. (11) and (12) are satisfied, (x_i, x_j) is a visible non-edge.

Eq. (11) means that, if the residual of regressing x_i on a set containing x_j is independent of the residual of regressing x_j on a set that does not contain x_i, the pair is not invisible, and x_i is not a parent of x_j. When Eqs. (11) and (12) are satisfied, both directions of parent relations are forbidden; thus (x_i, x_j) is a visible non-edge. Instead of Eq. (4), Eqs. (11) and (12) provide an alternative, and sometimes more convenient, way to certify visible non-edges.

Similar to Lemma 5, we can use the above condition to identify causal relationships in potentially invisible pairs as follows.

Lemma 9. *Consider distinct x_i, x_j, and $x_{k_1}, \dots, x_{k_m}$. Let $K = \{x_{k_1}, \dots, x_{k_m}\}$. If Eqs. (11), (12), and (6) are satisfied, (x_j, x_i) is a visible non-edge, and each x_{k_q} is either a parent of x_i or a parent of x_j.*

G.2 Proof of Proposition 1

If (x_i, x_j) is an invisible pair with respect to $X \setminus \{x_k\}$, there exists a UBP/UCP between (x_i, x_j) with respect to $X \setminus \{x_k\}$. Call this UBP/UCP P. If (x_i, x_j) is visible with respect to X, there are no UBPs/UCPs between (x_i, x_j) with respect to X. This implies that P is a UBP/UCP between (x_i, x_j) with respect to $X \setminus \{x_k\}$, and is not a UBP/UCP between (x_i, x_j) with respect to X. Thus, x_k must be a parent of x_j or a parent of x_i.

G.3 Proof of Lemma 4

($\Leftarrow$): If Eq. (6) is satisfied, (x_i, x_j) is an invisible pair with respect to $X \setminus \{x_{k_q}\}$ due to Lemma 3. When Eqs. (7) and (8) are satisfied, x_j is a visible parent of x_i with respect to X, due to Lemma 1. Thus, due to Proposition 1, each x_{k_q} must be a parent of x_j or a parent of x_i.

($\Rightarrow$): Since (x_i, x_j) is invisible with respect to $X \setminus \{x_{k_q}\}$, Eq. (6) holds, due to Lemma 3. Since x_j is a visible parent of x_i, Eqs. (7) and

$$\exists G_1, G_2 \in \mathcal{G}, M \subseteq X \setminus \{x_i, x_j\}, N \subseteq X \setminus \{x_i, x_j\} : x_i - G_1(M \cup \{x_j\}) \perp\!\!\!\perp x_j - G_2(N) \tag{13}$$

are satisfied, due to Proposition 2. The remaining task is to show that Eq. (8) holds. To do so, one way is to show that, for any $x_{k_q} \in K$, either $x_{k_q} \in M$ or $x_{k_q} \in N$ with M, N satisfying Eq. (13). Suppose that this is false, i.e., there exists some x_{k_q} such that $x_{k_q} \notin M$ and $x_{k_q} \notin N$, for M, N satisfying Eq. (13). Then, Eq. (13) implies that:

$$\exists G_1, G_2 \in \mathcal{G}, M \subseteq X \setminus \{x_i, x_{k_q}\}, N \subseteq X \setminus \{x_i, x_j, x_{k_q}\} : x_i - G_1(M) \perp\!\!\!\perp x_j - G_2(N), \tag{14}$$

which contradicts Eq. (6). Therefore, any $x_{k_q} \in K$ satisfies $x_{k_q} \in M$ or $x_{k_q} \in N$.

G.4 Proof of Lemma 5

($\Leftarrow$): If Eq. (6) is satisfied, (x_i, x_j) is invisible with respect to $X \setminus \{x_{k_q}\}$, due to Lemma 3. When Eq. (9) is satisfied, (x_i, x_j) is a visible non-edge, due to Lemma 2. Due to Proposition 1, each x_{k_q} must be a parent of x_j or a parent of x_i.

($\Rightarrow$): Since (x_i, x_j) is invisible with respect to $X \setminus \{x_{k_q}\}$, Eq. (6) holds, due to Lemma 3. Since (x_i, x_j) is a visible non-edge, we have

$$\exists G_1, G_2 \in \mathcal{G}, M \subseteq X \setminus \{x_i, x_j\}, N \subseteq X \setminus \{x_i, x_j\} : x_i - G_1(M) \perp\!\!\!\perp x_j - G_2(N), \tag{15}$$

due to Lemma 2. To show Eq. (9), we will show that for every $x_{k_q} \in K$, either $x_{k_q} \in M$ or $x_{k_q} \in N$, with M, N satisfying Eq. (15). Suppose that this is false, i.e., for any M, N that satisfies Eq. (15), there exists some x_{k_q} such that $x_{k_q} \notin M$ and $x_{k_q} \notin N$. Eq. (15) implies that:

$$\exists G_1, G_2 \in \mathcal{G}, M \subseteq X \setminus \{x_i, x_j, x_{k_q}\}, N \subseteq X \setminus \{x_i, x_j, x_{k_q}\} : x_i - G_1(M) \perp\!\!\!\perp x_j - G_2(N), \tag{16}$$

which contradicts Eq. (6). Therefore, every $x_{k_q} \in K$ is in M or in N.

