Detecting and Identifying Selection Structure in Sequential Data

Yujia Zheng [1] **Zeyu Tang** [1] **Yiwen Qiu** [1] **Bernhard Schölkopf** [2] **Kun Zhang** [1 3]

Abstract

We argue that the selective inclusion of data points based on latent objectives is common in practical situations, such as music sequences. Since this selection process often distorts statistical analysis, previous work primarily views it as a bias to be corrected and proposes various methods to mitigate its effect. However, while controlling this bias is crucial, selection also offers an opportunity to provide a deeper insight into the hidden generation process, as it is a fundamental mechanism underlying what we observe. In particular, overlooking selection in sequential data can lead to an incomplete or overcomplicated inductive bias in modeling, such as assuming a universal autoregressive structure for all dependencies. Therefore, rather than merely viewing it as a bias, we explore the causal structure of selection in sequential data to delve deeper into the complete causal process. Specifically, we show that selection structure is identifiable without any parametric assumptions or interventional experiments. Moreover, even in cases where selection variables coexist with latent confounders, we still establish the nonparametric identifiability under appropriate structural conditions. Meanwhile, we also propose a provably correct algorithm to detect and identify selection structures as well as other types of dependencies. The framework has been validated empirically on both synthetic data and real-world music.

1. Introduction

Selection arises from the preferential inclusion of certain data points based on latent variables that are dependent on some of the observed variables (Heckman, 1979). This selection process not only introduces bias into statistical inferences but is also a fundamental aspect of the data-generating process in various applications. For instance, in composing music, composers are guided by specific artistic goals or themes, leading them to selectively choose certain patterns of musical combinations (as combinations of basic elements) from their mind, thereby introducing dependencies among the basic elements in the music sequences (Schoenberg et al., 1967). These intentional but unmeasured selections, together with the contextual information, shape the structure of the compositions. A comprehensive understanding of the selection structure is essential for uncovering the underlying causal process and making use of it.

In sequential data, the understanding of selection plays a vital role. One essential question is whether selection leaves unique data dependence patterns that cannot be well explained by direct causal relations or latent confounding. Interestingly, as we will see in this paper, the answer is yes. Consequently, overlooking selection in such data can result in the introduction of incomplete or overcomplicated dependence models for the data. For instance, due to the sequential nature of the data, an autoregressive structure is usually assumed for the entire sequence (Radford et al., 2018), where each variable is considered primarily a function of its predecessors, perhaps with attention, and hence might be overcomplicated. Even when latent variables are incorporated to account for more complex dependencies, treating them solely as confounders amounts to a restricted generative view–as we will show, latent confounders and latent selection are distinguishable based on observational data, indicating their different footprints in the data. These perspectives neglect the potential influence of latent selection structures, such as the possibility that some dependencies among observed variables in the data sequence are due to a latent sampling criterion acting as the selection variable, rather than direct causal relations between variables. Confusing selection with latent confounding or direct causal relations can distort our understanding of the generative process. Therefore, discovering and integrating selection structure may be essential in the modeling of sequential data, implementing a more veridical inductive bias.

Previous research related to selection mainly focuses on controlling the impact of selection in data for inference or discovery (Spirtes et al., 1995; Hernán et al., 2004; Zhang, 2008; Bareinboim et al., 2014; Zhang et al., 2016; Correa et al., 2019; Forré & Mooij, 2020; Versteeg et al.,

[1]Carnegie Mellon University [2]Max Planck Institute for Intelligent Systems [3]MBZUAI.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

2022; Chen et al., 2024). These studies propose various conditions to mitigate the distortions caused by selection, aiming to achieve reliable statistical analysis even in the presence of selection bias, which holds significant practical value. At the same time, while these works constitute significant progress, they primarily view selection as a bias to be alleviated or a distortion to be overcome, rather than as an integral component of the data generation or selection process that requires further structural understanding. The FCI algorithm (Spirtes et al., 1995; Zhang, 2008) aims to discover ancestral relations up to an equivalence class in the presence of latent confounders and selection bias, but significant ambiguities remain for the selection structure. To truly discover the mechanism underlying the data, it is natural to explore deeper into the structure of selection, such as differentiating between dependencies caused by selection and those arising from confounding or direct causal relations. Recent work by Kaltenpoth & Vreeken (2023) introduces methods to detect whether or not selection bias is an issue for the considered population under specific parametric assumptions. Nevertheless, the results do not indicate any structure of the selection process, leaving a gap in comprehensively understanding the role of selection.

In this paper, we explore the selection structure in sequential data to gain insights into underlying mechanisms through the additional lens of selection. We first prove the identifiability of the selection structure without any parametric assumptions and interventional experiments (Theorem 3.2). Specifically, we can detect and identify different types of dependencies, including selection, direct causal relations, or both, in sequential data. Furthermore, even in cases with latent confounders, which add significant complexity, we prove that the selection structure can still be identified under additional graphical conditions (Theorem 3.4). Meanwhile, we also propose a provably correct algorithm to identify selection structure as well as other types of dependencies. We thus establish one of the first frameworks for uncovering the general structure of selection in sequential data from only observational data. The theoretical claims as well as the scalability of the algorithm have been validated empirically in various settings. Results on real-world music not only align with existing knowledge but also hint at new discoveries.

2. Preliminaries

The data $\mathbf{X} = \{X_1, X_2, \cdots, X_N\}$ consists of observed variables where each X_i represents an individual one. A crucial aspect of this data is its sequential structure, characterized by a property that if there exists any causal relation $X_i \to X_j$, where $X_i, X_j \in \mathbf{X}$, it holds that $i < j$. This structure is especially pertinent in temporal data, encapsulating the principle that future events cannot influence the past. The underlying causal structure is defined by a directed acyclic graph $\mathcal{G} = \{\mathbf{V}, \mathbf{E}\}$, where $\mathbf{V} = \{\mathbf{X}, \mathbf{S}, \mathbf{C}\}$ encapsulates all the observed variables $\mathbf{X}$, latent selection variables $\mathbf{S}$, and latent confounders $\mathbf{C}$. Each directed edge represents a direct causal relation.

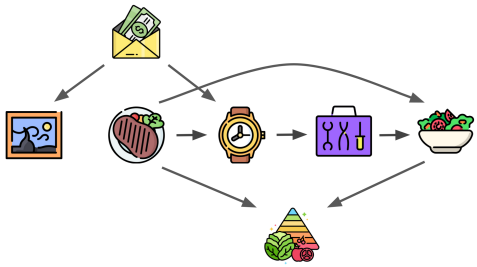


Figure 1. **Example of the selection structure** in the presence of latent confounders and direct relations. The figure depicts a sequence in a user's shopping history with a healthy lifestyle. Initially, the user buys an expensive painting from a gallery. Then the user visits a mall and first enjoys a steak meal at a restaurant. After that, the user spots a nearby watch store and then buys a pricey watch with its accompanying toolkit at the mall. Then, to adhere to a balanced diet, the user purchases salads from another nearby store within the mall. This data sequence was selected for a study investigating the daily habits of individuals identified as leading a healthy lifestyle. Therefore, the healthy lifestyle acts as a **selection variable**, of which one of the contributing factors is a balanced diet, exemplified by choices such as steak and salad. Concurrently, the user's financial status is a latent confounder (a common cause) for the user to buy both the painting and the watch. While there is no direct causal relation between having the steak and the painting, buying salad is directly caused by having the steak.

The selection is viewed as follows: Let $S_k \in \mathbf{S}$ denotes a selection variable, which is part of the data-generating process. We can only observe the data points for which the selection criterion is met, i.e., $S_k = 1$. As a consequence, the data distribution is actually the conditional one $P(\mathbf{X}|\mathbf{S})$. Moreover, selection variables depend on observed variables and introduce dependence between them. For example, in $X_i \to S_k \leftarrow X_j$, there is always a dependence between X_i and X_j since S_k is always being conditioned on. To differentiate selection variables from others, we illustrate them using double-circle nodes. Since our task is to distinguish different types of dependencies in sequential data, we assume that all selection variables depend on more than one observed variable. In addition, we denote a latent confounder as $C_k \in \mathbf{C}$, which is a hidden common cause of two (or more) observed variables. We allow the coexistence of latent confounders and selection variables, provided that there are no direct causal relations between these latent confounders and selection variables. Moreover, we define different types of dependencies between variables as follows, with an illustrative example including all of them in Figure 1.

Definition 2.1. (Selection Pair) A pair (X_i, X_j) is a *selection pair*, denoted as $(X_i, X_j)_{\mathrm{s}}$, if there is a selection

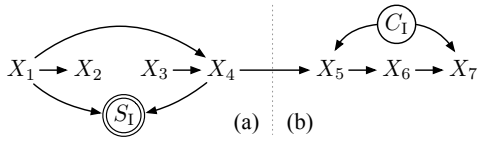


Figure 2. A running example for our identifiability theory.

variable S_k such that $X_i \to S_k \leftarrow X_j$.

Definition 2.2. (Direct Relation) A pair (X_i, X_j) is a *direct relation*, denoted as $(X_i, X_j)_{\mathrm{x}}$, if there exists a directed edge $X_i \to X_j$ in the causal graph. It is a higher-order direct relation if $j > i + 1$.

Definition 2.3. (Confounded Pair) A pair (X_i, X_j) is a *confounded pair*, denoted as $(X_i, X_j)_{\mathrm{c}}$, if there is a latent confounder C_k such that $X_i \leftarrow C_k \to X_j$.

Given only the observed variables $\mathbf{X}$, our goal is to identify these different types of dependencies. Specifically, we aim to determine the specific type for all *dependent pairs* among measured variables in the sequential data. We denote a pair (X_i, X_j) as a *dependent pair* $(X_i, X_j)_{\mathrm{d}}$, if it is a selection pair, direct relation, or confounded pair, i.e., $d \in \{\mathrm{s}, \mathrm{x}, \mathrm{c}\}$. Note that there could be multiple selection variables or latent confounders affecting a pair of variables, and we represent them as a single entity (e.g., S_k or C_k) for simplicity.

3. Identifiability of Selection Structure

Is the structure identifiable when selection coexists with other types of dependencies, as illustrated in Figure 2?

To answer this, we first establish the identifiability of selection structure without any parametric assumptions or interventional experiments (Section 3.1). Then we extend the identifiability result to cases with latent confounders (Section 3.2). After that, we illustrate alternative representations of the selection structures (Section 3.3). Finally, we propose a provably correct algorithm to identify selection structure as well as other types of dependencies (Section 3.4).

3.1. Identifiability Without Latent Confounders

We first show that the selection structure is identifiable in scenarios without latent confounders. For instance, in part (a) of Figure 2, we need to distinguish the selection pair $(X_1 \to S_I \leftarrow X_4)$ from the direct relation $(X_1 \to X_4)$. Under some structural conditions, we establish the nonparametric identifiability in sequential data by producing Theorem 3.2, which involves the following structural conditions:

Condition 3.1. (Structural Conditions).

i. For any $(X_i, X_j)_{\mathrm{s}}$, $(X_{j-1}, X_j)_{\mathrm{d}}$ exists with $j - 1 \neq i$, but not $(X_i, X_{j+1})_{\mathrm{d}}$, and X_j is not the effect of multiple higher-order direct relations. No $(X_i, X_{i+1})_{\mathrm{x}}$ if $j = i + 2$.

ii. For any $(X_i, X_j)_{\mathrm{d}\neq\mathrm{s}}$, $(X_{j-1}, X_j)_{\mathrm{d}}$ exists, and if X_j is also the cause of a selection variable, $(X_i, X_{j-1})_{\mathrm{d}}$ does not exist. No $(X_i, X_{j+1})_{\mathrm{d}}$ if $j = i + 2$.

Theorem 3.2. *Let the observed data be a large enough sample generated by a model defined in Section 2. In addition to the faithfulness assumption and Markov condition, suppose the following assumptions hold:*

i. (No future influencing the past): For any edge between observed variables, e.g., $X_i \to X_j$, $i < j$.

ii. (No latent confounders): $\mathbf{C} = \emptyset$.

iii. (Exclude degenerate structures): Structural conditions 3.1 are satisfied.

Then all selection pairs and direct relations in the causal graph $\mathcal{G}$ are identifiable.

Proof (Intuition and Sketch). The complete proof is in Appendix C.1. In particular, we develop an algorithm, Algorithm 1 (with its full version in Appendix A), and prove that it can identify all selection pairs and direct relations under the assumptions. We first show that Stage One of the algorithm (lines 3-5) detects all dependencies but not their specific types. This is achieved by conditional independence tests (CI tests) given a set of variables that blocks all other potential paths between X_i and X_j. Subsequently, Stage Two (lines 6-8) determines the specific type of dependence, whether it be a selection pair, a direct relation, or both.

Consider the two examples in Figure 3, since the latent selection variable is always conditioned on, the only structural difference between the direct relation $(X_i, X_j)_{\mathrm{x}}$ (Figure 3(b)) and the selection pair $(X_i, X_j)_{\mathrm{s}}$ (Figure 3(a)) lies in whether X_j is a collider on the path $\{X_i \cdots X_j \leftarrow \cdots X_k\}$. Thus, by blocking all other paths between X_i and X_k, we can identify whether X_j is a collider on the path by the designed CI tests. Specifically, X_i is independent of X_k conditioning on X_j and some other variables if we have $(X_i, X_j)_{\mathrm{s}}$ but not $(X_i, X_j)_{\mathrm{x}}$. Thus, we are able to distinguish selection pairs from direct relations. For a pair of variables that constitute both a selection pair and a direct relation (e.g., Figure 2(a)), we can identify them since they are dependent on each other, regardless of whether we condition on X_j or not. At the same time, given the flexibility of the potential structures, there are many different cases that need to be considered, and we include them all in Appendix C.1.

Discussion on Assumptions. The first two assumptions are from the considered setting. Assumption 3.2-i corresponds to the sequential structure of the data that is common in various scenarios, including but not limited to temporal data. Note that we do not assume a simple chain structure between variables as there could be arbitrary disconnected pairs in

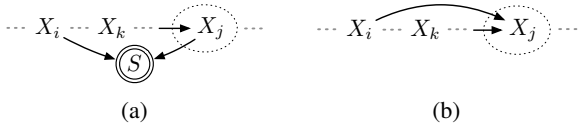


Figure 3. Illustration of the intuition for the proof of Theorem 3.2. In general, we distinguish the selection pair (a) from the direct relation (b) based on whether X_j is a collider on the path.

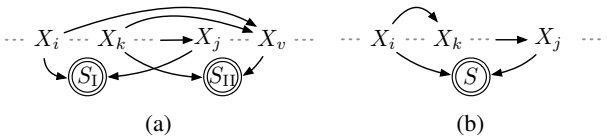


Figure 4. Examples of the degenerate structures in Theorem 3.2.

the sequence. Assumption 3.2-ii ensures that the only latent variables are selection variables.

However, since all selection variables are unobserved and we are only given the selected sub-population, uncovering any information regarding the latent selection is clearly non-trivial. For instance, Kaltenpoth & Vreeken (2023) employ parametric assumptions solely to detect the presence of selection bias in the entire dataset without any further structural information. In contrast, we aim to not only detect the existence but also identify the structure of selection, and thus some additional assumptions are necessary. Instead of introducing parametric constraints, we consider structural conditions to exclude several degenerate cases.

The intuition of these structural conditions (Condition 3.1 in Assumption 3.2-iii) stems naturally from the difference between different types of dependencies. As outlined in the proof sketch, the main focus is on determining whether X_j acts as a collider on the path $\{X_i \cdots X_j \leftarrow \cdots X_k\}$ (see Figure 3). For this, it is crucial that X_{j-1} is adjacent to X_j, indicated by the existence of $(X_{j-1}, X_j)_{\mathrm{d}}$. Without this adjacency, X_j cannot function as a collider. Additionally, we also need to block all other paths between X_i and X_k. Achieving this requires a carefully chosen conditioning set for d-separations (Pearl et al., 2000), accompanied by assumptions to exclude degenerate cases where the inclusion or exclusion of a variable in the conditioning set invariably leads to a d-connected path. For instance, if an observed variable X_v ($v \neq j$) is caused by both X_i and X_k and is part of a selection pair $(X_i, X_v)_{\mathrm{s}}$ (e.g., Figure 4(a)), a d-connected path between X_i and X_k will persist regardless of conditioning on X_v. Moreover, it is necessary to ensure the existence of such an X_k where (X_i, X_k) does not form a dependent pair (e.g., Figure 4(b)); otherwise, they will always be d-connected. Together with the algorithm (Algorithm 1), some structural conditions are needed to address cases like these. As a result, we introduce Condition 3.1 for the nonparametric identifiability of the selection structure given only observed data.

Implications. By establishing the nonparametric identifiability of the selection structure, we demonstrate the feasibility of going beyond the conventional generative view to examine specific selection processes in sequential data. This offers additional insights, potentially revealing overlooked aspects of certain inductive biases. For example, if the dependence between two variables is found to be due to selection rather than a direct causal relation, modeling all observed variables with an autoregressive structure could introduce significant bias. Consider a scenario in financial market analysis, where an algorithm examines stock return movements over time. In this sequence, X_i represents the stock return at time i, and X_j represents the stock price at a later time j (where $j > i$). A common assumption might be a direct relation, suggesting that X_i influences X_j. However, if the dataset only includes days where both X_i and X_j exceed a certain threshold – perhaps due to an internal data selection policy focusing on days with significant market activity – this introduces a selection structure. If selection is overlooked in data analysis, we might incorrectly interpret X_j as a variable directly influenced by X_i, and consequently model X_j as being generated by X_i along with other factors.