G.5 Proof of Lemma 8

Assume Eq. (11). We will prove that (x_i, x_j) is not invisible, and x_i is not a parent of x_j, by contradiction.

If (x_i, x_j) is invisible, Eq. (5) is satisfied, which contradicts Eq. (11). Therefore, (x_i, x_j) is not invisible.

Suppose that x_i is a parent of x_j. This means that x_i is a visible parent of x_j. Eq. (2) is satisfied, due to Lemma 1. However, this contradicts Eq. (11). Therefore, x_i is not a parent of x_j.

When both Eqs. (11) and (12) are satisfied, (x_i, x_j) is not invisible, thus one of the following three cases must happen: 1) x_i is a visible parent of x_j, 2) x_j is a visible parent of x_i, or 3) (x_i, x_j) is a visible non-edge. However, Eq. (11) implies x_i is not a parent of x_j, and Eq. (12) implies that x_j is not a parent of x_i. Therefore, (x_i, x_j) must be a visible non-edge.

G.6 Proof of Lemma 9

Eq. (6) implies that (x_i, x_j) is invisible with respect to $X \setminus \{x_{k_q}\}$, due to Lemma 3. Eqs. (11) and (12) imply that (x_i, x_j) is a visible non-edge with respect to X, by Lemma 8. Due to Proposition 1, each x_{k_q} must be a parent of x_i or x_j.

G.7 Proof of Lemma 6

As mentioned in the main text, the lemma is standard and not new. For completeness, we record its proof here.

We prove by contradiction.

1. Suppose there is a backdoor path $x_i \leftarrow \cdots \leftarrow v \rightarrow \cdots \rightarrow x_j$. Since x_k is an ancestor of x_i, the path $x_k \rightarrow \cdots \rightarrow x_i \leftarrow \cdots \leftarrow v \rightarrow \cdots \rightarrow x_j$ exists. By conditioning on x_i, which is a collider on this path, the path is open and thus x_k and x_j cannot be independent due to CFC. This contradicts the assumption $x_k \perp\!\!\!\perp x_j \mid x_i$.
2. Suppose x_j is an ancestor of x_i. Since x_k is an ancestor of x_i, the path $x_k \rightarrow \cdots \rightarrow x_i \leftarrow \cdots \leftarrow x_j$ exists. By conditioning on x_i, which is a collider on this path, the path is open and thus x_k and x_j cannot be independent due to CFC. This contradicts the assumption $x_k \perp\!\!\!\perp x_j \mid x_i$.

G.8 Proof of Corollary 6.1

Eq. (5) implies that a UBP/UCP must exist between x_i and x_j, due to Lemma 3. Lemma 6 rules out the possibilities of any UBP and the causal direction from x_j to x_i. Therefore, the only possible scenario is that there is a UCP from x_i to x_j, which means x_i is an ancestor of x_j.

G.9 Proof of Corollary 6.2

Lemma 6 implies that x_k is not an ancestor, and thus not a parent, of x_i. Therefore, the only possibility is that x_k is a parent of x_j.

G.10 Proof of Theorem 7

We state some facts that can limit the possible contents of the regression sets M and N in Lemmas 1 and 2.

Lemma 1 is equivalent to the following proposition.

Proposition 2. *Consider $X' \subseteq X$ and $x_i, x_j \in X'$. x_j is a visible parent of x_i with respect to X' if and only if Eq. (2) and*

$$\exists G_1, G_2 \in \mathcal{G}, M \subseteq X' \setminus \{x_i, x_j\}, N \subseteq X' \setminus \{x_i, x_j\} : x_i - G_1(M \cup \{x_j\}) \perp\!\!\!\perp x_j - G_2(N) \tag{17}$$

are satisfied. Eq. (17) means that, in Eq. (3), one must regress x_i on a set that contains x_j.

Proof. If Eqs. (2) and (17) are satisfied, Eqs. (2) and (3) are satisfied. Thus, x_j is a visible parent of x_i with respect to X' by Lemma 1.

Suppose that x_j is a visible parent of x_i with respect to X'. Eqs. (2) and (3) are satisfied by Lemma 1. From Eq. (3), we have:

$$\exists G'_1, G'_2 \in \mathcal{G}, M' \subseteq X' \setminus \{x_i\}, N' \subseteq X' \setminus \{x_i, x_j\} : x_i - G'_1(M') \perp\!\!\!\perp x_j - G'_2(N'). \tag{18}$$

For any $M' \subseteq X' \setminus \{x_i\}$ satisfying the above equation, we prove that x_j must belong to M' by contradiction. Suppose that $x_j \notin M'$. Then $M' \subseteq X' \setminus \{x_i, x_j\}$, which means that

$$\exists G'_1, G'_2 \in \mathcal{G}, M' \subseteq X' \setminus \{x_i, x_j\}, N' \subseteq X' \setminus \{x_i, x_j\} : x_i - G'_1(M') \perp\!\!\!\perp x_j - G'_2(N'). \tag{19}$$

By Lemma 2, this implies that (x_i, x_j) is a visible non-edge with respect to X'. However, this contradicts the fact that x_j is a parent of x_i. Therefore, $x_j \in M'$, which means Eq. (17) is true. $\square$

We have the following proposition to limit the content of the regression set M in Eq. (17).

Proposition 3. *Assume that x_k is not a parent of x_i and not a parent of x_j. The following equation*

$$\exists G_1, G_2 \in \mathcal{G}, M \subseteq X \setminus \{x_i, x_k\}, N \subseteq X \setminus \{x_i, x_j\} : x_i - G_1(M \cup \{x_k\}) \perp\!\!\!\perp x_j - G_2(N) \tag{20}$$

implies that

$$\exists G_1, G_2 \in \mathcal{G}, M \subseteq X \setminus \{x_i, x_k\}, N \subseteq X \setminus \{x_i, x_j\} : x_i - G_1(M) \perp\!\!\!\perp x_j - G_2(N). \tag{21}$$

In other words, one can limit the set M to the union of the set of parents of x_i and the set of parents of x_j when checking Eq. (17).

Proof. Eq. (20) implies that (x_i, x_j) is visible with respect to X. Since x_k is not a parent of x_i or a parent of x_j, if there are no UBPs/UCPs between (x_i, x_j) with respect to X, then there are also no UBPs/UCPs between (x_i, x_j) with respect to $X \setminus \{x_k\}$. Therefore, (x_i, x_j) is also visible with respect to $X \setminus \{x_k\}$, and we have

$$\exists G_1, G_2 \in \mathcal{G}, M \subseteq X \setminus \{x_i, x_k\}, N \subseteq X \setminus \{x_i, x_j, x_k\} : x_i - G_1(M) \perp\!\!\!\perp x_j - G_2(N), \tag{22}$$

which implies Eq. (21). ☐

Similarly, we have the following proposition.

Proposition 4. *Assume that x_k is not a parent of x_i and not a parent of x_j. The following equation*

$$\exists G_1, G_2 \in \mathcal{G}, M \subseteq X \setminus \{x_i\}, N \subseteq X \setminus \{x_i, x_j, x_k\} : x_i - G_1(M) \perp\!\!\!\perp x_j - G_2(N \cup \{x_k\}) \tag{23}$$

implies that

$$\exists G_1, G_2 \in \mathcal{G}, M \subseteq X \setminus \{x_i\}, N \subseteq X \setminus \{x_i, x_j, x_k\} : x_i - G_1(M) \perp\!\!\!\perp x_j - G_2(N). \tag{24}$$

In other words, one can also limit the set N to the union of the set of parents of x_i and the set of parents of x_j when checking Eq. (17).

Proof. Same as the proof of Proposition 3. ☐

We have the following Propositions 5 and 6 to limit the contents of the regression sets M and N when checking Eq. (4) in Lemma 2.

Proposition 5. *Assume that x_k is not a parent of x_i and not a parent of x_j. The following equation*

$$\exists G_1, G_2 \in \mathcal{G}, M \subseteq X \setminus \{x_i, x_j, x_k\}, N \subseteq X \setminus \{x_i, x_j\} : x_i - G_1(M \cup \{x_k\}) \perp\!\!\!\perp x_j - G_2(N) \tag{25}$$

implies that

$$\exists G_1, G_2 \in \mathcal{G}, M \subseteq X \setminus \{x_i, x_j, x_k\}, N \subseteq X \setminus \{x_i, x_j\} : x_i - G_1(M) \perp\!\!\!\perp x_j - G_2(N). \tag{26}$$

Proof. Eq. (25) implies that (x_i, x_j) is a visible non-edge with respect to X. Since x_k is not a parent of x_i or a parent of x_j, if there are no UBPs/UCPs between (x_i, x_j) with respect to X, then there are also no UBPs/UCPs between (x_i, x_j) with respect to $X \setminus \{x_k\}$. Therefore, (x_i, x_j) is also a visible non-edge with respect to $X \setminus \{x_k\}$, and we have

$$\exists G_1, G_2 \in \mathcal{G}, M \subseteq X \setminus \{x_i, x_j, x_k\}, N \subseteq X \setminus \{x_i, x_j, x_k\} : x_i - G_1(M) \perp\!\!\!\perp x_j - G_2(N), \tag{27}$$

which implies Eq. (26). ☐

Proposition 6. *Assume that x_k is not a parent of x_i and not a parent of x_j. The following equation*

$$\exists G_1, G_2 \in \mathcal{G}, M \subseteq X \setminus \{x_i, x_j\}, N \subseteq X \setminus \{x_i, x_j, x_k\} : x_i - G_1(M) \perp\!\!\!\perp x_j - G_2(N \cup \{x_k\}) \tag{28}$$

implies that

$$\exists G_1, G_2 \in \mathcal{G}, M \subseteq X \setminus \{x_i, x_j\}, N \subseteq X \setminus \{x_i, x_j, x_k\} : x_i - G_1(M) \perp\!\!\!\perp x_j - G_2(N). \tag{29}$$

Proof. Same as the proof of Proposition 5. ☐

Now we are ready to prove the soundness and completeness of CAM-UV-X as follows.