Furthermore, the knowledge of latent selection structures also has the potential to guide various machine learning tasks by providing more complete causal insights into the underlying mechanisms. For tasks that inherently require a trustworthy modeling process, such as interpretability, including selection structure is naturally beneficial, as it forms a crucial part of the overall causal process. Simultaneously, even for tasks that may not seem directly to benefit from a causal structure, such as pretraining, understanding the selection structure can still be helpful. In most tasks, striking a balance between inductive bias and data is essential. The revealed selection structure contributes to an inductive bias that more accurately reflects the hidden causal processes, thus enhancing the efficiency of data utilization. While a model trained with a less comprehensive inductive bias, such as a universal autoregressive structure, can still be effective, it often requires a substantial amount of data and incurs significant training costs. Incorporating potential selection structures makes it possible to adopt a more appropriate inductive bias, achieving our goals in a more cost-effective manner.

3.2. Identifiability With Latent Confounders

In the previous section, we established the identifiability of the selection structure in sequential data (Theorem 3.2). While this provides a way to delve deeper into latent selection, the setting of the absence of latent confounders might not hold in certain scenarios, given that latent confounders often exist. For example, we can identify part (a) of Figure 2 but not part (b). Motivated by that, we extend our theory and demonstrate the identifiability of the selection structure

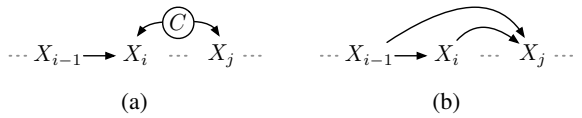


Figure 5. Intuition for identifying confounded pairs in Theorem 3.4. In general, we distinguish the confounded pair (a) from the spurious direct relations (b) by identifying the collider X_i.

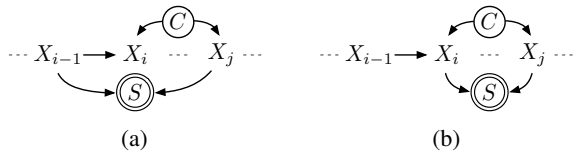


Figure 6. Examples of the degenerate structures in Theorem 3.4.

even in the presence of latent confounders.

Condition 3.3. For any $(X_i, X_j)_{\mathrm{c}}$, there exists no $(X_{i-1}, X_{j-1})_{\mathrm{d}}$, $(X_{i-1}, X_j)_{\mathrm{s}}$, $(X_{i-1}, X_{i+1})_{\mathrm{x}}$, $(X_i, X_j)_{\mathrm{s}}$, $(X_{i+1}, X_{j-1})_{\mathrm{d}}$, and $(X_{i+1}, X_{j+1})_{\mathrm{x}}$. Neither X_i nor X_j is in another confounded pair or higher-order direct relation.

Theorem 3.4. *Let the observed data be a large enough sample generated by a model defined in Section 2. In addition to the faithfulness assumption and Markov condition, suppose the following assumptions hold:*

i. (No future influencing the past): For any edge between observed variables, e.g., $X_i \to X_j$, $i < j$.

ii. (Exclude degenerate structures): Structural conditions 3.1 and 3.3 are satisfied.

Then all selection pairs, direct relations, and confounded pairs in the causal graph $\mathcal{G}$ are identifiable.

Proof (Intuition and Sketch). The complete proof is in Appendix C.2. Similar to Theorem 3.2, we prove that the algorithm (Algorithm 1) identifies the dependent pairs at Stage One and distinguishes selection pairs from others after Stage Two. However, due to the possibility of latent confounders, there exist some spurious direct relations corresponding to the unidentified confounded pairs after Stage Two. This is because, when conditioning on X_i for $(X_i, X_j)_{\mathrm{c}}$ (e.g., Figure 5), X_{i-1} and X_j will be d-connected by the path $\{X_{i-1} \to X_i \leftarrow C \to X_j$. Because both the confounded pair and direct relation have an arrowhead pointing to X_j, if there exists $(X_i, X_j)_{\mathrm{c}}$, both (X_i, X_j) and (X_{i-1}, X_j) will be identified as direct relations after Stage Two according to Theorem 3.2. Thus, we introduce Stage Three to identify spurious direct relations, specifically to distinguish Figure 5(a) from Figure 5(b), which is achieved by designed CI tests. Meanwhile, since the introduced latent confounders largely broaden the space of possible structures, we elaborate all potential cases in the complete proof (Appendix C.2).

Discussion on Assumptions. Compared to Theorem 3.2, we remove the assumption of no latent confounders in Theorem 3.4, thus increasing the applicability of our theory. However, this extension comes with a trade-off: the inclusion of latent confounders alongside latent selection variables significantly increases the complexity of the problem. To address this, it is necessary to have additional structural conditions (i.e., Conditions 3.3 in Assumption 3.4-ii).

As mentioned in the proof sketch, to identify the added latent confounders, we need to distinguish them from the spurious direct relations after Stage Two, i.e., differentiating Figures 5(a) with 5(b). Notably, in Figure 5(a), X_i acts as a collider on the path $\{X_{i-1} \to X_i \leftarrow C \to X_j\}$, a structure absent in Figure 5(b). This allows us to detect confounded pairs by testing for additional independence created by this collider structure. A crucial requirement is that X_{i-1} and X_i do not form dependent pairs with X_j, as this would invariably lead to dependence between X_{i-1} and X_j. For example, if $(X_{i-1}, X_j)_{\mathrm{d}}$ exist, X_{i-1} and X_j remain d-connected due to the non-inclusion of X_{i-1} in the conditioning set (Figure 6(a)). This is also the case if $(X_i, X_j)_{\mathrm{d}\neq\mathrm{c}}$ exists (Figure 6(b)), since X_i should not be in the conditioning set for testing the extra independence introduced by X_i acting as a collider on the path. Moreover, we need to make sure that the only d-connected path between X_{i-1} and X_j is $\{X_{i-1} \to X_i \leftarrow C \to X_j\}$ in Figure 5(a). This necessitates structural conditions, similar to those in Theorem 3.2, to ensure that alternative paths can be effectively blocked using the designed CI tests. Additionally, the introduction of latent confounders must not obstruct the identification of other dependency types. For instance, conditions regarding confounders are needed for the algorithm to be able to identify the selection structure, such as the existence of X_k in Stage Two, as outlined in the previous discussion on Theorem 3.2. In conjunction with our proposed algorithm (Algorithm 1), we introduce Condition 3.3 as an added structural constraint to facilitate the extension of our theory to scenarios that include latent confounders.

Implications. Theorem 3.4 shows that, even in the case where latent confounders exist, we can still identify selection structures in a manner. This is exciting since both selection variables and confounders are unobserved, yet we still prove that it is actually possible to distinguish them based on only observed data. Notably, this identifiability is established without imposing any parametric constraints or necessitating interventional experiments.

In real-world scenarios, especially in sequential data, latent confounders often coexist with selection processes. For instance, policy shifts, serving as selection variables, can influence which patients are admitted at different times, thereby impacting health metrics sequentially in longitudinal health studies. At the same time, unobserved health conditions, i.e., latent confounders, may also influence these metrics. Accurately acknowledging the presence of both selection

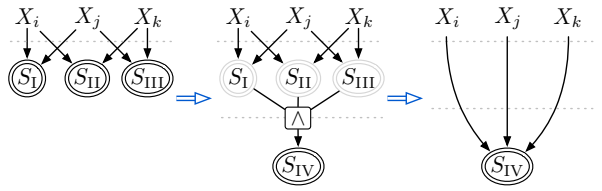


Figure 7. Alternative representations of the selection structure.

variables and latent confounders is key to developing a more nuanced and complete understanding of the data and its underlying dynamics. Compared to Theorem 3.2, which establishes identifiability of selection structures in the absence of latent confounders, Theorem 3.4 broadens the scope by encompassing scenarios that include latent confounders, enhancing both its theoretical significance and practical utility.

3.3. Alternative Representations of Selection Structure

We have demonstrated that selection structures are identifiable from observed data, both in the presence and absence of latent confounders. At the same time, in line with the principle of simplicity or Occam's razor (Zhang, 2013), we may consider merging multiple selection variables into a single entity when it simplifies the overall structure.

Remark 3.5. For a set of observed variables, where every variable pair is a selection pair, the conditional dependence structure is equivalent to one where all observed variables in this set are direct causes of a single selection variable.

This remark follows directly from the default conditioning of selection variables. If all variable pairs are selection pairs, conditioning on the selection variables d-connects each observed variable with others in the set. This is equivalent to a structure where all observed variables are d-connected by conditioning on a single selection variable acting as the common collider. In a scenario where every pair of variables within a set forms a selection pair, it is possible to construct a higher-level selection variable. This is achieved by merging all individual selection criteria using a logical AND conjunction, thereby consolidating multiple selection variables into one (e.g., Figure 7). For instance, in a health study assessing variables such as Heart Rate, Blood Pressure, and Cholesterol Levels, where each metric pair is affected by distinct factors like lifestyle, diet, and genetics, these factors can be unified into a singular selection variable. This composite variable, formed through the logical AND of all individual factors, simplifies the structure across the set of variables, reflecting a unified selection structure. Furthermore, if more than two observed variables depend on a single selection variable, we can always represent them in a pair-wise way, since the dependence patterns are exactly the same.

Additionally, structuring the selection process in this way may also help uncover a hierarchical nature of selection. This may reflect cognitive processes in decision-making.

Algorithm 1: Identification of the Selection Structure (*Sketch with full details in Appendix A*).

Input : Data $\mathbf{X}$
Output : A causal graph $\mathcal{G}$
Initialize lists $\mathcal{L}$ and $\mathcal{R}$ as empty;
Initialize an empty edge set $\mathcal{E}$ and a set of vertices $\mathcal{V}$;
// **Stage One**
ForEach *variable pair* (X_i, X_j) **Do**
  Identify potential dependent pairs by conducting designed CI tests on X_i and X_j;
  Update $\mathcal{L}$ and $\mathcal{R}$ with identified pairs;
// **Stage Two**
ForEach *variable in* $\mathcal{L}$ *and* $\mathcal{R}$ **Do**
  Distinguish between selection pairs and direct relations by conducting designed CI tests;
  Update the edge set $\mathcal{E}$ accordingly;
// **Stage Three** (only for Theorem 3.4)
ForEach *direct relation that may be confounded* **Do**
  Identify confounded pairs while removing spurious dependencies;
  Adjust $\mathcal{E}$ based on confounded pairs;
Return : $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$

Take Figure 7 for example, a composer crafting a musical piece might set a specific goal (S_{IV}), such as style or topic, for the whole piece, then divide it into different lower-level goals (S_{I}, S_{II}, and S_{III}) for individual passages.

3.4. Identification Algorithm

Alongside establishing identifiability, we propose an algorithm for the identification of selection structures, as well as direct relations and confounded pairs, using only observed data. A brief outline of it is presented as Algorithm 1, while comprehensive details can be found in Appendix A with an illustrative example in Appendix B. It should be noted that Stage Three of the algorithm is only for scenarios that include latent confounders, as addressed in Theorem 3.4. Consequently, this stage is not needed in situations where latent confounders are absent, such as Theorem 3.2. Our algorithm achieves a relatively lower complexity of $O(N^2)$, where N denotes the number of observed variables, by searching in a pair-wise manner and focusing solely on essential CI tests required for differentiating various cases. This stands in contrast to the often exponential complexity observed in other constraint-based structure learning algorithms (e.g., the complexity of PC or FCI in the worst case (Spirtes et al., 2000)). However, we do not claim that our method is more computationally efficient, as it targets a new and distinct task.

4. Experiments

In this section, we empirically validate our proposed theory and algorithm on both synthetic and real-world datasets.

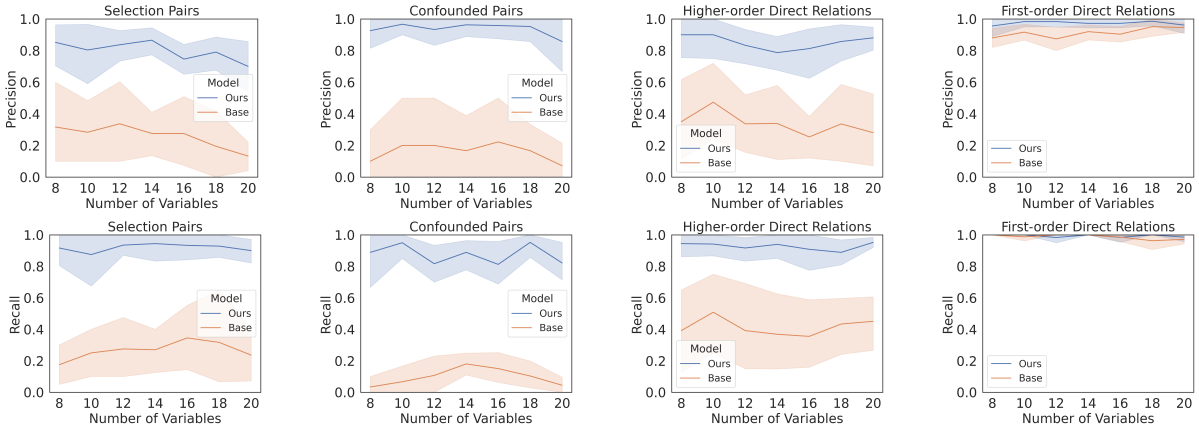


Figure 8. Comparison results on synthetic datasets.

4.1. Synthetic Data

Setup. The synthetic datasets are generated according to the selection process on the data generated by linear Gaussian Structural Causal Models (SCMs). We incorporate selection variables as part of the SCMs for preferential sampling, specifically, selecting samples from the unbiased population conditional on whether the values of selection variables exceed their mean. We sample $10,000$ data points from a sufficiently large unbiased population for simulations. For the ground-truth structure, we first randomly generate a basic chain structure, and then randomly drop half of the edges, after which the remaining edges are the first-order direct relations. Then, we randomly add selection pairs, higher-order direct relations, and/or confounded pairs. The number of each type of dependency is randomly sampled from $[0, N/2]$, where N is the number of observed variables. None of the selection variables and latent confounders are included in the observed data. The implementation of the algorithm directly follows the full details elaborated in Appendix A with Fisher's z-test (Fisher et al., 1921).

Identification. To validate our theoretical claims and demonstrate the necessity of the proposed conditions, we conduct a comparative analysis. This involves contrasting our model, in which all assumptions are met, with a base model that does not satisfy structural conditions. We vary the number of observed variables from $\{8, 10, \cdots, 20\}$, and randomly repeat every experiment 10 times. From Figure 8, it is clear that our algorithm can identify selection pairs as well as other types of dependencies under the proposed conditions, which corroborates our identifiability theory. At the same time, both the base model and ours successfully identify almost all first-order direct relations. This is reasonable since the identification of these first-order direct relations is less complex. However, the base model's precision in identifying first-order direct relations is lower than its recall, particularly when compared to our model. This discrepancy arises because, in the absence of appropriate structural conditions, the algorithm may erroneously classify some confounded pairs as first-order direct relations, leading to an increase in false positives. This observation further underscores the necessity of the proposed conditions.

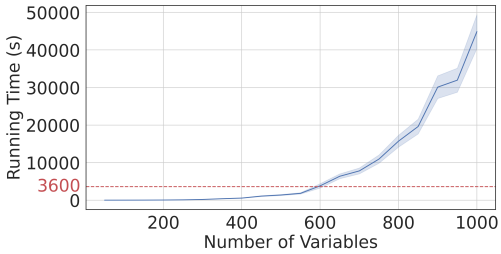
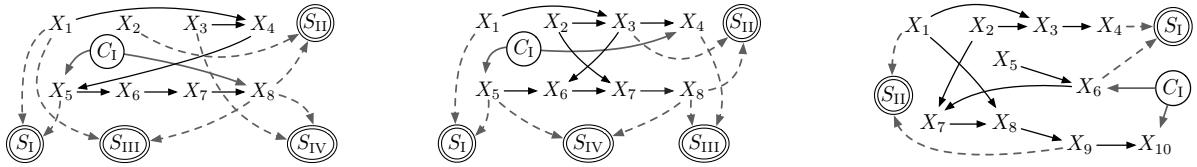


Figure 9. Running time for large graphs. The algorithm tackles a graph with roughly 600 variables within one hour (red dotted line).

Scalability. We also evaluate the scalability of our algorithm with $\{50, 100, \ldots, 1000\}$ observed variables. Since the number of selection pairs does not influence the number of CI tests needed for the algorithm, we limit it to up to 10 for sufficient samples after selection. All experiments are from 3 random trials with only CPUs and 12 GB of memory. From Figure 9, we observe that our algorithm can easily scale to hundreds of observed variables, handling about 600 variables in an hour (red dotted line), which underscores the scalability of our approach with low complexity of $O(N^2)$.

4.2. Real-World Data

Setup. In this subsection, we present experimental results on real-world music data. Music can be represented in different formats, and one can consider three main classes of representations: sheet music, symbolic representation, and audio (Müller, 2015). In our experiments, we consider guitar soundtrack clips from *Hotel California* performed by Eagles, and string section soundtrack clips from *Panis Angelicus*, a verse from the hymn *Sacris solemniis* and set



(a) *Hotel California* guitar solo motif (b) *Hotel California* guitar interplay (c) *Panis Angelicus* string section

Figure 10. Experimental results on real-world music data.

to music by César Franck.

Data Processing. We use a single node to represent each music segment of a length of one bar on sheet music for *Hotel California* guitar soundtracks (approximately 3.25 seconds per bar), and a length of two bars for *Panis Angelicus* (approximately 3 seconds per bar).

The music feature analysis literature has established significant connections between the characteristics in the frequency domain and various musical features, including timbre, pitch and harmony, emotion and mood (Kim et al., 2010; Yang & Chen, 2011; Müller, 2015). Therefore, in order to obtain different samples from the same music piece, we first perform Fast Fourier Transform (FFT) on the audio waveform data (e.g., a `.wav` or `.mp3` file). Then, we apply lower and upper cutoff frequencies, 20Hz and 20kHz, respectively, followed by the Mel scale conversion to better align with the human hearing range (O'Shaughnessy & Douglas, 1999). We sample one-dimensional features in the frequency domain for the music piece corresponding to each node. We demonstrate the causal relations among nodes from different music clips in Figure 10.

Results. Figure 10(a) presents the result corresponding to the soundtrack clip of one repetition of the *Hotel California* guitar solo motif, which is a short pattern of chord movements that are characteristic of the song. Figure 10(b) presents the result corresponding to the soundtrack clip of another guitar in the interplay. This clip is synchronized with the guitar solo motif presented in Figure 10(a). As we can see from the comparison between Figures 10(a) and 10(b), although the two causal graphs are not identical, one can observe similar selection and causal dependence patterns. Particularly, the selection based on the 1st and 5th bar of the iconic *Hotel California* harmonic motif, denoted by X_1 and X_5 forming a selected pair $X_1 \to S_{\mathrm{I}} \leftarrow X_5$ in both Figures 10(a) and 10(b), closely aligns with the overall progression of the Andalusian cadence (a popular musical progression often associated with Flamenco music). Such alignment is also observed via $X_1 \to S_{\mathrm{III}} \leftarrow X_8$ in Figure 10(a), and via $X_5 \to S_{\mathrm{IV}} \leftarrow X_8$ in Figure 10(b).

Figure 10(c) presents the result corresponding to the soundtrack clip of the string section in the hymn *Panis Angelicus*. We do not directly utilize the strophe (in Latin) of the hymn in our experiments. Here, we reference it to facilitate a more intuitive interpretation and validation of our results. We observe that the identified selection pattern based on the soundtrack clip exhibits alignment with the strophe to a certain extent. For instance, for the selected pair $X_4 \to S_{\mathrm{I}} \leftarrow X_6$, the corresponding lines are "figuris terminum" (for X_4) and "Manducat Dominum" (for X_6), holding significant meanings within the context of the hymn's religious and sacramental themes.

5. Conclusion

We argue that the phenomenon of sample selection may be involved in many types of data. Overlooking the selection process or confusing it either with latent confounding effects or direct causal relations may lead to an incomplete or overcomplicated inductive bias, especially in sequential data. Motivated by this, we establish a set of theoretical results demonstrating the identifiability of selection structure in sequential data. Instead of adopting the traditional view that mainly treats selection as a bias to be mitigated, we consider it as an informative mechanism underlying the data, seeking to delve deeper into the causal structure of selection. Specifically, we first prove that the selection structure in sequential data is identifiable without any parametric assumptions or interventional experiments. Moreover, even in the presence of latent confounders, we show that the nonparametric identifiability still holds true under appropriate structural conditions. At the same time, we propose a provably correct algorithm for the identification of selection structures as well as other types of dependencies. The validity of the theoretical claims and the efficacy of the algorithm have been rigorously evaluated under various settings, utilizing both synthetic and real-world data.

Since the structure of selection may provide an often-overlooked insight into the world underlying the data, the discovery results with an identifiability guarantee could act as a new and more veridical inductive bias for machine learning tasks, suggesting exciting directions on both efficiently utilizing the data and reliably uncovering the underlying truth. Currently, the lack of exploration beyond understanding is a limitation; accordingly, future work could be leveraging the uncovered structure to solve open problems such as building a trustworthy foundation model.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

Acknowledgement

We thank Peter Spirtes and anonymous reviewers for insightful comments and discussion. We would also like to acknowledge the support from NSF Grant 2229881, the National Institutes of Health (NIH) under Contract R01HL159805, and grants from Apple Inc., KDDI Research Inc., Quris AI, and Florin Court Capital.

References

Bareinboim, E., Tian, J., and Pearl, J. Recovering from selection bias in causal and statistical inference. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28, 2014.

Chen, L., Zoeter, O., and Mooij, J. M. Modeling latent selection with structural causal models. *arXiv preprint arXiv:2401.06925*, 2024.

Correa, J. D., Tian, J., and Bareinboim, E. Identification of causal effects in the presence of selection bias. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 2744–2751, 2019.

Fisher, R. A. et al. 014: On the" probable error" of a coefficient of correlation deduced from a small sample. 1921.

Forré, P. and Mooij, J. M. Causal calculus in the presence of cycles, latent confounders and selection bias. In *Uncertainty in Artificial Intelligence*, pp. 71–80. PMLR, 2020.

Heckman, J. J. Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, pp. 153–161, 1979.

Hernán, M. A., Hernández-Díaz, S., and Robins, J. M. A structural approach to selection bias. *Epidemiology*, pp. 615–625, 2004.

Kaltenpoth, D. and Vreeken, J. Identifying selection bias from observational data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., Speck, J. A., and Turnbull, D. Music emotion recognition: A state of the art review. In *Proc. ismir*, volume 86, pp. 937–952, 2010.

Müller, M. *Fundamentals of music processing: Audio, analysis, algorithms, applications*, volume 5. Springer, 2015.

O'Shaughnessy and Douglas. *Speech Communications: Human and Machine*. 1999.

Pearl, J. et al. Models, reasoning and inference. *Cambridge, UK: CambridgeUniversityPress*, 19(2):3, 2000.

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al. Improving language understanding by generative pre-training. 2018.

Schoenberg, A., Strang, G., and Stein, L. *Fundamentals of musical composition*. Faber and Faber, 1967. URL https://cir.nii.ac.jp/crid/1130282269094586112.

Spirtes, P., Meek, C., and Richardson, T. Causal inference in the presence of latent variables and selection bias. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pp. 499–506, 1995.

Spirtes, P., Glymour, C. N., and Scheines, R. *Causation, prediction, and search*. 2000.

Versteeg, P., Mooij, J., and Zhang, C. Local constraint-based causal discovery under selection bias. In *Conference on Causal Learning and Reasoning*, pp. 840–860. PMLR, 2022.

Yang, Y.-H. and Chen, H. H. *Music emotion recognition*. CRC Press, 2011.

Zhang, J. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16-17): 1873–1896, 2008.

Zhang, J. A comparison of three Occam's razors for markovian causal models. *The British journal for the philosophy of science*, 2013.

Zhang, K., Zhang, J., Huang, B., Schölkopf, B., and Glymour, C. On the identifiability and estimation of functional causal models in the presence of outcome-dependent selection. In *UAI*, 2016.

Zheng, Y., Huang, B., Chen, W., Ramsey, J., Gong, M., Cai, R., Shimizu, S., Spirtes, P., and Zhang, K. Causal-learn: Causal discovery in python. *Journal of Machine Learning Research*, 25(60):1–8, 2024.

A. Complete Algorithm

Algorithm 2: Identification of the Selection Structure

Input : Data $\mathbf{X}$
Output : A causal graph $\mathcal{G}$

▷ **Initialization**
Set N as the number of observed variables in $\mathbf{X}$;
Initialize two empty lists as $\mathcal{L} := [\cdot]$ and $\mathcal{R} := [\cdot]$, an empty edge set $\mathcal{E}$, and a set of vertices $\mathcal{V} := \{1, 2, \cdots, N\}$;

▷ **Stage One**
ForEach l *in* $\{N-1, N-2, \cdots, 1\}$ **Do**
  Set $i := 1$, $j := i + l$;
  While $j \leq N$ **Do**
    If $X_i \not\perp\!\!\!\perp X_j | \mathbf{X}_{Pre(j)\setminus i}$ **Then**
      If $\forall k \in \mathcal{R}[i], X_i \not\perp\!\!\!\perp X_j | \{\mathbf{X}_{Pre(j)\setminus i}, X_{j+1}, X_k\}$ $(X_{N+1} = \emptyset)$ **Then**
        $\mathcal{L}$.append(i);
        $\mathcal{R}[i]$.append(j);
    Set $i := i + 1$, $j := i + l$;

▷ **Stage Two**
Set $u := 0$;
ForEach $i \in ReverseSorted(\mathcal{L})$ **Do**
  ForEach $j \in ReverseSorted(\mathcal{R}[i])$ **Do**
    Set $k = j - 1$;
    While $k \in \mathcal{R}[i]$ *and* $k > i + 1$ **Do**
      $k := k - 1$;
    Let $\mathbf{T} = \{X_v$, where $(X_p, X_v)_s \in \mathcal{E}\}$;
    If $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{Pre(k)\setminus i}, X_j, \mathbf{T}\}$ *or* $k = i$ **Then**
      If $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{Pre(k)\setminus i}, X_{j+1}, \mathbf{T}\}$ $(X_{N+1} = \emptyset)$ *or* $k = i$ **Then**
        Add to $\mathcal{E}$ an edge $i \to j$;
      Else
        Add to $\mathcal{E}$ an edge $i \to j$;
        Add to $\mathcal{E}$ edges $i \to s_u$ and $j \to s_u$;
        Set $u := u + 1$;
        If $j - 1 \in \mathcal{R}[i]$ **Then**
          $\mathcal{R}[i]$.remove$(j-1)$;
    Else
      Add to $\mathcal{E}$ edges $i \to s_u$ and $j \to s_u$;
      Set $u := u + 1$;

▷ **Stage Three** (for latent confounders)
Set $u := 0$;
ForEach j *in* $\{N, N-1 \cdots, 3\}$ **Do**
  ForEach i *in* $\{N-1, N-2 \cdots, 2\}$ **Do**
    Let $\mathbf{T} = \{X_v$, where $(X_{i-1}, X_v)_s$, $(X_i, X_v)_s$, or $(X_{i+1}, X_v)_s$, $v > j\}$;
    If $i \to j$ *and* $i - 1 \to j$ **Then**
      If $X_{i-1} \not\perp\!\!\!\perp X_j | \{\mathbf{X}_{Pre(j)\setminus\{i-1,i\}}, X_{j+1}, \mathbf{T}\}$ $(X_{N+1} = \emptyset)$ **Then**
        Add to $\mathcal{E}$ edges $i - 1 \to j$ and $i \to j$;
      Else
        Add to $\mathcal{E}$ edges $c_u \to i$ and $c_u \to j$;
        Remove $i - 1 \to j$ and $i \to j$ from $\mathcal{E}$;
        Set $u := u + 1$;

Return : $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$.

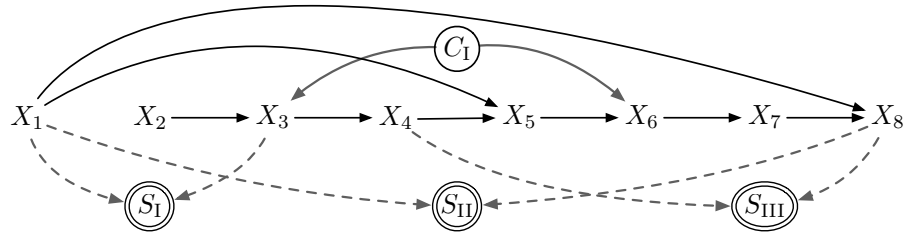


Figure 11. Example for the illustration of the algorithm.

B. Illustration of the Algorithm

In this section, we illustrate the complete algorithm (Algorithm 2) with a step-by-step example. Suppose our data $\mathbf{X} = \{X_1, X_2, \cdots, X_8\}$ is generated from the causal structure shown in Figure 11. Note that the complete proof of the correctness of the algorithm is included in Appendices C.1 and C.2. Here we just explain the procedure of the three stages based on the example in Figure 11:

- **Stage One:** For each pair of variables, the algorithm first tests if $X_i \not\perp\!\!\!\perp X_j | \mathbf{X}_{\text{Pre}(j)\setminus i}$ (line 6). According to the order of tests (lines 3 and 10), we first test (X_1, X_8). For line 6, we get the results $X_1 \not\perp\!\!\!\perp X_8 | \mathbf{X}_{\text{Pre}(8)\setminus 1}$. Since this is the first pair, $\mathcal{R}[1]$ is empty, and we also have $X_{j+1} = X_9 = \emptyset$. Thus, for line 7, we still have $X_1 \not\perp\!\!\!\perp X_8 | \mathbf{X}_{\text{Pre}(8)\setminus 1}$, and then we update $\mathcal{R}[1]$ as $\mathcal{R}[1] = \{8\}$.

  However, for $i = 1$ and $j = 7$, there is $X_1 \not\perp\!\!\!\perp X_7 | \mathbf{X}_{\text{Pre}(7)\setminus 1}$ even though (X_1, X_7) is not a dependent pair. This is because there exists a selection pair $(X_1, X_8)_{\text{s}}$, and if X_8 is not conditioned, the path $\{X_1 \to S_{\text{II}} \leftarrow X_8 \leftarrow X_7\}$ is d-connected, leading to the conditional dependence $X_1 \not\perp\!\!\!\perp X_7 | \mathbf{X}_{\text{Pre}(7)\setminus 1}$. Moreover, when we move to line 7, we still have $X_1 \perp\!\!\!\perp X_7 | \{\mathbf{X}_{\text{Pre}(7)\setminus 1}, X_8\}$ since (X_1, X_8) is also a direct relation on the path. The conditioning on X_8 although blocks the path $\{X_1 \to S_{\text{II}} \leftarrow X_8 \leftarrow X_7\}$, but also opens the path $\{X_1 \to X_8 \leftarrow X_7\}$. Therefore, we have $\mathcal{R}[1] = \{8, 7\}$, and include (X_1, X_7) as a spurious dependent pair.

  Then we move to (X_2, X_8), which is similar to (X_1, X_8) and will be included as a dependent pair. However, it gets complicated again when it comes to (X_1, X_6). Similarly, because of $\{X_1 \to S_{\text{II}} \leftarrow X_8 \leftarrow X_7 \leftarrow X_6\}$, there is $X_1 \not\perp\!\!\!\perp X_6 | \mathbf{X}_{\text{Pre}(6)\setminus 1}$ even though (X_1, X_6) is not a dependent pair. Nevertheless, (X_1, X_6) will not be included this time. This is because, for (X_1, X_6), the test result in line 7 is $X_1 \perp\!\!\!\perp X_6 | \{\mathbf{X}_{\text{Pre}(6)\setminus 1}, X_7, X_8\}$ ($X_{j+1} = 7$ and $8 \in \mathcal{R}[1]$), which is different from the case for (X_1, X_7). The reason is that, even if X_8 is both a collider and a non-collider, we have X_6 in the conditioning set, which blocks the path $\{X_1 \to X_8 \leftarrow X_7 \leftarrow X_6\}$. The path $\{X_1 \to S_{\text{II}} \leftarrow X_8 \leftarrow X_7 \leftarrow X_6\}$ is blocked by conditioning on X_8.

  The preceding pairs of variables are all similar to the cases discussed above until we meet (X_2, X_6). This pair is special since we have (X_3, X_6) as a confounded pair in the ground-truth structure. Thus, when conditioning on X_3, X_2 and X_6 are always d-connected. Therefore, we first have $X_2 \not\perp\!\!\!\perp X_6 | \mathbf{X}_{\text{Pre}(6)\setminus 1}$. Then, we also have $X_2 \not\perp\!\!\!\perp X_6 | \{\mathbf{X}_{\text{Pre}(6)\setminus 1}, X_7\}$, which indicates that (X_2, X_6) is included as a spurious dependent pair.

  The other pairs are all similar to the cases we have discussed. At the end of Stage One, we have the following results: $\mathcal{R}[1] = \{8, 7, 5, 3\}$, $\mathcal{R}[2] = \{6, 3\}$, $\mathcal{R}[3] = \{6, 4\}$, $\mathcal{R}[4] = \{8, 5\}$, $\mathcal{R}[5] = \{6\}$. $\mathcal{R}[6] = \{7\}$, and $\mathcal{R}[7] = \{8\}$.

- **Stage Two:** Then we move to the second stage with $\mathcal{L}$ and $\mathcal{R}$ from the first stage. We start from a descending order for both $\mathcal{L}$ and $\mathcal{R}$. The first pair is (X_7, X_8), which is identified as a direct relation since $k = i$. This is similar for (X_6, X_7) and (X_5, X_6), and both of them are identified as (first-order) direct relations.

  Now we move to (X_4, X_8), for which we have $k = 7$. For line 18, we have $X_4 \perp\!\!\!\perp X_7 | \{\mathbf{X}_{\text{Pre}(7)\setminus 4}, X_8\}$, where $\mathbf{T} = \emptyset$. Thus, (X_4, X_8) is identified as a selection pair. Then we identify (X_4, X_5) as a direct relation similar to the case discussed above.