Soundness in identifying visible non-edges

A pair (x_i, x_j) is identified as a visible non-edge in CAM-UV-X if 1) it is identified as a visible non-edge by CAM-UV, or 2) it is identified as invisible by CAM-UV, and is re-identified as a visible non-edge by `checkVisible`.

- For the first case, since CAM-UV is sound in identifying visible pairs with Assumption 3, (x_i, x_j) is also a visible non-edge in the ground truth.
- For the second case, this means that either 1) $e > \alpha$ in line 6, or 2) $iNotParent = True$ and $jNotParent = True$ in line 17 of `checkVisible`.
  - If $e > \alpha$, Eq. (4) with $X' = X$ is satisfied due to Assumption 3. This implies (x_i, x_j) is a visible non-edge in the ground truth, due to Lemma 2.
  - If $iNotParent = True$ and $jNotParent = True$, Eqs. (11) and (12) are satisfied, due to Assumption 3. This implies that (x_i, x_j) is a visible non-edge in the ground truth, due to Lemma 8.

Therefore, CAM-UV-X is sound in identifying visible non-edges.

Completeness in identifying visible non-edges.

Suppose that (x_i, x_j) is a visible non-edge in the ground truth. Due to Lemma 2 , Eq. (4) with $X' = X$ holds in the data. After the execution of CAM-UV, there are three cases:

- (x_i, x_j) is concluded as a visible non-edge: CAM-UV-X leaves the pair as is. Therefore, (x_i, x_j) is also a visible non-edge in the output of CAM-UV-X.
- (x_i, x_j) is concluded as a visible edge: this case does not happen, since CAM-UV is sound in identifying visible pairs with Assumption 3.
- (x_i, x_j) is concluded as an invisible pair: `checkVisible` is executed for (i, j). In lines 5 and 6 of `checkVisible`, independence is checked for all subsets of Q, which is the union of P_i, P_j, and the set of nodes whose parent–child relationship to either x_i or x_j is not clear. Since CAM-UV is sound for identifying visible non-edges with Assumption 3, this set does not miss any parents of x_i or x_j. Since Eq. (4) holds with $X' = X$, there exist some sets $M \subseteq X \setminus \{x_i, x_j\}$ and $N \subseteq X \setminus \{x_i, x_j\}$ that realize the independence between residuals, and produce a p-HSIC value that is greater than α, due to Assumption 3.
  - If M and N do not contain non-parents of x_i and x_j, this means $M, N \subseteq Q$, and exhaustively searching through Q as in `checkVisible` ensures the finding of M and N.
  - If M or N contain some non-parent of x_i and x_j, Propositions 5 and 6 imply that there exist some sets M' and N' in Q such that the independence between residuals is realized, and produce a p-HSIC value that is greater than α, by Assumption 3. Exhaustively searching through Q as in `checkVisible` ensures the finding of M' and N'.

  Therefore, `checkVisible` is guaranteed to find some set $M, N \subseteq Q$ such that $e > \alpha$ and line 6 is satisfied. Thus, CAM-UV-X outputs (x_i, x_j) as a visible non-edge.

Therefore, CAM-UV-X is complete in identifying visible non-edges.

Soundness in identifying visible edges

An edge $x_j \to x_i$ is identified as visible in CAM-UV-X if 1) it is identified as visible by CAM-UV, or 2) it is identified as invisible by CAM-UV, and is re-identified as visible by `checkVisible`.

- For the first case, since CAM-UV is sound in identifying visible pairs with Assumption 3, $x_j \to x_i$ is also a visible edge in the ground truth.
- For the second case, this means that $e \leq \alpha$ in line 6 for all sets $M, N \subseteq Q$, $a_1 > \alpha$ (since $iNotParent$ must be $True$) for some set $M, N \subseteq Q$, and $a_2 \leq \alpha$ (since $jNotParent$ must remain $False$) for all sets $M, N \subseteq Q$ in `checkVisible`.

  - + Since $a_1 > \alpha$ for some set $M, N \subseteq Q$, Eq. (17) with $X' = Q \cup \{x_i, x_j\}$ is satisfied, due to Assumption 3.
  - + Since $e \leq \alpha$ and $a_2 \leq \alpha$ for all sets checked, this means that for all $M \subseteq Q$ and $N \subseteq Q \cup \{x_i\}$, independence is not established for all functions in $\mathcal{G}$, due to Assumption 3. Q is the union of P_i, P_j, and the set of nodes whose parent–child relationship to either x_i or x_j is not clear. Since CAM-UV is sound for identifying visible non-edges with Assumption 3, this set does not miss any parents of x_i or x_j. This implies that Eq. (2) with $X' = Q \cup \{x_i, x_j\}$ is satisfied.