  Next we move to (X_3, X_6), for which we have $k = 5$ and $\mathbf{T} = \{X_8\}$. Because $X_3 \not\perp\!\!\!\perp X_5 | \{\mathbf{X}_{\text{Pre}(5)\setminus 3}, X_6, X_8\}$ (line 18, $\mathbf{T} = \{X_8\}$) and $X_3 \perp\!\!\!\perp X_5 | \{\mathbf{X}_{\text{Pre}(5)\setminus 3}, X_7\}$ (line 19), we identify (X_3, X_6) as a direct relation. However, since (X_3, X_6) is a confounded pair, we now include a spurious direct relation in $\mathcal{E}$, which will be identified in Stage Three.

  After that, we identify (X_3, X_4) as a first-order direct relation, then we consider (X_2, X_6), for which we have $k = 5$ and $\mathbf{T} = \{X_8\}$. Note that, (X_2, X_6) is actually not a dependent pair. This is included in Stage Two only because

(X_3, X_6) is a confounded pair, and the collider X_3 on the path was always included in the conditioning set in Stage One. For Stage Two, because $X_2 \not\perp\!\!\!\perp X_5 | \{\mathbf{X}_{\text{Pre}(5)\setminus 2}, X_6, X_8\}$ (line 18) and $X_2 \perp\!\!\!\perp X_5 | \{\mathbf{X}_{\text{Pre}(5)\setminus 2}, X_7, X_8\}$ (line 19), (X_2, X_6) is still identified as a direct relation as a spurious one again.

Then after identifying (X_2, X_3) as a first-order direct relation, we move to (X_1, X_8). For this pair, we have $K = 6$ since $7 \in \mathcal{R}[1]$. Because $X_1 \not\perp\!\!\!\perp X_6 | \{\mathbf{X}_{\text{Pre}(6)\setminus 1}, X_8\}$ (line 18) and $X_1 \not\perp\!\!\!\perp X_6 | \{\mathbf{X}_{\text{Pre}(6)\setminus 1}, X_8\}$ (line 19), we identify (X_1, X_8) as both a direct relation and a selection pair. As a result, we also remove 7 from $\mathcal{R}[1]$ (line 26), so now we have $\mathcal{R}[1] = \{8, 5, 3\}$, and we remove to (X_1, X_5).

For (X_1, X_5), we have $k = 4$ and $\mathbf{T} = \{X_8\}$. Consequentially, we have $X_1 \not\perp\!\!\!\perp X_4 | \{\mathbf{X}_{\text{Pre}(4)\setminus 1}, X_5, X_8\}$ (line 18) and $X_1 \perp\!\!\!\perp X_4 | \{\mathbf{X}_{\text{Pre}(4)\setminus 1}, X_6, X_8\}$ (line 19). Thus, (X_1, X_5) is identified as a direct relation. Then we move to (X_1, X_3) with $k = 2$ and $\mathbf{T} = \{X_8\}$. For line 6, we have $X_1 \perp\!\!\!\perp X_2 | \{\mathbf{X}_{\text{Pre}(2)\setminus 1}, X_3, X_8\}$, which implies that we identify it as a selection pair.

Finally, we finished all tests in Stage Two. In summary, we have identified all selection pairs and direct relations, while misidentifying a confounded pair $(X_3, X_6)_{\text{c}}$ as a direct relation and introducing a spurious direct relation for (X_2, X_6) because of the confounded pair $(X_3, X_6)_{\text{c}}$.

- **Stage Three:** At this stage, we need to identify all confounded pairs while removing all spurious structures. It is worth noting that, according to the previous discussion, all spurious structures after Stage Two are caused by confounded pairs. Thus, if our setting does not allow the existence of latent confounders (Theorem 3.2), Stage Three will be removed from the algorithm.

  According to line 34, we only need to test (X_3, X_6) since we have both (X_3, X_6) and (X_2, X_6) are identified as direct relations in the previous stage. At this point, we have $\mathbf{T} = \{X_8\}$ (line 33) and $X_2 \perp\!\!\!\perp X_6 | \{\mathbf{X}_{\text{Pre}(6)\setminus\{2,3\}}, X_7, X_8\}$. Note that both X_7 and X_8 are essential in the conditioning set. If we do not condition on both, there will be a d-connected path $\{X_2 \to X_3 \to S_{\text{I}} \leftarrow X_1 \to S_{\text{II}} \leftarrow X_8 \leftarrow X_7 \leftarrow X_6\}$ that makes X_2 and X_6 conditionally dependent. This is because, if we only add X_8 to the conditioning set together with $\mathbf{X}_{\text{Pre}(6)\setminus\{2,3\}}$, there will be another d-connected path $\{X_2 \to X_3 \to S_{\text{I}} \leftarrow X_1 \to X_8 \leftarrow X_7 \leftarrow X_6\}$ between X_2 and X_6 since X_8 is a collider on the path. Only when we also add X_7 in the condition set, we will have $X_2 \perp\!\!\!\perp X_6 | \{\mathbf{X}_{\text{Pre}(6)\setminus\{2,3\}}, X_7, X_8\}$. Thus, according to lines 38-40, we identify (X_3, X_6) as a confounded pair, as well as remove $X_3 \to X_6$ and $X_2 \to X_6$ from the edge set $\mathcal{E}$, which are the only remaining spurious direct relations. After that, we have identified all confounded pairs while removing all spurious structures.

Now, we have identified all selection pairs, direct relations, and confounded pairs in Figure 11. The implementation of tests is based on *causal-learn* (Zheng et al., 2024).

B.1. Additional Discussion on the Algorithm

Beyond Structural Conditions. In the above example, we have shown how to identify different types of dependencies based on purely observational data, even when there may exist hidden confounders and selection variables. In real-world scenarios, it is important to acknowledge that some structural conditions may not consistently apply universally. The current algorithm is proposed to uncover the complete structure under the premise of identifiability, where all conditions are met. However, for practical applications, we can integrate condition checkers into the algorithm to allow potential ambiguities in the output when certain conditions are not satisfied. Meanwhile, the algorithm could be designed to raise a warning when it is unsure about certain parts of the output structure, and thus users could still make use of the remaining parts and deal with the uncertainty with their desired strategies (e.g., additional experiments).

Beyond Sequential Data. By discovering hidden selection structures from observational data, a more veridical inductive bias can be unveiled to apply in various tasks. In particular, the modeling of sequential data benefits a lot from that selection structure, since the commonly assumed autoregressive structure is usually overcomplicated and dense, and there could be latent confounders and selection variables that introduce different types of dependencies. However, the idea behind our proposed algorithm is also applicable beyond the setting where data has a sequential nature. The sequential structure is leveraged to distinguish selection from other types of dependencies in certain structures and does not need to be satisfied universally. For instance, if there is no selection among a subset of observed variables, their structure can be identified using previous constraint-based methods without assuming the sequential nature. Moreover, even when selection exists in the presence of potential latent confounders in certain structures, we can still identify it without the sequentiality assumption. Consider a non-sequential case with $A \to B$, $C \leftarrow D$, and an unknown dependence between B and C. If this dependence

arises from B causing C, C causing B, or latent confounders for them, there must be at least one edge along the path between B and C that has an arrowhead into B or C. This contrasts with scenarios where latent selection variables influence B and C, as these introduce additional colliders, distinguishable by testing for conditional dependence given the colliders under the faithfulness assumption. For certain dependencies that cannot be distinguished without the sequential structure, the algorithm may just remain and alert the indeterminacies in the output, or other conditions could be introduced if needed.

C. Proofs

C.1. Proof of Theorem 3.2

Theorem 3.2. *Let the observed data be a large enough sample generated by a model defined in Section 2. In addition to the faithfulness assumption and Markov condition, suppose the following assumptions hold:*

i. (No future influencing the past): For any edge between observed variables, e.g., $X_i \to X_j$, $i < j$.

ii. (No latent confounders): $\mathbf{C} = \emptyset$.

iii. (Exclude degenerate structures): Structural conditions 3.1 are satisfied.

Then all selection pairs and direct relations in the causal graph $\mathcal{G}$ are identifiable.

Proof. We prove the identifiability of the causal graph by showing that the algorithm 2 (only lines 1-29 since there are no latent confounders according to assumption ii) identifies all selection pairs and direct relations. Consistent with the two stages of the algorithm, the proof is divided into the following three stages:

- **Stage One**: Identifying dependent pairs together with several spurious structures.
- **Stage Two**: Identifying selection pairs, dependent pairs that are both selection pairs and direct relations, and direct relations while removing spurious structures.

Stage One. We first show that, after Stage One (lines 3-10), the identified sets of variable pairs, denoted as $\mathbf{E}_1$, only consist of the following types of cases:

- All selection pairs $(X_i, X_j)_{\mathrm{s}}$ in the ground-truth causal graph.
- All direct relations $(X_i, X_j)_{\mathrm{x}}$ in the ground-truth causal graph.
- All dependent pairs that are both selection pairs and direct relations, i.e., $(X_i, X_j)_{\{\mathrm{s,x}\}}$, in the ground-truth causal graph.
- All pairs of variables (X_i, X_j) that there exists (X_i, X_{j+1}) that is both a selection pair and a direct relation in the ground-truth causal graph.

For a variable pair (X_i, X_j), we discuss all potential cases for the paths between X_i and X_j as follows. We denote X_k as the observed variable closest to X_i.

I: (X_i, X_j) is a dependent pair. In this case, X_i is d-connected with X_j given any conditioning set, thus $X_i \not\perp\!\!\!\perp X_j | \mathbf{X}_{\mathrm{Pre}(j)\setminus i}$ and $(X_i, X_j)_{\mathrm{d}} \in \mathbf{E}_1$. This covers the cases where (X_i, X_j) is a selection pair, a direct relation, and a dependent pair which is both a selection pair and a direct relation. Note that there do not exist any confounded pairs according to assumption ii.

II: $\{X_i \cdots \leftarrow X_k \cdots X_j\}$, where $k < i$. Since $X_k \in \mathbf{X}_{\mathrm{Pre}(j)\setminus i}$, this path is d-connected only if X_k is a collider on the path. However, X_k cannot be a collider because of $\{X_i \leftarrow X_k\}$. Thus, $X_i \perp\!\!\!\perp X_j | \mathbf{X}_{\mathrm{Pre}(j)\setminus i}$ and $(X_i, X_j)_{\mathrm{d}} \notin \mathbf{E}_1$.

III: $\{X_i \cdots \to X_k \cdots X_j\}$, where $k < i$. Because of assumption i, $\{X_i \to X_k\}$ is impossible. Thus, there must be a confounded pair $\{X_k, X_i\}_{\mathrm{c}}$. Because of assumption ii, this is impossible and thus $(X_i, X_j)_{\mathrm{d}} \notin \mathbf{E}_1$.

IV: $\{X_i \cdots \leftrightarrow X_k \cdots X_j\}$, where $k < i$ and $\leftrightarrow$ means both tails and arrow heads exist. Because of assumption i, $\{X_i \rightarrow X_k\}$ is impossible. Thus, there must be a confounded pair $\{X_k, X_i\}_{\mathrm{c}}$. Because of assumption ii, this is impossible and thus $(X_i, X_j)_{\mathrm{d}} \notin \mathbf{E}_1$.

V: $\{X_i \cdots \leftarrow X_k \cdots X_j\}$, where $k > i$. Based on this, we discuss the following cases:

V.i: The case where $i < k < j$. Because of assumption i, $\{X_i \leftarrow X_k\}$ is impossible. Thus, there must be a selection pair $\{X_i, X_k\}_{\mathrm{s}}$, and X_k must not be a collider on the path. Since $\mathbf{X}_{\text{Pre}(j)\setminus i}$ is the condition set and $X_k \in \mathbf{X}_{\text{Pre}(j)\setminus i}$, we have $X_i \perp\!\!\!\perp X_j | \mathbf{X}_{\text{Pre}(j)\setminus i}$ ($X_k \in \mathbf{X}_{\text{Pre}(j)\setminus i}$) and thus $(X_i, X_j)_{\mathrm{d}} \notin \mathbf{E}_1$.

V.ii: The case where $k > j$. Because of assumption i, $\{X_i \leftarrow X_k\}$ is impossible. Thus, there must be a selection pair $\{X_i, X_k\}_{\mathrm{s}}$, and X_k must not be a collider. Since $k > j$, X_k is not in the conditioning set $\mathbf{X}_{\text{Pre}(j)\setminus i}$. However, the index k for any selection pair $\{X_i, X_k\}_{\mathrm{s}}$, where $k > j$, has already been included in $\mathcal{R}[i]$ according to the order of tests in the algorithm (lines 3-10). Therefore, there is the conditional independence $X_i \perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus i}, X_{j+1}, X_k\}$, which implies $(X_i, X_j)_{\mathrm{d}} \notin \mathbf{E}_1$.

VI: $\{X_i \cdots \rightarrow X_k \cdots X_j\}$, where $k > i$. Based on this, we discuss the following cases:

VI.i: The case where $i < k < j$. Since $\mathbf{X}_{\text{Pre}(j)\setminus i}$ is the condition set and $X_k \in \mathbf{X}_{\text{Pre}(j)\setminus i}$, X_k must be a collider for the path to be d-connected. On the path, let us denote the X_p as the observed variable closest to X_k other than X_i. Thus, we must have $\{X_k \leftarrow \cdots X_p\}$ on the path. We further have the following cases:

- The case where $p < k$. If $p < k$, we have $p < i$ as $p \neq i$. Since $\{X_k \leftarrow \cdots X_p\}$, (X_k, X_p) is either a confounded pair or direct relation. Because of assumption ii, (X_k, X_p) can only be a direct relation. Therefore, X_p must not be a collider. So we have $X_i \perp\!\!\!\perp X_j | \mathbf{X}_{\text{Pre}(j)\setminus i}$, which implies $(X_i, X_j)_{\mathrm{d}} \notin \mathbf{E}_1$.
- The case where $p > k$. In this case, (X_k, X_p) can only be a confounded pair. Because of assumption ii, this is impossible and thus $(X_i, X_j)_{\mathrm{d}} \notin \mathbf{E}_1$.

VI.ii: The case where $k > j$. If $k > j$ ($X_k \notin \mathbf{X}_{\text{Pre}(j)\setminus i}$), the only case where the path between X_i and X_j is not d-separated by X_k is the case where X_k is the cause of a selection variable. However, since we have $\{X_i \cdots \rightarrow X_k\}$, it is not possible for (X_i, X_k) to be a selection pair. For any other selection pairs, in order to have $X_i \not\perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus i}\}$, (X_j, X_k) must be a selection pair. According to assumption iii, we have $k \neq j+1$. Therefore, the path is d-separated by either k, where $k \in \mathcal{R}[i]$, or $j+1$. Consequently, there must be the conditional independence $X_i \not\perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus i}, X_{j+1}, X_k\}$ and thus $(X_i, X_j)_{\mathrm{d}} \notin \mathbf{E}_1$.

VII: $\{X_i \cdots \leftrightarrow X_k \cdots X_j\}$, where $k > i$. Based on this, we discuss the following cases:

VII.i: The case where $i < k < j$. Since $\mathbf{X}_{\text{Pre}(j)\setminus i}$ is the condition set and $X_k \in \mathbf{X}_{\text{Pre}(j)\setminus i}$, X_k must be a collider for the path to be d-connected. At the same time, (X_i, X_k) must be a selection pair. We further discuss the following cases:

- The case where $k = i+1$. According to assumption iii, (X_i, X_k) cannot be a selection pair if $k = i+1$, which is a contradiction.
- The case where $k \neq i+1$. On the path, let us denote the X_p as the observed variable closest to X_k other than X_i. Thus, we must have $\{X_k \leftarrow \cdots X_p\}$ on the path. We further have the following cases:
  - The case where $p < k$. If $p < k$, we have $p < i$ as $p \neq i$. Since $\{X_k \leftarrow \cdots X_p\}$, (X_p, X_k) is either a confounded pair or a direct relation. Because of assumption ii, (X_p, X_k) can only be a direct relation. Therefore, X_p must not be a collider, and $X_i \perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus i}\}$, which implies $(X_i, X_j)_{\mathrm{d}} \notin \mathbf{E}_1$.
  - The case where $p > k$. In this case, (X_k, X_p) can only be a confounded pair. Because of assumption ii, this is impossible and thus $(X_i, X_j)_{\mathrm{d}} \notin \mathbf{E}_1$.

VII.ii: The case where $k > j$. If $k > j$ ($X_k \notin \mathbf{X}_{\text{Pre}(j)\setminus i}$), the only case where the path between X_i and X_j is not d-separated by X_k is the case where X_k is the cause of a selection variable. However, the index k for any selection pair $\{X_i, X_k\}_{\mathrm{s}}$, where $k > j$, has already been included in $R[i]$ according to the order of tests in the algorithm (lines 3-10). According to assumption iii, X_k cannot be caused by more than two observed variables via higher-order direct relations. Therefore, the only case for X_i to be conditionally dependent of X_j given $\{\mathbf{X}_{\text{Pre}(j)\setminus i}, X_{j+1}, X_k\}$ for all $k \in \mathcal{R}[i]$ is the structure that $(X_i, X_{j+1})_{\{s,x\}}$. At the same time, we also have $X_i \not\perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus i}\}$. As a result, we have $(X_i, X_j)_{\mathrm{d}} \in \mathbf{E}_1$ where (X_i, X_{j+1}) is both a selection pair and a direct relation.

Stage Two. Next, we prove that we can identify selection pairs, direct relations, and dependent pairs that are both selection pairs and direct relations in $\mathcal{E}_1$ after lines 11-29 in the algorithm.

We discuss the following cases for all types in $\mathcal{E}_1$:

I: (X_i, X_j) is a selection pair but not a direct relation. According to the algorithm, it will be identified as a selection pair but not a direct relation if and only if $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$ (the definition of $\mathbf{T}$ is in line 17 in the algorithm).

We first show that all these selection pairs will be identified. If (X_i, X_j) is a selection pair but not a direct relation, X_j must be a non-collider on the path since there is no latent confounder according to assumption ii, which may lead to the d-separation and thus the conditional independence. Thus, we need to show that there exists no other path that leads to conditional dependence. We prove it by contradiction.

Suppose that there exist a path other than $\{X_i \to S \leftarrow X_j \cdots X_k\}$ that makes $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$. We discuss all potential cases for that path:

I.i: $\{X_i \cdots \leftarrow X_p \cdots X_k\}$, where $p < i$ and X_p denotes the observed variable that is the closest to X_i on the path. Since $X_p \in \mathbf{X}_{\text{Pre}(k)\setminus i}$, this path is d-connected if and only if X_p is a collider on the path. However, X_p cannot be a collider because of $\{X_i \leftarrow X_p\}$. Thus, this case is impossible.

I.ii: $\{X_i \cdots \to X_p \cdots X_k\}$, where $p < i$ and X_p denotes the observed variable that is the closest to X_i on the path. Because of assumption i, $\{X_i \to X_p\}$ is impossible. Thus, there must be a confounded pair $\{X_p, X_i\}_{\text{c}}$. Because of assumption ii, there are no latent confounders and thus this case is impossible.

I.iii: $\{X_i \cdots \leftrightarrow X_p \cdots X_k\}$, where $p < i$ and X_p denotes the observed variable that is the closest to X_i on the path. Because of assumption i, $\{X_p \to X_i\}$ is impossible. Thus, there must be a confounded pair $\{X_p, X_i\}_{\text{c}}$. Because of assumption ii, there are no latent confounders and thus this case is impossible.

I.iv: $\{X_i \cdots \leftarrow X_p \cdots X_k\}$, where $i < p < k$ and X_p denotes the observed variable that is the closest to X_i on the path. Because of assumption i, $\{X_i \leftarrow X_p\}$ is impossible. Thus, there must be a selection pair $(X_i, X_p)_{\text{s}}$, and X_p must not be a collider. Since $\{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$ is the condition set and $X_p \in \mathbf{X}_{\text{Pre}(k)\setminus i}$, we have $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$ ($X_p \in \mathbf{X}_{\text{Pre}(k)\setminus i}$) and thus this case is impossible.

I.v: $\{X_i \cdots \to X_p \cdots X_k\}$, where $i < p < k$. Since $\mathbf{X}_{\text{Pre}(k)\setminus i}$ is in the condition set and $X_p \in \mathbf{X}_{\text{Pre}(k)\setminus i}$, X_p must be a collider for the path to be d-connected. On the path, let us denote the X_q as the observed variable closest to X_p other than X_i. Thus, we must have $\{X_p \leftarrow \cdots X_q\}$ on the path. We further have the following cases:

- The case where $q < p$. If $q < p$, we have $q < i$ as $q \neq i$. Since $\{X_p \leftarrow \cdots X_q\}$, (X_q, X_p) is either a confounded pair or direct relation. Because of assumption ii, it can only be a direct relation. If (X_q, X_p) is a direct relation, X_q must not be a collider, and $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$ ($X_q \in \mathbf{X}_{\text{Pre}(k)\setminus i}$) and thus this case is impossible.
- The case where $q > p$. In this case, (X_p, X_q) can only be a confounded pair. Because of assumption ii, there are no latent confounders and thus this case is impossible.

I.vi: $\{X_i \cdots \leftrightarrow X_p \cdots X_k\}$, where $i < p < k$. Since $\mathbf{X}_{\text{Pre}(k)\setminus i}$ is in the condition set and $X_p \in \mathbf{X}_{\text{Pre}(k)\setminus i}$, X_p must be a collider for the path to be d-connected. At the same time, (X_i, X_p) must be a selection pair. We further discuss the following cases:

- The case where $p = i + 1$. According to assumption iii, (X_i, X_p) cannot be a selection pair if $p = i + 1$, which is a contradiction.
- The case where $p \neq i + 1$. On the path, let us denote the X_q as the observed variable closest to X_p other than X_i. Thus, we must have $\{X_p \leftarrow \cdots X_q\}$ on the path. We further have the following cases:
  - The case where $q < p$. If $q < p$, we have $q < i$ as $q \neq i$. Since $\{X_p \leftarrow \cdots X_q\}$, (X_q, X_p) is either a confounded pair or a direct relation. Because of assumption ii, (X_q, X_p) must be a direct relation, X_q must not be a collider, and the path is, again, not d-connected. Thus, this case is impossible.
  - The case where $q > p$. In this case, (X_p, X_q) can only be a confounded pair. Because of assumption ii, there should not be any latent confounders. Thus, this case is impossible.

I.vii: $\{X_i \cdots \leftarrow X_p\}$, where $k = p$. This is impossible because of the algorithm, specifically, how k is selected (lines 20-25).

I.viii: $\{X_i \cdots \to X_p\}$, where $k = p$. This is also impossible because of the algorithm.

I.ix: $\{X_i \cdots \leftrightarrow X_p\}$, where $k = p$. Similarly, this is impossible according to the algorithm.

I.x: $\{X_i \cdots \leftarrow X_p \cdots X_k\}$, $k < p < j$. Because of assumption i, $\{X_i \leftarrow X_p\}$ is impossible. Thus, there must be a selection pair $(X_i, X_p)_{\mathrm{s}}$. Together with the search procedure (lines 11-16), it is thus impossible for the selected k to be smaller than p, which is a contradiction.

I.xi: $\{X_i \cdots \rightarrow X_p \cdots X_k\}$, $k < p < j$. Because of $\{X_i \cdots \rightarrow X_p\}$, (X_i, X_p) is a higher-order direct relation. Since X_p is not in the condition set, (X_i, X_p) must also be a selection pair for the path to be d-connected. However, according to assumption iii, there exits no $(X_i, X_{p+1})_{\mathrm{d}}$. This contradicts the search procedure to select k (lines 11-16), and thus this case is impossible.

I.xii: $\{X_i \cdots \leftrightarrow X_p \cdots X_k\}$, $k < p < j$. Because of assumption i, $\{X_i \leftarrow X_p\}$ is impossible. Thus, there must be a selection pair $(X_i, X_p)_{\mathrm{s}}$. Similarly, it is thus impossible for the selected k to be smaller than p, which is a contradiction.

I.xiii: $\{X_i \cdots \leftarrow X_p \cdots X_k\}$, $j < p$. Because of assumption i, $\{X_i \leftarrow X_p\}$ is impossible. Thus, there must be a selection pair $(X_i, X_p)_{\mathrm{s}}$. Moreover, according to assumption iii, X_p cannot be the effect of multiple higher-order directions. Thus, X_p must not be a collider on the path, so we can d-separate the path by conditioning on it. Since X_p is already in $\mathcal{T}$ according to the algorithm, we have $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\mathrm{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$ and thus this case is impossible.

I.xiv: $\{X_i \cdots \rightarrow X_p \cdots X_k\}$, $j < p$. In this case, (X_i, X_p) is either a higher-order direct relation or a confounded pair. Because of assumption ii, there are no latent confounders and thus it can only be a higher-order direct relation. We further discuss the following case:

- X_p is not a cause of a selection variable. In this case, X_p must be a collider on the path, and the path is d-separated when not conditioning on X_p. So we have $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\mathrm{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$, which indicates that this case is impossible.
- X_p is a cause of a selection variable. According to assumption iii, it cannot be the effect of multiple higher-order directions. Therefore, if X_p is the cause of another selection variable together with any observed variable with index $u \in [k, j)$, the path is always d-separated by conditioning on both X_p and X_j. At the same time, if X_p is the cause of another selection variable together with any observed variable with an index larger than j, that variable will also be in the condition set ($\mathbf{T}$) according to the algorithm (line 17). Recursively, all these variables will be in the condition set, and the path will be d-separated, which indicates that this case is impossible.

I.xv: $\{X_i \cdots \leftrightarrow X_p \cdots X_k\}$, $j < p$. Because of assumption i, $\{X_i \leftarrow X_p\}$ is impossible. Thus, there must be a selection pair $(X_i, X_p)_{\mathrm{s}}$. At the same time, we have $(X_i, X_p)_{\mathrm{d}\neq\mathrm{c}}$, i.e., it is both a selection pair and a direct relation. According to assumption iii, it cannot be the effect of multiple higher-order directions. Therefore, if X_p is the cause of another selection variable together with any observed variable with index $u \in [k, j)$, the path is always d-separated by conditioning on both X_p and X_j. At the same time, if X_p is the cause of another selection variable together with any observed variable with an index larger than j, that variable will also be in the condition set ($\mathbf{T}$) according to the algorithm (line 17). Recursively, all these variables will be in the condition set, and the path will be d-separated, which indicates that this case is impossible.

Therefore, there does not exist any path other than $\{X_i \rightarrow S \leftarrow X_j \cdots X_k\}$ that makes $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\mathrm{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$. So all selection pairs $(X_i, X_j)_{\mathrm{s}}$ are identified.

We then discuss the other cases in $\mathcal{E}_1$ and show that it is impossible for any other types of variable pairs in $\mathcal{E}_1$ to have the test results that $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\mathrm{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$.

- (X_i, X_j) is a direct relation but not a selection pair. In this case, since X_j is collider on the path $\{X_i \cdots \rightarrow X_j\}$, for any $v \in \mathbf{T}$, there is always $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\mathrm{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$. Thus, this case is impossible.
- (X_i, X_j) is both a selection pair and a direct relation. In this case, since X_j is still a collider on the path $\{X_i \cdots \rightarrow X_j\}$, for any $v \in \mathbf{T}$, there is always $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\mathrm{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$, which indicates that this case is impossible.
- (X_i, X_j) is a spurious dependent pair that there exists (X_i, X_{j+1}) that is both a selection pair and a direct relation in the ground-truth causal graph. In this case, (X_i, X_j) is not a dependent pair according to assumption iii, and will be removed by the algorithm (line 26). Therefore, this case is impossible.

II: (X_i, X_j) is a direct relation but not a selection pair. We first show that all these variable pairs will be identified. In this case, X_j is a collider on the path. Thus, we have $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\mathrm{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$. However, when X_j is not in the condition set, we may have the d-separation and thus the conditional independence $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\mathrm{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$

(the definition of $\mathbf{T}$ is in line 17 in the algorithm). Thus, we need to show that there exists no other path that leads to conditional dependence. We prove it by contradiction.

Suppose that there exist a path other than $\{X_i \to X_j \cdots X_k\}$ that makes $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$. Denote X_p as the observed variable that is the closest to X_i on the path, we discuss all potential cases for that path:

II.i: The discussions for the following cases are similar to those in the cases where (X_i, X_j) is a selection pair but not a direct relation:

- $\{X_i \cdots \leftarrow X_p \cdots X_k\}$, where $p < i$, $i < p < k$, or $k < p < j$.
- $\{X_i \cdots \to X_p \cdots X_k\}$, where $p < i$ or $k < p < j$.
- $\{X_i \cdots \leftrightarrow X_p \cdots X_k\}$, where $p < i$, $i < p < k$, or $k < p < j$.
- $\{X_i \cdot X_p\}$, where $p = k$.

Thus, we mainly focus on the other cases as follows.

II.ii: $\{X_i \cdots \to X_p \cdots X_k\}$, where $i < p < k$. Since $\mathbf{X}_{\text{Pre}(k)\setminus i}$ is in the condition set and $X_p \in \mathbf{X}_{\text{Pre}(k)\setminus i}$, X_p must be a collider for the path to be d-connected. On the path, let us denote the X_q as the observed variable closest to X_p other than X_i. Thus, we must have $\{X_p \leftarrow \cdots X_q\}$ on the path. We further have the following cases:

- The case where $q < p$. If $q < p$, we have $q < i$ as $q \neq i$. Since $\{X_p \leftarrow \cdots X_q\}$, (X_p, X_q) is either a confounded pair or a direct relation. Because of assumption ii, it can only be a direct relation. Therefore, X_q must not be a collider, and $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, \mathbf{T}\}$ ($X_{q-1} \in \mathbf{X}_{\text{Pre}(k)\setminus i}$) and thus this case is impossible.
- The case where $q > p$. In this case, (X_p, X_q) can only be a confounded pair. Because of assumption ii, there are no latent confounders and thus this case is impossible.

II.iii: $\{X_i \cdots \leftarrow X_p \cdots X_k\}$, $j < p$. Because of assumption iii, $\{X_i \leftarrow X_p\}$ is impossible. Thus, there must be a selection pair $(X_i, X_p)_{\text{s}}$. Moreover, according to assumption iii, X_p cannot be the effect of multiple higher-order directions. Thus, X_p must not be a collider on the path, so we can d-separate the path by conditioning on it. Since X_p is already in $\mathbf{T}$ according to the algorithm, we have $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$ and thus this case is impossible.

II.iv: $\{X_i \cdots \to X_p \cdots X_k\}$, $j < p$. In this case, (X_i, X_p) is either a higher-order direct relation or a confounded pair. Because of assumption ii, there are no latent confounders, so it can only be a higher-order direct relation. We further discuss the following case:

- X_p is not a cause of a selection variable. In this case, X_p must be a collider on the path, and the path is d-separated when not conditioning on X_p. So we have $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$, which indicates that this case is impossible.
- X_p is a cause of a selection variable. According to assumption iii, it cannot be the effect of multiple higher-order directions. Meanwhile, the index p has already been included in $R[i]$ according to the order of tests in the algorithm (lines 12 and 13). At the same time, it is impossible for p to be $j + 1$ according to assumption iii. This is because if X_{j+1} is a cause of a selection variable while (X_i, X_{j+1}) is a higher-order direct relation, it is impossible to have $(X_i, X_j)_{\text{d}}$. Therefore, the path is always d-separated by conditioning on $\{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$ and thus this case is impossible.

II.v: $\{X_i \cdots \leftrightarrow X_p \cdots X_k\}$, $j < p$. Because of assumption i, $\{X_i \leftarrow X_p\}$ is impossible. Thus, there must be a selection pair $(X_i, X_p)_{\text{s}}$. At the same time, we have $(X_i, X_p)_{\text{d}\neq\text{c}}$, i.e., it is both a selection pair and a direct relation. According to assumption iii, it cannot be the effect of multiple higher-order directions. Meanwhile, it is impossible for p to be $j + 1$ according to assumption iii. This is because if X_{j+1} is a cause of a selection variable while (X_i, X_{j+1}) is a higher-order direct relation, it is impossible to have $(X_i, X_j)_{\text{d}}$. Therefore, the path is always d-separated by conditioning on $\{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$ and thus this case is impossible.

We then discuss the other cases in $\mathcal{E}_1$ and show that it is impossible for any other types of variable pairs in $\mathcal{E}_1$ to be identified as a direct relation.

- (X_i, X_j) is a selection pair but not a directed relation. As already shown before, it will always be identified as a selection pair but not a direct relation. Thus, this case is impossible
- (X_i, X_j) is both a selection pair and a direct relation. In this case, since X_j is still a collider on the path $\{X_i \cdots \to X_k\}$, there is always $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$. At the same time, since (X_i, X_j) is also a selection pair, not conditioning on X_j will also open the path. Thus, we also have $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$, which implies that this variable pair will not be identified as just a direct relation according to the algorithm (lines 21-24).

- (X_i, X_j) is a spurious dependent pair that there exists (X_i, X_{j+1}) that is both a selection pair and a direct relation in the ground-truth causal graph. In this case, (X_i, X_j) is not a dependent pair according to assumption iii, and will be removed by the algorithm (line 26). Therefore, this case is impossible.

III: (X_i, X_j) is both a selection pair and a direct relation. In this case, since X_j is still a collider on the path $\{X_i \cdots \rightarrow X_k\}$, there is always $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$. At the same time, since (X_i, X_j) is also a selection pair, not conditioning on X_j will also open the path. Thus, we also have $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$. We then discuss the other cases in $\mathcal{E}_1$ and show that, except for those variable pairs that are both a selection pair and a direct relation, it is impossible for any other types of variable pairs in $\mathcal{E}_1$ to be identified as it.

- (X_i, X_j) is a selection pair but not a directed relation. As already shown before, it will always be identified as a selection pair but not a direct relation. Thus, this case is impossible
- (X_i, X_j) is a direct relation but not a selection pair. In this case, according to the discussion before, there is always the conditional independence $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$. Thus, it is impossible for this case to be identified as a variable pair i.e. both a selection pair and a direct relation.
- (X_i, X_j) is a spurious dependent pair that there exists (X_i, X_{j+1}) that is both a selection pair and a direct relation in the ground-truth causal graph. In this case, (X_i, X_j) is not a dependent pair according to assumption iii, and will be removed by the algorithm (line 26). Therefore, this case is impossible.