  Since Eqs. (17) and (2) are satisfied with $X' = Q \cup \{x_i, x_j\}$, x_j is a visible parent of x_i with respect to $Q \cup \{x_i, x_j\}$ in the ground truth, due to Proposition 2. This implies that x_j is also a visible parent of x_i with respect to X in the ground truth.

Therefore, CAM-UV-X is sound in identifying visible edges.

Completeness in identifying visible edges

Suppose that $x_j \rightarrow x_i$ is a visible edge in the ground truth. Due to Proposition 2, Eqs. (2) and (17) with $X' = X$ hold in the data. After the execution of CAM-UV, there are three cases:

- $x_j \rightarrow x_i$ is concluded as a visible edge: CAM-UV-X leaves the edge as is. Therefore, $x_j \rightarrow x_i$ is also a visible edge in the output of CAM-UV-X.
- $x_j \rightarrow x_i$ is concluded as a visible non-edge: this case does not happen, since CAM-UV is sound in identifying visible pairs with Assumption 3.
- $x_j \rightarrow x_i$ is concluded as an invisible pair: `checkVisible` is executed for (i, j). Due to Eq. (2) with $X' = X$ and Assumption 3, $e \leq \alpha$ and $a_2 \leq \alpha$ for all sets $M, N \subseteq Q$. Thus, $jNotParent$ remains $False$, and lines 7 and 18 of `checkVisible` are guaranteed to be not executed. Q is the union of P_i, P_j, and the set of nodes whose parent–child relationship to either x_i or x_j is not clear. Since CAM-UV is sound for identifying visible non-edges with Assumption 3, this set does not miss any parents of x_i or x_j. Since Eq. (17) with $X' = X$ holds, there exist some sets M and N that realize the independence between residuals.
  - + If M and N do not contain non-parents of x_i and x_j, this means $M, N \subseteq Q$, and searching through Q ensures the finding of M and N.
  - + If M or N contain some non-parent of x_i and x_j, Propositions 3 and 4 imply that there exist some set M' and N' in Q such that the independence between residuals is realized. Searching through Q ensures the finding of M' and N'.

  Therefore, CAM-UV-X is guaranteed to find some sets such that regression in those sets will produce $a_1 > \alpha$ in line 11 of `checkVisible`, due to Assumption 3. Therefore, $iNotParent$ will be changed to $True$. Coupling this with the fact that $jNotParent$ remains $False$, and the fact that lines 7 and 18 of `checkVisible` are not executed, one can conclude that line 23 of `checkVisible` is guaranteed to be executed. Thus, CAM-UV-X outputs $x_j \rightarrow x_i$ as a visible edge.

Therefore, CAM-UV-X is complete in identifying visible edges.

Soundness in identifying invisible pairs.

When CAM-UV-X identifies a pair (x_i, x_j) as invisible, the pair cannot be a visible edge in the ground truth, since this would contradict the proven completeness of CAM-UV-X in identifying visible edges. It also cannot be a visible non-edge in the ground truth, since this would contradict the proven completeness of CAM-UV-X in identifying visible non-edges. Therefore, the pair must be invisible in the ground truth. This means CAM-UV-X is sound in identifying invisible pairs.

Completeness in identifying invisible pairs.

Suppose (x_i, x_j) is an invisible pair in the ground truth. In the output of CAM-UV-X, the pair cannot be a visible edge, since this would contradict the proven soundness of CAM-UV-X in identifying visible edges. It also cannot be a visible non-edge in the output, since this would contradict the proven soundness of CAM-UV-X in identifying visible non-edges. Therefore, the pair must be invisible in the output of CAM-UV-X. This means CAM-UV-X is complete in identifying invisible pairs.

H STEP-BY-STEP EXECUTION OF CAM-UV ON FIGS. 1A AND 1B

We work out step-by-step the execution of the CAM-UV algorithm for the graphs in Figs. 1a and 1b. We assume Assumption 3.

H.1 Fig. 1a

- Algorithm 1:
  - Phase 1:
    - $t = 2$: The sets $\{x_1, x_2\}$, $\{x_1, x_3\}$, and $\{x_2, x_3\}$ are considered as K. We search for the candidate sink x_b in K.
      - $K = \{x_1, x_3\}$ or $K = \{x_2, x_3\}$: Since these pairs are invisible, $x_b - G_1(M_b \cup K \setminus \{x_b\})$ and $x_j - G_2(M_j)$ for $j \in K \setminus \{x_b\}$ are not independent, regardless of x_b. Thus, line 15 will fail since $e \leq \alpha$. There is no change in M_i.
      - $K = \{x_1, x_2\}$: CAM-UV correctly finds $x_b = x_2$. However, $x_b - G_1(M_b \cup K \setminus \{x_b\}) = x_2 - G_1(x_1)$ and $x_j - G_2(M_j) = x_1$ is not independent, due to the unblocked backdoor path $x_1 \leftarrow U_1 \rightarrow x_3 \rightarrow x_2$. Thus, line 15 will fail since $e \leq \alpha$. There is no change in M_i.