IV: (X_i, X_j) is a spurious dependent pair that there exists (X_i, X_{j+1}) that is both a selection pair and a direct relation in the ground-truth causal graph. In this case, (X_i, X_j) is not a dependent pair according to assumption iii, and will be removed by the algorithm (line 26).

Therefore, after Stage Two, we identify all selection pairs and direct relations. As a result, we have shown that all selection pairs and direct relations in the ground-truth causal graph are identifiable. □

C.2. Proof of Theorem 3.4

Theorem 3.4. *Let the observed data be a large enough sample generated by a model defined in Section 2. In addition to the faithfulness assumption and Markov condition, suppose the following assumptions hold:*

i. (No future influencing the past): For any edge between observed variables, e.g., $X_i \rightarrow X_j$, $i < j$.

ii. (Exclude degenerate structures): Structural conditions 3.1 and 3.3 are satisfied.

Then all selection pairs, direct relations, and confounded pairs in the causal graph $\mathcal{G}$ are identifiable.

Proof. We prove the identifiability of the causal graph by showing that the algorithm 2 identifies all selection pairs, direct relations, and confounded pairs. Consistent with the three stages of the algorithm, the proof is divided into the following three stages:

- **Stage One**: Identifying dependent pairs (selection pairs, direct relations, and confounded pairs) together with several spurious structures.
- **Stage Two**: Identifying selection pairs, dependent pairs that are both selection pairs and direct relations, and direct relations, while remaining some spurious direct relations corresponding to unidentified confounded pairs.
- **Stage Three**: Identifying confounded pairs while removing all spurious structures.

Stage One. We first show that, after lines 1-10, the identified sets of variable pairs, denoted as $\mathbf{E}_1$, only consists of the following types of cases:

- All selection pairs $(X_i, X_j)_{\text{s}}$ in the ground-truth causal graph.
- All direct relations $(X_i, X_j)_{\text{x}}$ in the ground-truth causal graph.
- All confounded pairs $(X_i, X_j)_{\text{c}}$ in the ground-truth causal graph.

- All pairs of variables (X_i, X_j) that there exist confounded pairs $(X_{i+1}, X_j)_{\mathrm{c}}$ in the ground-truth causal graph.
- All dependent pairs that are both selection pairs and direct relations, i.e., $(X_i, X_j)_{\{s,x\}}$, in the ground-truth causal graph.
- All pairs of variables (X_i, X_j) that there exists (X_i, X_{j+1}) that is both a selection pair and a direct relation in the ground-truth causal graph.

For a variable pair (X_i, X_j), we discuss all potential cases for the paths between X_i and X_j as follows. We denote X_k as the observed variable closest to X_i.

I: (X_i, X_j) is a dependent pair. In this case, X_i is d-connected with X_j given any conditioning set, thus $X_i \not\perp\!\!\!\perp X_j | \mathbf{X}_{\text{Pre}(j)\setminus i}$ and $(X_i, X_j)_{\mathrm{d}} \in \mathbf{E}_1$. This covers the cases where (X_i, X_j) is a selection pair, a direct relation, a confounded pair, and a dependent pair which is both a selection pair and a direct relation.

II: $\{X_i \cdots \leftarrow X_k \cdots X_j\}$, where $k < i$. Since $X_k \in \mathbf{X}_{\text{Pre}(j)\setminus i}$, this path is d-connected if and only if X_k is a collider on the path. However, X_k cannot be a collider because of $\{X_i \leftarrow X_k\}$. Thus, $X_i \perp\!\!\!\perp X_j | \mathbf{X}_{\text{Pre}(j)\setminus i}$ and $(X_i, X_j)_{\mathrm{d}} \notin \mathbf{E}_1$.

III: $\{X_i \cdots \rightarrow X_k \cdots X_j\}$, where $k < i$. Because of assumption i, $\{X_i \rightarrow X_k\}$ is impossible. Thus, there must be a confounded pair $\{X_k, X_i\}_{\mathrm{c}}$. Because of assumption 3.3, X_k cannot be caused by another confounder or observed variable via higher-order direct relation. Therefore, X_k is not a collider on the path, and we have $X_i \perp\!\!\!\perp X_j | \mathbf{X}_{\text{Pre}(j)\setminus i}$ ($X_k \in \mathbf{X}_{\text{Pre}(j)\setminus i}$) and thus $(X_i, X_j)_{\mathrm{d}} \notin \mathbf{E}_1$.

IV: $\{X_i \cdots \leftrightarrow X_k \cdots X_j\}$, where $k < i$ and $\leftrightarrow$ means both tails and arrow heads exist. Because of assumption i, $\{X_i \rightarrow X_k\}$ is impossible. Thus, there must be a confounded pair $\{X_k, X_i\}_{\mathrm{c}}$. Because of assumption 3.3, (X_k, X_i) cannot be a selection pair or direct relation, and X_k cannot be caused by another confounder or observed variable via higher-order direct relation. We further discuss the following cases:

IV.i: X_k is a collider. For X_k to be a collider, the only case is $\{X_{k-1} \rightarrow X_k \leftarrow C \rightarrow X_i\}$, where C as the corresponding latent confounder in $\mathbf{C}$. In that case, $X_{k-1} \in \mathbf{X}_{\text{Pre}(j)\setminus i}$ cannot be a collider on the path. Thus, we have $X_i \perp\!\!\!\perp X_j | \mathbf{X}_{\text{Pre}(j)\setminus i}$, which implies $(X_i, X_j)_{\mathrm{d}} \notin \mathbf{E}_1$.

IV.ii: X_k is not a collider. Then we have $X_i \perp\!\!\!\perp X_j | \mathbf{X}_{\text{Pre}(j)\setminus i}$ ($X_k \in \mathbf{X}_{\text{Pre}(j)\setminus i}$) and thus $(X_i, X_j)_{\mathrm{d}} \notin \mathbf{E}_1$.

V: $\{X_i \cdots \leftarrow X_k \cdots X_j\}$, where $k > i$. Based on this, we discuss the following cases:

V.i: The case where $i < k < j$. Because of assumption i, $\{X_i \leftarrow X_k\}$ is impossible. Thus, there must be a selection pair $\{X_i, X_k\}_{\mathrm{s}}$, and X_k must not be a collider on the path. Since $\mathbf{X}_{\text{Pre}(j)\setminus i}$ is the condition set and $X_k \in \mathbf{X}_{\text{Pre}(j)\setminus i}$, we have $X_i \perp\!\!\!\perp X_j | \mathbf{X}_{\text{Pre}(j)\setminus i}$ ($X_k \in \mathbf{X}_{\text{Pre}(j)\setminus i}$) and thus $(X_i, X_j)_{\mathrm{d}} \notin \mathbf{E}_1$.

V.ii: The case where $k > j$. Because of assumption i, $\{X_i \leftarrow X_k\}$ is impossible. Thus, there must be a selection pair $\{X_i, X_k\}_{\mathrm{s}}$, and X_k must not be a collider. Since $k > j$, X_k is not in the conditioning set $\mathbf{X}_{\text{Pre}(j)\setminus i}$. However, the index k for any selection pair $\{X_i, X_k\}_{\mathrm{s}}$, where $k > j$, has already been included in $\mathcal{R}[i]$ according to the order of tests in the algorithm (lines 3-10). Therefore, there is the conditional independence $X_i \perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus i}, X_{j+1}, X_k\}$, which implies $(X_i, X_j)_{\mathrm{d}} \notin \mathbf{E}_1$.

VI: $\{X_i \cdots \rightarrow X_k \cdots X_j\}$, where $k > i$. Based on this, we discuss the following cases:

VI.i: The case where $i < k < j$. Since $\mathbf{X}_{\text{Pre}(j)\setminus i}$ is the condition set and $X_k \in \mathbf{X}_{\text{Pre}(j)\setminus i}$, X_k must be a collider for the path to be d-connected. On the path, let us denote the X_p as the observed variable closest to X_k other than X_i. Thus, we must have $\{X_k \leftarrow \cdots X_p\}$ on the path. We further have the following cases:

- The case where $p < k$. If $p < k$, we have $p < i$ as $p \neq i$. Since $\{X_k \leftarrow \cdots X_p\}$, (X_k, X_p) is either a confounded pair or direct relation. We further discuss the following cases:
  - If (X_p, X_k) is a confounded pair, because X_p cannot be caused by another confounder (assumption 3.3), the only case for X_p to also be a collider is $\{X_k \leftarrow C \rightarrow X_p \leftarrow X_{p-1}$. Thus, X_{p-1} must not be a collider on the path, and $X_i \perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus i}\}$, which implies $(X_i, X_j)_{\mathrm{d}} \notin \mathbf{E}_1$.
  - If (X_p, X_k) is a direct relation, X_p must not be a collider. So we have $X_i \perp\!\!\!\perp X_j | \mathbf{X}_{\text{Pre}(j)\setminus i}$, which implies $(X_i, X_j)_{\mathrm{d}} \notin \mathbf{E}_1$.

- The case where $p > k$. In this case, (X_k, X_p) can only be a confounded pair. Because of assumption 3.3, X_k and X_p cannot be caused by another confounder or higher-order direct relation. Therefore, there must be $\{X_i \to X_k\}$ and $k = i + 1$ for the path to be d-connected. We further discuss the following cases:
  - If $k < p < j$, because of assumption 3.3, X_p cannot be caused by another confounder or higher-order direct relation. If X_p is a collider, X_{p-1} must not be a collider on the path as $\{X_{p-1} \to X_p\}$. Thus, we have $X_i \perp\!\!\!\perp X_j | \mathbf{X}_{\text{Pre}(j)\setminus i}$, which implies $(X_i, X_j)_{\text{d}} \notin \mathbf{E}_1$.
  - If $p = j$, then $X_i \not\perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus i}\}$ because of $(X_{i+1}, X_j)_{\text{c}}$, which implies $(X_i, X_j)_{\text{d}} \in \mathbf{E}_1$.
  - If $p > j$ ($X_p \notin \mathbf{X}_{\text{Pre}(j)\setminus i}$), the only case where the path between X_i and X_j is not d-separated by X_p is the case where X_p is the cause of a selection variable. However, the index p has already been included in $R[i]$ according to the order of tests in the algorithm (lines 3-10). According to assumption 3.1, X_p cannot be caused by more than two observed variables via higher-order direct relations. Therefore, the only case for X_i to be conditionally dependent of X_j given $\{\mathbf{X}_{\text{Pre}(j)\setminus i}, X_{j+1}, X_k\}$ for all $k \in \mathcal{R}[i]$ is the structure that $\{X_i \to S \leftarrow X_p\}$ and $p = j + 1$. However, this contradicts with assumption 3.3 since $\{X_{i+1}, X_p\}$ is a confounded pair.

VI.ii: The case where $k > j$. If $k > j$ ($X_k \notin \mathbf{X}_{\text{Pre}(j)\setminus i}$), the only case where the path between X_i and X_j is not d-separated by X_k is the case where X_k is the cause of a selection variable. However, since we have $\{X_i \cdots \to X_k\}$, it is not possible for (X_i, X_k) to be a selection pair. For any other selection pairs, in order to have $X_i \not\perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus i}\}$, (X_j, X_k) must be a selection pair. According to assumption 3.1, we have $k \neq j + 1$. Therefore, the path is d-separated by either k, where $k \in \mathcal{R}[i]$, or $j + 1$. Consequently, there must be the conditional independence $X_i \not\perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus i}, X_{j+1}, X_k\}$ and thus $(X_i, X_j)_{\text{d}} \notin \mathbf{E}_1$.

VII: $\{X_i \cdots \leftrightarrow X_k \cdots X_j\}$, where $k > i$. Based on this, we discuss the following cases:

VII.i: The case where $i < k < j$. Since $\mathbf{X}_{\text{Pre}(j)\setminus i}$ is the condition set and $X_k \in \mathbf{X}_{\text{Pre}(j)\setminus i}$, X_k must be a collider for the path to be d-connected. At the same time, (X_i, X_k) must be a selection pair. We further discuss the following cases:

- The case where $k = i + 1$. According to assumption 3.1, (X_i, X_k) cannot be a selection pair if $k = i + 1$, which is a contradiction.
- The case where $k \neq i + 1$. On the path, let us denote the X_p as the observed variable closest to X_k other than X_i. Thus, we must have $\{X_k \leftarrow \cdots X_p\}$ on the path. We further have the following cases:
  - The case where $p < k$. If $p < k$, we have $p < i$ as $p \neq i$. Since $\{X_k \leftarrow \cdots X_p\}$, (X_p, X_k) is either a confounded pair or a direct relation. If (X_p, X_k) is a confounded pair, because X_p cannot be caused by another confounder (assumption 3.3), the only case for X_p to also be a collider is $\{X_k \leftarrow C \to X_p \leftarrow X_{p-1}\}$. Thus, X_{p-1} must not be a collider, and $X_i \perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus i}\}$, which implies $(X_i, X_j)_{\text{d}} \notin \mathbf{E}_1$; If (X_p, X_k) is a direct relation, X_p must not be a collider, and $X_i \perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus i}\}$, which implies $(X_i, X_j)_{\text{d}} \notin \mathbf{E}_1$.
  - The case where $p > k$. In this case, (X_k, X_p) can only be a confounded pair. However, because $k \neq i + 1$ and $\{X_i \cdots \leftrightarrow X_k\}$, (X_i, X_k) is either a confounded pair or higher-order direct relation, both of which are impossible according to assumption 3.3. Thus, $(X_i, X_j)_{\text{d}} \notin \mathbf{E}_1$.

VII.ii: The case where $k > j$. If $k > j$ ($X_k \notin \mathbf{X}_{\text{Pre}(j)\setminus i}$), the only case where the path between X_i and X_j is not d-separated by X_k is the case where X_k is the cause of a selection variable. However, the index k for any selection pair $\{X_i, X_k\}_{\text{s}}$, where $k > j$, has already been included in $R[i]$ according to the order of tests in the algorithm (lines 3-10). According to assumption 3.1, X_k cannot be caused by more than two observed variables via higher-order direct relations. Therefore, the only case for X_i to be conditionally dependent of X_j given $\{\mathbf{X}_{\text{Pre}(j)\setminus i}, X_{j+1}, X_k\}$ for all $k \in \mathcal{R}[i]$ is the structure that $(X_i, X_{j+1})_{\{s,x\}}$. At the same time, we also have $X_i \not\perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus i}\}$. As a result, we have $(X_i, X_j)_{\text{d}} \in \mathbf{E}_1$ where (X_i, X_{j+1}) is both a selection pair and a direct relation.

Stage Two. Next, we prove that we can identify selection pairs, direct relations, and dependent pairs that are both selection pairs and direct relations in $\mathcal{E}_1$ after lines 11-29 in the algorithm. After this stage, there are still some spurious dependent pairs because of unidentified confounded pairs. Note that, according to the algorithm, for any X_i, the tests start from variables in $\mathcal{R}[i]$ with the largest index and continue in a descending order. In addition, X_k for k after lines 15 and 16 is a variable where (X_i, X_k) is not a pair in $\mathcal{E}_1$.

We discuss the following cases for all types in $\mathcal{E}_1$.

I: (X_i, X_j) is a selection pair but not a direct relation. According to the algorithm, it will be identified as a selection pair but not a direct relation if and only if $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$ (the definition of $\mathbf{T}$ is in line 17 in the algorithm).

We first show that all these selection pairs will be identified. If (X_i, X_j) is a selection pair but not a direct relation, according to assumption 3.3, if (X_i, X_j) is a selection pair, it cannot be a confounded pair. Thus, X_j must be a non-collider on the path, which may lead to the d-separation and thus the conditional independence. Thus, we need to show that there exists no other path that leads to conditional dependence. We prove it by contradiction.

Suppose that there exist a path other than $\{X_i \to S \leftarrow X_j \cdots X_k\}$ that makes $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$. We discuss all potential cases for that path:

I.i: $\{X_i \cdots \leftarrow X_p \cdots X_k\}$, where $p < i$ and X_p denotes the observed variable that is the closest to X_i on the path. Since $X_p \in \mathbf{X}_{\text{Pre}(k)\setminus i}$, this path is d-connected if and only if X_p is a collider on the path. However, X_p cannot be a collider because of $\{X_i \leftarrow X_p\}$. Thus, this case is impossible.

I.ii: $\{X_i \cdots \to X_p \cdots X_k\}$, where $p < i$ and X_p denotes the observed variable that is the closest to X_i on the path. Because of assumption i, $\{X_i \to X_p\}$ is impossible. Thus, there must be a confounded pair $\{X_p, X_i\}_{\text{c}}$. Because of assumption 3.3, X_p cannot be caused by another confounder or observed variable via higher-order direct relation. Therefore, X_p is not a collider and this case is impossible

I.iii: $\{X_i \cdots \leftrightarrow X_p \cdots X_k\}$, where $p < i$ and X_p denotes the observed variable that is the closest to X_i on the path. Because of assumption i, $\{X_p \to X_i\}$ is impossible. Thus, there must be a confounded pair $\{X_p, X_i\}_{\text{c}}$. Because of assumption 3.3, (X_p, X_i) cannot be a selection pair or direct relation. Thus, it is impossible to have $\{X_i \cdots \leftrightarrow X_p\}$ on the path. Consequently, this case is impossible.

I.iv: $\{X_i \cdots \leftarrow X_p \cdots X_k\}$, where $i < p < k$ and X_p denotes the observed variable that is the closest to X_i on the path. Because of assumption i, $\{X_i \leftarrow X_p\}$ is impossible. Thus, there must be a selection pair $(X_i, X_p)_{\text{s}}$, and X_p must not be a collider. Since $\{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$ is the condition set and $X_p \in \mathbf{X}_{\text{Pre}(k)\setminus i}$, we have $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$ ($X_p \in \mathbf{X}_{\text{Pre}(k)\setminus i}$) and thus this case is impossible.

I.v: $\{X_i \cdots \to X_p \cdots X_k\}$, where $i < p < k$. Since $\mathbf{X}_{\text{Pre}(k)\setminus i}$ is in the condition set and $X_p \in \mathbf{X}_{\text{Pre}(k)\setminus i}$, X_p must be a collider for the path to be d-connected. On the path, let us denote the X_q as the observed variable closest to X_p other than X_i. Thus, we must have $\{X_p \leftarrow \cdots X_q\}$ on the path. We further have the following cases:

- The case where $q < p$. If $q < p$, we have $q < i$ as $q \neq i$. Since $\{X_p \leftarrow \cdots X_q\}$, (X_q, X_p) is either a confounded pair or direct relation. We further discuss the following cases:
  - If (X_q, X_p) is a confounded pair, because X_q cannot be caused by another confounder (assumption 3.3), the only case for X_q to also be a collider is $\{X_p \leftarrow C \to X_q \leftarrow X_{q-1}$. Thus, X_{q-1} must not be a collider, and $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$ ($X_{q-1} \in \mathbf{X}_{\text{Pre}(k)\setminus i}$) and thus this case is impossible.
  - If (X_q, X_p) is a direct relation, X_q must not be a collider, and $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$ ($X_q \in \mathbf{X}_{\text{Pre}(k)\setminus i}$) and thus this case is impossible.
- The case where $q > p$. In this case, (X_p, X_q) can only be a confounded pair. Because of assumption 3.3, X_p and X_q cannot be caused by another confounder or higher-order direct relation. Therefore, there must be $\{X_i \to X_p\}$ and $p = i + 1$ for the path to be d-connected. We further discuss the following cases:
  - If $p < q < k$, because of assumption 3.3, X_q cannot be caused by another confounder or higher-order direct relation. If X_q is a collider, X_{q-1} must not be a collider on the path as $\{X_{q-1} \to X_q\}$, and $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$ ($X_{q-1} \in \mathbf{X}_{\text{Pre}(k)\setminus i}$) and thus this case is impossible.
  - If $q = k$, then we have $(X_{i+1}, X_k)_{\text{c}}$, which is a contradiction to the definition of k ($k \in \mathcal{R}[i]$ according to the algorithm)
  - If $k < q < j$, because of assumption 3.3, $(X_{p-1}, X_{q+1})_{d \neq c}$ should not exist. However, since $p - 1 = i$, this contradicts the definition of k ($k \in \mathcal{R}[i]$ according to the algorithm) or the existing selection pair $(X_i, X_j)_{\text{s}}$.
  - If $q = j$, this contradicts assumption 3.3 that there exists no $(X_i, X_j)_{\text{s}}$ ($i = p - 1$).
  - If $q > j$, if not conditioning on X_q, the only case where the path between X_i and X_k is not d-separated by X_q is the case where X_q is the cause of a selection variable. However, the index q has already been included in $R[i]$ according to the order of tests in the algorithm (Lines 12 and 13). According to assumption 3.1, X_q cannot be caused by more than two observed variables via higher-order direct relations. As a result, we have $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$ and thus this case is impossible.

I.vi: $\{X_i \cdots \leftrightarrow X_p \cdots X_k\}$, where $i < p < k$. Since $\mathbf{X}_{\text{Pre}(k)\setminus i}$ is in the condition set and $X_p \in \mathbf{X}_{\text{Pre}(k)\setminus i}$, X_p must be a collider for the path to be d-connected. At the same time, (X_i, X_p) must be a selection pair. We further discuss the following cases:

- The case where $p = i + 1$. According to assumption 3.1, (X_i, X_p) cannot be a selection pair if $p = i + 1$, which is a contradiction.
- The case where $p \neq i + 1$. On the path, let us denote the X_q as the observed variable closest to X_p other than X_i. Thus, we must have $\{X_p \leftarrow \cdots X_q\}$ on the path. We further have the following cases:
  - The case where $q < p$. If $q < p$, we have $q < i$ as $q \neq i$. Since $\{X_p \leftarrow \cdots X_q\}$, (X_q, X_p) is either a confounded pair or direct relation. If (X_q, X_p) is a confounded pair, because X_q cannot be caused by another confounder (assumption 3.3), the only case for X_q to also be a collider is $\{X_p \leftarrow C \rightarrow X_q \leftarrow X_{q-1}\}$. Thus, X_{q-1} must not be a collider, and the path is not d-connected; If (X_q, X_p) is a direct relation, X_q must not be a collider, and the path is, again, not d-connected. Thus, this case is impossible.
  - The case where $q > p$. In this case, (X_p, X_q) can only be a confounded pair. However, because $p \neq i + 1$ and $\{X_i \cdots \leftrightarrow X_p\}$, (X_i, X_p) is either a confounded pair or higher-order direct relation, both of which are impossible according to assumption 3.3. Thus, this case is impossible.

I.vii: $\{X_i \cdots \leftarrow X_p\}$, where $k = p$. This is impossible because of the algorithm, specifically, how k is selected (lines 15 and 16).

I.viii: $\{X_i \cdots \rightarrow X_p\}$, where $k = p$. This is also impossible because of the algorithm.

I.ix: $\{X_i \cdots \leftrightarrow X_p\}$, where $k = p$. Similarly, this is impossible according to the algorithm.

I.x: $\{X_i \cdots \leftarrow X_p \cdots X_k\}$, $k < p < j$. Because of assumption i, $\{X_i \leftarrow X_p\}$ is impossible. Thus, there must be a selection pair $(X_i, X_p)_{\mathrm{s}}$. Together with the search procedure (lines 20-25), it is thus impossible for the selected k to be smaller than p, which is a contradiction.

I.xi: $\{X_i \cdots \rightarrow X_p \cdots X_k\}$, $k < p < j$. Because of $\{X_i \cdots \rightarrow X_p\}$, (X_i, X_p) is either a higher-order direct relation or a confounded pair. Since X_p is not in the condition set, (X_i, X_p) must also be a selection pair for the path to be d-connected. However, according to assumption 3.1, there exits no $(X_i, X_{p+1})_{\mathrm{d}}$. This contradicts the search procedure to select k (lines 15 and 16), and thus this case is impossible.

I.xii: $\{X_i \cdots \leftrightarrow X_p \cdots X_k\}$, $k < p < j$. Because of assumption i, $\{X_i \leftarrow X_p\}$ is impossible. Thus, there must be a selection pair $(X_i, X_p)_{\mathrm{s}}$. Similarly, it is thus impossible for the selected k to be smaller than p, which is a contradiction.

I.xiii: $\{X_i \cdots \leftarrow X_p \cdots X_k\}$, $j < p$. Because of assumption i, $\{X_i \leftarrow X_p\}$ is impossible. Thus, there must be a selection pair $(X_i, X_p)_{\mathrm{s}}$. Moreover, according to assumption 3.1, X_p cannot be the effect of multiple higher-order directions. Thus, X_p must not be a collider on the path, so we can d-separate the path by conditioning on it. Since X_p is already in $\mathcal{T}$ according to the algorithm, we have $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\mathrm{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$ and thus this case is impossible.

I.xiv: $\{X_i \cdots \rightarrow X_p \cdots X_k\}$, $j < p$. In this case, (X_i, X_p) is either a higher-order direct relation or a confounded pair. We further discuss the following case:

- X_p is not a cause of a selection variable. In this case, X_p must be a collider on the path, and the path is d-separated when not conditioning on X_p. So we have $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\mathrm{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$, which indicates that this case is impossible.
- X_p is a cause of a selection variable. According to assumption 3.1, it cannot be the effect of multiple higher-order directions. Therefore, if X_p is the cause of another selection variable together with any observed variable with index $u \in [k, j)$, the path is always d-separated by conditioning on both X_p and X_j. At the same time, if X_p is the cause of another selection variable together with any observed variable with an index larger than j, that variable will also be in the condition set ($\mathbf{T}$) according to the algorithm (line 17). Recursively, all these variables will be in the condition set, and the path will be d-separated, which indicates that this case is impossible.

I.xv: $\{X_i \cdots \leftrightarrow X_p \cdots X_k\}$, $j < p$. Because of assumption i, $\{X_i \leftarrow X_p\}$ is impossible. Thus, there must be a selection pair $(X_i, X_p)_{\mathrm{s}}$. At the same time, since (X_i, X_p) cannot be both a selection pair and confounded pair, we have $(X_i, X_p)_{\mathrm{d}\neq\mathrm{c}}$, i.e., it is both a selection pair and a direct relation. According to assumption 3.1, it cannot be the effect of multiple higher-order directions. Therefore, if X_p is the cause of another selection variable together with any observed variable with index $u \in [k, j)$, the path is always d-separated by conditioning on both X_p and X_j. At the same time, if X_p is the cause of another selection variable together with any observed variable with an index larger than j, that variable will also be in the condition set ($\mathbf{T}$) according to the algorithm (line 17). Recursively, all these variables will be in the condition set, and the path will be d-separated, which indicates that this case is impossible.

Therefore, there does not exist any path other than $\{X_i \to S \leftarrow X_j \cdots X_k\}$ that makes $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$. So all selection pairs $(X_i, X_j)_{\text{s}}$ are identified.

We then discuss the other cases in $\mathcal{E}_1$ and show that it is impossible for any other types of variable pairs in $\mathcal{E}_1$ to have the test results that $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$.

- (X_i, X_j) is a direct relation but not a selection pair. In this case, since X_j is collider on the path $\{X_i \cdots \to X_j\}$, for any $v \in \mathbf{T}$, there is always $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$. Thus, this case is impossible.
- (X_i, X_j) is both a selection pair and a direct relation. In this case, since X_j is still a collider on the path $\{X_i \cdots \to X_j\}$, for any $v \in \mathbf{T}$, there is always $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$, which indicates that this case is impossible.
- (X_i, X_j) is a confounded pair. In this case, X_j is a collider on the path (all confounders are latent and thus will never be conditioned). Similar to the previous two cases, for any $v \in \mathbf{T}$, there is always $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$, and this case is impossible.
- (X_i, X_j) is a spurious dependent pair that there exists a confounded pair $(X_{i+1}, X_j)_{\text{c}}$ in the ground-truth causal graph. In this case, X_{i+1} is a collider on the path, and thus conditioning on it does not d-separate the path. At the same time, X_j is also a collider. Since both X_{i+1} and X_j are in the condition set, there is always $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$, which indicates that this case is impossible.
- (X_i, X_j) is a spurious dependent pair that there exists (X_i, X_{j+1}) that is both a selection pair and a direct relation in the ground-truth causal graph. In this case, (X_i, X_j) is not a dependent pair according to assumption 3.1, and will be removed by the algorithm (line 26). Therefore, this case is impossible.

II: (X_i, X_j) is a direct relation but not a selection pair. We first show that all these variable pairs will be identified. In this case, X_j is a collider on the path. Thus, we have $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$. However, when X_j is not in the condition set, we may have the d-separation and thus the conditional independence $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$ (the definition of $\mathbf{T}$ is in line 17 in the algorithm). Thus, we need to show that there exists no other path that leads to conditional dependence. We prove it by contradiction.

Suppose that there exist a path other than $\{X_i \to X_j \cdots X_k\}$ that makes $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$. Denote X_p as the observed variable that is the closest to X_i on the path, we discuss all potential cases for that path:

II.i: The discussions for the following cases are similar to those in the cases where (X_i, X_j) is a selection pair but not a direct relation:

- $\{X_i \cdots \leftarrow X_p \cdots X_k\}$, where $p < i$, $i < p < k$, or $k < p < j$.
- $\{X_i \cdots \to X_p \cdots X_k\}$, where $p < i$ or $k < p < j$.
- $\{X_i \cdots \leftrightarrow X_p \cdots X_k\}$, where $p < i$, $i < p < k$, or $k < p < j$.
- $\{X_i \cdot X_p\}$, where $p = k$.

Thus, we mainly focus on the other cases as follows.

II.ii: $\{X_i \cdots \to X_p \cdots X_k\}$, where $i < p < k$. Since $\mathbf{X}_{\text{Pre}(k)\setminus i}$ is in the condition set and $X_p \in \mathbf{X}_{\text{Pre}(k)\setminus i}$, X_p must be a collider for the path to be d-connected. On the path, let us denote the X_q as the observed variable closest to X_p other than X_i. Thus, we must have $\{X_p \leftarrow \cdots X_q\}$ on the path. We further have the following cases:

- The case where $q < p$. If $q < p$, we have $q < i$ as $q \neq i$. Since $\{X_p \leftarrow \cdots X_q\}$, (X_p, X_q) is either a confounded pair or direct relation. We further discuss the following cases:
  - If (X_p, X_q) is a confounded pair, because X_q cannot be caused by another confounder (assumption 3.3), the only case for X_q to also be a collider is $\{X_p \leftarrow C \to X_q \leftarrow X_{q-1}$. Thus, X_{q-1} must not be a collider, and $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, \mathbf{T}\}$ ($X_{q-1} \in \mathbf{X}_{\text{Pre}(k)\setminus i}$) and thus this case is impossible.
  - If (X_p, X_q) is a direct relation, X_q must not be a collider, and $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, \mathbf{T}\}$ ($X_{q-1} \in \mathbf{X}_{\text{Pre}(k)\setminus i}$) and thus this case is impossible.
- The case where $q > p$. In this case, (X_p, X_q) can only be a confounded pair. Because of assumption 3.3, X_p and X_q cannot be caused by another confounder or higher-order direct relation. Therefore, there must be $\{X_i \to X_p\}$ and $p = i + 1$ for the path to be d-connected. We further discuss the following cases:
  - If $p < q < k$, because of assumption 3.3, X_q cannot be caused by another confounder or higher-order direct relation. If X_q is a collider, X_{q-1} must not be a collider on the path as $\{X_{q-1} \to X_q\}$, and $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, \mathbf{T}\}$ ($X_{q-1} \in \mathbf{X}_{\text{Pre}(k)\setminus i}$) and thus this case is impossible.

- If $q = k$, then we have $(X_{i+1}, X_k)_{\mathrm{c}}$, which is a contradiction to the definition of k ($k \in \mathcal{R}[i]$ according to the algorithm)
- If $k < q < j$, because of assumption 3.3, $(X_{p-1}, X_{q+1})_{d \neq c}$ should not exist. However, since $p - 1 = i$, this contradicts the definition of k ($k \in \mathcal{R}[i]$ according to the algorithm) or the existing selection pair $(X_i, X_j)_{\mathrm{s}}$.
- If $q = j$, this contradicts assumption 3.3 that there exists no $(X_i, X_j)_{\mathrm{s}}$ ($i = p - 1$).
- If $q > j$, if not conditioning on X_q, the only case where the path between X_i and X_k is not d-separated by X_q is the case where X_q is the cause of a selection variable. However, the index q has already been included in $R[i]$ according to the order of tests in the algorithm (lines 12 and 13). According to assumption 3.1, X_q cannot be caused by more than two observed variables via higher-order direct relations. At the same time, it is impossible to have $q = j + 1$, since there should not exist $(X_i, X_j)_{\mathrm{d}}$ if (X_{i+1}, X_{j+1}) is a confounded pair according to assumption ii. Therefore, the path is d-separated by conditioning on $\{\mathbf{X}_{\mathrm{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$. As a result, we have $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\mathrm{Pre}(k)\setminus i}, \mathbf{T}\}$ and thus this case is impossible.

II.iii: $\{X_i \cdots \leftarrow X_p \cdots X_k\}$, $j < p$. Because of assumption 3.1, $\{X_i \leftarrow X_p\}$ is impossible. Thus, there must be a selection pair $(X_i, X_p)_{\mathrm{s}}$. Moreover, according to assumption 3.1, X_p cannot be the effect of multiple higher-order directions. Thus, X_p must not be a collider on the path, so we can d-separate the path by conditioning on it. Since X_p is already in $\mathbf{T}$ according to the algorithm, we have $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\mathrm{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$ and thus this case is impossible.

II.iv: $\{X_i \cdots \rightarrow X_p \cdots X_k\}$, $j < p$. In this case, (X_i, X_p) is either a higher-order direct relation or a confounded pair. We further discuss the following case:

- X_p is not a cause of a selection variable. In this case, X_p must be a collider on the path, and the path is d-separated when not conditioning on X_p. So we have $X_i \perp\!\!\!\perp X_k | \{\mathbf{X}_{\mathrm{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$, which indicates that this case is impossible.
- X_p is a cause of a selection variable. According to assumption 3.1, it cannot be the effect of multiple higher-order directions. Meanwhile, the index p has already been included in $R[i]$ according to the order of tests in the algorithm (lines 12 and 13). At the same time, it is impossible for p to be $j + 1$ according to assumption 3.1. This is because if X_{j+1} is a cause of a selection variable while (X_i, X_{j+1}) is either a higher-order direct relation or a confounded pair, it is impossible to have $(X_i, X_j)_{\mathrm{d}}$. Therefore, the path is always d-separated by conditioning on $\{\mathbf{X}_{\mathrm{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$ and thus this case is impossible.