      M_i remains empty for each i and t increases to 3.
    - $t = 3$: $K = \{x_1, x_2, x_3\}$. The algorithm correctly chooses $x_b = x_2$. However, $x_b - G_1(M_b \cup K \setminus \{x_b\}) = x_2 - G(x_1, x_3)$ and $x_j - G_2(M_j) = x_3$ is not independent, due to the unobserved backdoor path $x_3 \leftarrow U_2 \rightarrow x_2$. Therefore, line 15 will fail again, since $e \leq \alpha$.

    M_i remains empty for each i. Phase 1 of the Algorithm 1 ends.
  - Phase 2: Since M_i is empty for each i, Phase 2 ends.

  Algorithm 1 ends with every M_i being empty.
- Algorithm 2: For each pair (i, j), line 5 is satisfied. Therefore, the algorithm concludes that every pair is invisible. CAM-UV ends.

The final output is an adjacency matrix where each off-diagonal element is NaN.

H.2 Fig. 1b

- Algorithm 1:
  - Phase 1:
    - $t = 2$: The sets $\{x_1, x_2\}$, $\{x_1, x_3\}$, $\{x_1, x_4\}$, $\{x_2, x_3\}$, $\{x_2, x_4\}$, and $\{x_3, x_4\}$ are considered as K. The candidate sink x_b of K is searched.
      - $K = \{x_1, x_2\}$: Regardless of which x_b is, $x_b - G_1(M_b \cup K \setminus \{x_b\}) = x_b - G_1(\{x_1, x_2\} \setminus \{x_b\})$ and $x_j - G_2(M_j) = x_j$ for $j \in K \setminus \{x_b\}$ are not independent, since there are unblocked BPs/CPs when x_3 and x_4 are not added to the regression. Thus, line 15 will fail since $e \leq \alpha$. There is no change in M_i.
      - The remaining pairs are all invisible. Therefore, $x_b - G_1(M_b \cup K \setminus \{x_b\}) = x_b - G_1(K \setminus \{x_b\})$ and $x_j - G_2(M_j)$ is not independent. Thus, line 15 will fail since $e \leq \alpha$. There is no change in M_i.

      M_i remains empty for each i and t increases to 3.
    - $t = 3$:
      - $K = \{x_1, x_2, x_3\}$:
        - $x_b = x_2$: $x_b - G_1(M_b \cup K \setminus \{x_b\}) = x_2 - G(x_1, x_3)$ and $x_j - G_2(M_j) = x_3$ is not independent, due to the unobserved backdoor path $x_3 \leftarrow U_4 \rightarrow x_2$. Therefore, line 15 will fail, since $e \leq \alpha$.
        - $x_b = x_1$: $x_b - G_1(M_b \cup K \setminus \{x_b\}) = x_1 - G(x_2, x_3)$ and $x_j - G_2(M_j) = x_3$ is not independent, due to the unobserved backdoor path $x_1 \leftarrow U_1 \rightarrow x_3$. Therefore, line 15 will fail, since $e \leq \alpha$.
        - $x_b = x_3$: $x_b - G_1(M_b \cup K \setminus \{x_b\}) = x_3 - G(x_1, x_2)$ and $x_j - G_2(M_j) = x_2$ is not independent, due to the unobserved backdoor path $x_3 \leftarrow U_4 \rightarrow x_2$. Therefore, line 15 will fail, since $e \leq \alpha$.
      - $K = \{x_2, x_3, x_4\}$:
        - $x_b = x_2$: $x_b - G_1(M_b \cup K \setminus \{x_b\}) = x_2 - G(x_3, x_4)$ and $x_j - G_2(M_j) = x_3$ is not independent, due to the unobserved backdoor path $x_3 \leftarrow U_4 \rightarrow x_2$. Therefore, line 15 will fail, since $e \leq \alpha$.

- ○ $x_b = x_3$: $x_b - G_1(M_b \cup K \setminus \{x_b\}) = x_3 - G(x_2, x_4)$ and $x_j - G_2(M_j) = x_2$ is not independent, due to the unobserved backdoor path $x_3 \leftarrow U_4 \rightarrow x_2$. Therefore, line 15 will fail, since $e \leq \alpha$.
- ○ $x_b = x_4$: $x_b - G_1(M_b \cup K \setminus \{x_b\}) = x_4 - G(x_2, x_3)$ and $x_j - G_2(M_j) = x_2$ is not independent, due to the unobserved backdoor path $x_2 \leftarrow U_3 \rightarrow x_4$. Therefore, line 15 will fail, since $e \leq \alpha$.

⋆ $K = \{x_1, x_3, x_4\}$:

- ○ $x_b = x_1$: $x_b - G_1(M_b \cup K \setminus \{x_b\}) = x_1 - G(x_3, x_4)$ and $x_j - G_2(M_j) = x_3$ is not independent, due to the unobserved backdoor path $x_3 \leftarrow U_1 \rightarrow x_1$. Therefore, line 15 will fail, since $e \leq \alpha$.
- ○ $x_b = x_3$: $x_b - G_1(M_b \cup K \setminus \{x_b\}) = x_3 - G(x_1, x_4)$ and $x_j - G_2(M_j) = x_1$ is not independent, due to the unobserved backdoor path $x_3 \leftarrow U_1 \rightarrow x_1$. Therefore, line 15 will fail, since $e \leq \alpha$.
- ○ $x_b = x_4$: $x_b - G_1(M_b \cup K \setminus \{x_b\}) = x_4 - G(x_1, x_3)$ and $x_j - G_2(M_j) = x_1$ is not independent, due to the unobserved causal path $x_1 \rightarrow U_2 \rightarrow x_4$. Therefore, line 15 will fail, since $e \leq \alpha$.

M_i remains empty for each i and t increases to 4.

–– $t = 4$:

⋆ $K = \{x_1, x_2, x_3, x_4\}$.

- ○ $x_b = x_1$: $x_b - G_1(M_b \cup K \setminus \{x_b\}) = x_1 - G(x_2, x_3, x_4)$ and $x_j - G_2(M_j) = x_3$ is not independent, due to the unobserved backdoor path $x_3 \leftarrow U_1 \rightarrow x_1$. Therefore, line 15 will fail, since $e \leq \alpha$.
- ○ $x_b = x_2$: $x_b - G_1(M_b \cup K \setminus \{x_b\}) = x_2 - G(x_1, x_3, x_4)$ and $x_j - G_2(M_j) = x_3$ is not independent, due to the unobserved backdoor path $x_3 \leftarrow U_4 \rightarrow x_2$. Therefore, line 15 will fail, since $e \leq \alpha$.
- ○ $x_b = x_3$: $x_b - G_1(M_b \cup K \setminus \{x_b\}) = x_3 - G(x_1, x_2, x_4)$ and $x_j - G_2(M_j) = x_1$ is not independent, due to the unobserved backdoor path $x_1 \leftarrow U_1 \rightarrow x_3$. Therefore, line 15 will fail, since $e \leq \alpha$.
- ○ $x_b = x_4$: $x_b - G_1(M_b \cup K \setminus \{x_b\}) = x_4 - G(x_1, x_2, x_3)$ and $x_j - G_2(M_j) = x_2$ is not independent, due to the unobserved backdoor path $x_4 \leftarrow U_3 \rightarrow x_2$. Therefore, line 15 will fail, since $e \leq \alpha$.

M_i remains empty for each i. Phase 1 of the Algorithm 1 ends.

+ Phase 2: Since M_i is empty for each i, Phase 2 ends.

Algorithm 1 ends with every M_i being empty.

- Algorithm 2: For each pair (i, j), line 5 is satisfied. Therefore, the algorithm concludes that every pair is invisible. CAM-UV ends.

The final output is an adjacency matrix where each off-diagonal element is NaN.

I DETAILS OF EXPERIMENTS

I.1 S-B method

As discussed in the main text, the theory of Schultheiss and Bühlmann (2024) is not designed for causal search. Below is a description of the S-B method, which is our attempt to create a baseline for identifying the parent–child relationships between observed variables using Schultheiss and Bühlmann (2024).

- Initialize the adjacency matrix A with all elements as NaN.
- For each $j = 1, \ldots, p$:
  - + For a given significance level α, apply Algorithm 3 of Schultheiss and Bühlmann (2024) with $\tilde{\alpha} = \alpha$ and number of splits $B = 25$ to obtain $\hat{W}_j \subseteq \{1, \ldots, p\}$, the set of indices i of variables X_i whose causal effects on X_j are well-specified.
  - + For each variable X_i with $i \in \hat{W}_j$, we estimate the causal effect of X_i on X_j while all remaining variables in Eq. (1), including the hidden ones, are fixed. If this causal effect is 0, then X_i is not a parent of X_j. Otherwise, X_i is a parent of X_j. The procedure is as follows.
    - –– We learn the conditional expectation function $\mathbb{E}[X_j \mid X_{-j}]$ by learning a regression model $f(X_{-j})$ to predict X_j using X_{-j}. Here, X_{-j} are the observed variables excluding X_j. In the experiments, we use XGBoost (Chen and Guestrin, 2016) as the regression model.
    - –– We sample a location m from $1, \ldots, n$, with n being the number of observed samples, independently M times. We use $M = 100$ in the experiments. For each location m, we calculate a value δ_m as follows.

- ⋆ For each location m, we randomly sample a location a ($a \neq m$) from $1, \ldots, n$.
  - ○ We calculate

$$\delta_m = \hat{f}\left(X_i = (x_i)_a, X_{k|k\neq i,j} = (x_{k|k\neq i,j})_m\right) - \hat{f}\left(X_i = (x_i)_m, X_{k|k\neq i,j} = (x_{k|k\neq i,j})_m\right).$$

    $(x_i)_a$ and $(x_i)_m$ are the values of X_i at the a-th and m-th samples, respectively. $(x_{k|k\neq i,j})_m$ are the values of the variables $X_{k|k\neq i,j}$ in the m-th sample.
  - ○ δ_m is our estimate of the difference

$$\mathbb{E}[X_j \mid X_i = (x_i)_a, X_{k|k\neq i,j} = (x_{k|k\neq i,j})_m] - \mathbb{E}[X_j \mid X_i = (x_i)_m, X_{k|k\neq i,j} = (x_{k|k\neq i,j})_m].$$

    Due to Theorem 1 of Schultheiss and Bühlmann (2024), this difference in conditional expectation values is the causal effect on the target variable X_j when X_i is changed from the value $(x_i)_m$ to the value $(x_i)_a$, while all remaining variables in Eq. (1), including the hidden ones, are fixed at their corresponding values in the m-th sample.
  - ○ If $\delta_m \neq 0$, X_i is a parent of X_j. Otherwise, X_i is not a parent of X_j.
- –– We use a t-test with significance level α for testing the null hypothesis of zero mean of the distribution of $\delta_m (m = 1, \ldots, M)$.
  - ⋆ If the null hypothesis is rejected, we conclude that X_i is a parent of X_j and set $A(j,i) = 1$ and $A(i,j) = 0$.
  - ⋆ If the null hypothesis is not rejected, we conclude that X_i is not a parent of X_j, and set $A(j,i) = 0$.

I.2 Computational environment

All experiments were performed on a desktop computer with an Intel Core i7-13700K processor and 64 GB of RAM.

I.3 Metrics

We define the metrics used in the experiments. TP is the number of true positives. TN is the number of true negatives. FN is the number of false negatives. FP is the number of false positives.

In identifying the adjacency matrix A, for the case where the estimated $A(i,j)$ is NaN, we add $0.5 \times P/(P+N)$, $0.5 \times N/(P+N)$, $0.5 \times P/(P+N)$, and $0.5 \times N/(P+N)$ to TP, TN, FN, and FP, respectively, to reflect the expected values of a random guess. Here, P and N are the total number of positives and negatives in the ground truth, respectively.

The $precision$, $recall$, and F_1 are calculated as follows. $Precision$ is $TP/(TP+FP)$. $Recall$ is $TP/(TP+FN)$. F_1 is $2 * (precision * recall)/(precision + recall)$.

I.4 ER graphs with Gaussian noise

We generated random graphs from the ER model with 10 observed variables and edge probability 0.2. We randomly selected 20 pairs of observed variables and introduced a hidden confounder between each pair, and for another 20 pairs of observed variables, we added a hidden intermediate variable. We generated 50 random graphs in this way. For each random graph, we generated one dataset with the same process as in Section 5.1. Each dataset contains 500 samples. We ran the algorithms with significance level $\alpha = 0.05$, 0.1, and 0.2. The results are shown in Fig. I.3. CAM-UV-X achieves performance comparable to that of CAM-UV.

I.5 Details on the experiment with the sociological data

We removed outliers in the data by retaining samples in which the value of x_2 (son's income) is at or below the 0.95 quantile of its sample distribution and the value of x_6 (number of siblings) is at or below the 0.95 quantile of its sample distribution. The final number of samples is 1262.

We ran CAM-UV and CAM-UV-X with $\alpha = 0.05$. For CAM-UV, we applied the following prior knowledge: each of x_2 (son's income), x_4 (son's occupation), and x_5 (son's education) cannot be parents of either x_1 (father's occupation), x_3 (father's education), or x_6 (number of siblings). We ran CAM-UV-X on the adjacency matrix estimated by CAM-UV, without any prior knowledge.

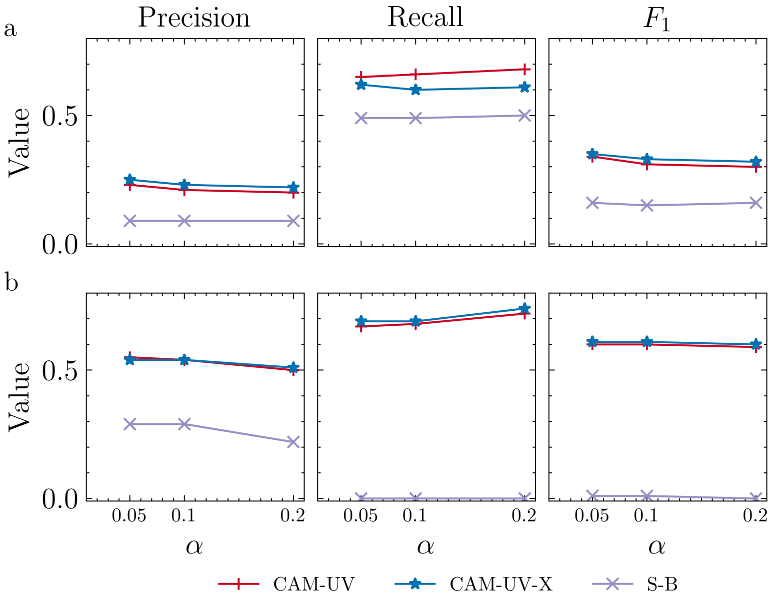

Figure I.3: Performance in ER random graphs. a: identifying the adjacency matrix, b: identifying ancestor relationships.