II.v: $\{X_i \cdots \leftrightarrow X_p \cdots X_k\}$, $j < p$. Because of assumption i, $\{X_i \leftarrow X_p\}$ is impossible. Thus, there must be a selection pair $(X_i, X_p)_{\mathrm{s}}$. At the same time, since (X_i, X_p) cannot be both a selection pair and confounded pair, we have $(X_i, X_p)_{\mathrm{d}\neq\mathrm{c}}$, i.e., it is both a selection pair and a direct relation. According to assumption 3.1, it cannot be the effect of multiple higher-order directions. Meanwhile, it is impossible for p to be $j + 1$ according to assumption 3.1. This is because if X_{j+1} is a cause of a selection variable while (X_i, X_{j+1}) is either a higher-order direct relation or a confounded pair, it is impossible to have $(X_i, X_j)_{\mathrm{d}}$. Therefore, the path is always d-separated by conditioning on $\{\mathbf{X}_{\mathrm{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$ and thus this case is impossible.

We then discuss the other cases in $\mathcal{E}_1$ and show that, except for those spurious direct relations corresponding to unidentified confounded pairs, it is impossible for any other types of variable pairs in $\mathcal{E}_1$ to be identified as a direct relation.

- (X_i, X_j) is a selection pair but not a directed relation. As already shown before, it will always be identified as a selection pair but not a direct relation. Thus, this case is impossible
- (X_i, X_j) is both a selection pair and a direct relation. In this case, since X_j is still a collider on the path $\{X_i \cdots \rightarrow X_k\}$, there is always $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\mathrm{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$. At the same time, since (X_i, X_j) is also a selection pair, not conditioning on X_j will also open the path. Thus, we also have $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\mathrm{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$, which implies that this variable pair will not just be identified as a direct relation according to the algorithm (lines 21-24).
- (X_i, X_j) is a confounded pair. In this case, X_j is always a collider on the path, and thus it is similar to the case where (X_i, X_j) is a direct relation. According to the discussion above, it will also be identified as a direct relation, and thus there will be spurious direct relations corresponding to unidentified confounded pairs.
- (X_i, X_j) is a spurious dependent pair that there exists a confounded pair $(X_{i+1}, X_j)_{\mathrm{c}}$ in the ground-truth causal graph. In this case, X_{i+1} is a collider on the path, and thus conditioning on it does not d-separate the path. At the same time, X_j is also a collider. Since X_{i+1} is always in the condition set, X_j is always a collider on the path. Thus, similar to the previous discussion, (X_i, X_j) will be identified as a spurious direct relation corresponding to the unidentified confounded pair.

- (X_i, X_j) is a spurious dependent pair that there exists (X_i, X_{j+1}) that is both a selection pair and a direct relation in the ground-truth causal graph. In this case, (X_i, X_j) is not a dependent pair according to assumption 3.1, and will be removed by the algorithm (line 26). Therefore, this case is impossible.

III: (X_i, X_j) is both a selection pair and a direct relation. In this case, since X_j is still a collider on the path $\{X_i \cdots \rightarrow X_k\}$, there is always $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_j, \mathbf{T}\}$. At the same time, since (X_i, X_j) is also a selection pair, not conditioning on X_j will also open the path. Thus, we also have $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$. We then discuss the other cases in $\mathcal{E}_1$ and show that, except for those variable pairs that are both a selection pair and a direct relation, it is impossible for any other types of variable pairs in $\mathcal{E}_1$ to be identified as it.

- (X_i, X_j) is a selection pair but not a directed relation. As already shown before, it will always be identified as a selection pair but not a direct relation. Thus, this case is impossible
- (X_i, X_j) is a direct relation but not a selection pair. In this case, according to the discussion before, there is always the conditional independence $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$. Thus, it is impossible for this case to be identified as a variable pair i.e. both a selection pair and a direct relation.
- (X_i, X_j) is a confounded pair. In this case, X_j is always a collider on the path, and thus it is similar to the case where (X_i, X_j) is a direct relation. According to the discussion above, there is always the conditional independence $X_i \not\perp\!\!\!\perp X_k | \{\mathbf{X}_{\text{Pre}(k)\setminus i}, X_{j+1}, \mathbf{T}\}$. Thus, this case is impossible
- (X_i, X_j) is a spurious dependent pair that there exists a confounded pair $(X_{i+1}, X_j)_{\text{c}}$ in the ground-truth causal graph. In this case, X_{i+1} is a collider on the path, and thus conditioning on it does not d-separate the path. At the same time, X_j is also a collider. Since X_{i+1} is always in the condition set, X_j is always a collider on the path. Thus, similar to the previous discussion, (X_i, X_j) will be identified as a spurious direct relation corresponding to the unidentified confounded pair.
- (X_i, X_j) is a spurious dependent pair that there exists (X_i, X_{j+1}) that is both a selection pair and a direct relation in the ground-truth causal graph. In this case, (X_i, X_j) is not a dependent pair according to assumption 3.1, and will be removed by the algorithm (line 26). Therefore, this case is impossible.

IV: (X_i, X_j) is a confounded pair. In this case, X_j is always a collider on the path, and X_i is always a non-collider on the path i.e in the condition set for all tests in the algorithm, Thus, it is similar to the case where (X_i, X_j) is a direct relation. According to the discussion above, it will also be identified as a direct relation, and thus there will be spurious direct relations corresponding to unidentified confounded pairs.

V: (X_i, X_j) is a spurious dependent pair that there exists a confounded pair $(X_{i+1}, X_j)_{\text{c}}$ in the ground-truth causal graph. In this case, X_{i+1} is a collider on the path, and thus conditioning on it does not d-separate the path. At the same time, X_j is also a collider. Since X_{i+1} is always in the condition set, X_j is always a collider on the path. Thus, similar to the previous discussion, (X_i, X_j) will be identified as a spurious direct relation corresponding to the unidentified confounded pair.

VI: (X_i, X_j) is a spurious dependent pair that there exists (X_i, X_{j+1}) that is both a selection pair and a direct relation in the ground-truth causal graph. In this case, (X_i, X_j) is not a dependent pair according to assumption 3.1, and will be removed by the algorithm (line 26).

Therefore, after Stage Two, we can identify selection pairs, dependent pairs that are both selection pairs and direct relations, and direct relations, while maintaining some spurious direct relations corresponding to unidentified confounded pairs.

Stage Three. Finally, we need to identify confounded pairs and remove the spurious direct relations caused by the previously unidentified confounded pairs. After the second stage, the only cases where (X_i, X_j) can be a confounded pair is where we have identified both (X_{i-1}, X_j) and (X_i, X_j) as direct relations. Based on the test order in the algorithm (lines 31 and 32), if (X_i, X_j) is a spurious dependent pair caused by the confounded pair (X_i, X_{j+1}), it would be removed before the test for itself. Thus, we discuss the following potential cases:

I: (X_{i-1}, X_j) is a direct relation in the ground-truth, and (X_i, X_j) is a confounded pair in the ground-truth. This case is impossible according to assumption 3.3.

II: (X_{i-1}, X_j) is a confounded pair in the ground-truth, and (X_i, X_j) is a direct relation in the ground-truth. This case is impossible according to assumption 3.3.

III: Both (X_{i-1}, X_j) and (X_i, X_j) are direct relations in the ground-truth. In this case, we always have X_{i-1} dependent of X_j given any condition set. Thus, these two direct relations will be identified by the algorithm,

IV: Both (X_{i-1}, X_j) and (X_i, X_j) are confounded pairs in the ground-truth. This case is impossible according to assumption 3.3.

V: (X_{i-1}, X_j) is not a dependent pair in the ground-truth, and (X_i, X_j) is a confounded pair in the ground-truth. In this case, X_i is a collider on the path between X_{i-1} and X_j. So if there are no other d-connected paths between them, we will have $X_{i-1} \perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}, X_{j+1}, \mathbf{T}\}$.

Suppose there exists another path other than $\{X_{i-1} \to X_i \leftarrow C \to X_j\}$ that makes $X_{i-1} \not\perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}\}$ or $X_{i-1} \not\perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}, \mathbf{T}\}$. Denote X_p as the observed variable that is the closest to X_{i-1} on the path, we discuss all potential cases for that path:

V.i: $\{X_{i-1} \cdots \leftarrow X_p \cdots X_j\}$, where $p < i-1$. In this case, it is impossible for X_p to be a collider on the path. Since $X_p \in \mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}$, there is $X_{i-1} \perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}, X_{j+1}, \mathbf{T}\}$. Thus, this case is impossible.

V.ii: $\{X_{i-1} \cdots \to X_p \cdots X_j\}$, where $p < i-1$. In this case, according to assumption i, it is impossible to have $\{X_{i-1} \to X_p\}$. Thus, there must be a confounded pair $(X_p, X_{i-1})_{\text{c}}$. According to assumption 3.3, X_p cannot be caused by another latent confounder or observed variable via higher-order direct relation. Therefore, if this path is d-connected, we must have $X_{p-1} \to X_p$ on the path. However, if this is the case, X_{p-1} cannot be the collider on the path. Because both X_p and X_{p-1} are in the condition set, we have $X_{i-1} \perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}, X_{j+1}, \mathbf{T}\}$. Thus, this case is impossible.

V.iii: $\{X_{i-1} \cdots \leftrightarrow X_p \cdots X_j\}$, where $p < i-1$. In this case, according to assumption i, it is impossible to have $\{X_{i-1} \to X_p\}$. Thus, there must be a confounded pair $(X_p, X_{i-1})_{\text{c}}$. Then, because of assumption 3.3, (X_p, X_{i-1}) cannot be a selection pair or direct relation. Thus, it is impossible to have $\{X_{i-1} \cdots \leftrightarrow X_p\}$ on the path. Consequently, this case is impossible.

V.iv: $\{X_{i-1} \cdots \leftarrow X_p \cdots X_j\}$, where $i < p < j$. Because of assumption i, $\{X_{i-1} \leftarrow X_p\}$ is impossible. Thus, there must be a selection pair $(X_{i-1}, X_p)_{\text{s}}$, and X_p must not be a collider. Since $\{\mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}\}$ is the condition set and $X_p \in \mathbf{X}_{\text{Pre}(k)\setminus\{i-1,i\}}$, we have $X_{i-1} \perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}, X_{j+1}, \mathbf{T}\}$ and thus this case is impossible.

V.v: $\{X_{i-1} \cdots \to X_p \cdots X_j\}$, where $i < p < j$. Since $\mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}$ is in the condition set and $X_p \in \mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}$, X_p must be a collider on the path for the path to be d-connected. In addition, because of assumption 3.3, there must be $p \neq i+1$. On the path, let us denote the X_q as the observed variable closest to X_p other than X_i. Thus, we must have $\{X_p \leftarrow \cdots X_q\}$ on the path. We further have the following cases:

- The case where $q < p$. Since $\{X_p \leftarrow \cdots X_q\}$, (X_q, X_p) is either a confounded pair or direct relation. Since X_i is already involved in a confounded pair, we have $q \neq i$ according to assumption 3.3. Since there is also $X_q \neq i-1$, we have $q < i-1$. We further discuss the following cases:
  - If (X_q, X_p) is a confounded pair, because X_q cannot be caused by another confounder (assumption 3.3), the only case for X_q to also be a collider is $\{X_p \leftarrow C \to X_q \leftarrow X_{q-1}$. Thus, X_{q-1} must not be a collider on the path, and $X_{i-1} \perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}, X_{j+1}, \mathbf{T}\}$ ($X_{q-1} \in \mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}$) and thus this case is impossible.
  - If (X_q, X_p) is a direct relation, X_q must not be a collider on the path, and $X_{i-1} \perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}, X_{j+1}, \mathbf{T}\}$ ($X_q \in \mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}$) and thus this case is impossible.
- The case where $q > p$. In this case, (X_p, X_q) can only be a confounded pair. Because of assumption 3.3, X_p and X_q cannot be in another confounded pair or higher-order direct relation. Since $p > i+1$, it is impossible to have $(X_{i-1}, X_p)_{\text{d}}$ or $(X_i, X_p)_{\text{d}}$. Thus X_p cannot be a collider on the path, which implies $X_{i-1} \perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}, X_{j+1}, \mathbf{T}\}$ ($X_p \in \mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}$). As a result, this case is impossible.

V.vi: $\{X_{i-1} \cdots \leftrightarrow X_p \cdots X_j\}$, where $i < p < j$. Since $\mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}$ is in the condition set and $X_p \in \mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}$, X_p must be a collider on the path for the path to be d-connected. In addition, because of assumption 3.3, there must be $p \neq i+1$. So we have $p > i+1$. Because there is $\{X_{i-1} \cdots \leftrightarrow X_p\}$, (X_{i-1}, X_p) must be a selection pair. As a result, (X_{i-1}, X_p) cannot be a confounded pair and thus must also be a direct relation. According to assumption 3.1, X_p cannot be the effect of more than one higher-order direct relation. Therefore, another direct relation can only be $(X_{p-1}, X_p)_{\text{d}}$. Then X_{p-1} must not be a collider on the path (note that $p-1 \neq i$), and thus this case is impossible.

V.vii: $\{X_{i-1} \cdots \leftarrow X_p \cdots X_j\}$, where $j < p$. Because of assumption i, $\{X_{i-1} \leftarrow X_p\}$ is impossible. Thus, there must be a selection pair $(X_{i-1}, X_p)_{\text{s}}$. Moreover, according to assumption 3.1, X_p cannot be the effect of more than

one higher-order direct relation. Thus, X_p must not be a collider on the path, so we can d-separate the path by conditioning on it. Since X_p is already in $\mathcal{T}$ according to the algorithm, we have $X_{i-1} \perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}, \mathbf{T}\}$ and thus this case is impossible.

V.viii: $\{X_{i-1} \cdots \rightarrow X_p \cdots X_j\}$, where $j < p$. In this case, (X_{i-1}, X_p) is either a higher-order direct relation or a confounded pair. We further discuss the following case:

- X_p is not a cause of a selection variable. In this case, X_p must be a collider on the path, and the path is d-separated when not conditioning on X_p. So we have $X_{i-1} \perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}, \mathbf{T}\}$, which indicates that this case is impossible.
- X_p is a cause of a selection variable. According to assumption 3.1, it cannot be the effect of multiple higher-order directions. Therefore, if X_p is the cause of another selection variable, the path will be d-separated by conditioning on both X_p and X_{j+1}. Note that conditioning on X_{j+1} does not make the path d-connected, since X_{j+1} cannot be d-connected with X_{i-1}, X_i, or X_{i+1} via a higher-order direct relation or a confounded pair according to assumption 3.3. Thus, there is always $X_{i-1} \perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}, \mathbf{T}\}$, which implies that this case is impossible.

V.ix: $\{X_{i-1} \cdots \leftrightarrow X_p \cdots X_j\}$, where $j < p$. Because of assumption i, $\{X_{i-1} \leftarrow X_p\}$ is impossible. Thus, there must be a selection pair $(X_{i-1}, X_p)_{\text{s}}$. At the same time, since (X_{i-1}, X_p) cannot be both a selection pair and confounded pair, we have $(X_{i-1}, X_p)_{\text{d}\neq\text{c}}$, i.e., it is both a selection pair and a direct relation. According to assumption 3.1, it cannot be the effect of multiple higher-order directions. Therefore, if X_p is the cause of another selection variable, the path will be d-separated by conditioning on both X_p and X_{j+1}. Note that conditioning on X_{j+1} does not make the path d-connected, since X_{j+1} cannot be d-connected with X_{i-1}, X_i, or X_{i+1} via a higher-order direct relation or a confounded pair according to assumption 3.3. Thus, there is always $X_{i-1} \perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}, \mathbf{T}\}$, which implies that this case is impossible.

V.x: $\{X_{i-1} \cdots \leftarrow X_p \cdots X_j\}$, where $p = i$. In this case, according to assumption i, it is impossible to have $\{X_{i-1} \rightarrow X_p\}$. Thus, there must be a confounded pair $(X_p, X_{i-1})_{\text{c}}$. However, since $p = i$ and $(X_i, X_j)_{\text{c}}$, X_p cannot be caused by another latent confounder. Thus, this case is impossible.

V.xi: $\{X_{i-1} \cdots \rightarrow X_p \cdots X_j\}$, where $p = i$. In this case, let us denote X_q as the closest observed variable to X_i (X_p) on the path other than X_{i-1}. Since both X_{i-1} and X_i are not in the condition set, the discussion on the potential cases for X_q is similar to that for X_p, except for the case where we have $\{X_{i-1} \rightarrow X_i \rightarrow X_q \cdots X_j\}$, where $q = i + 1$.

If $q = i + 1$, the path is $\{X_{i-1} \rightarrow X_i \rightarrow X_{i+1} \cdots X_j\}$. On this path, if X_{i+1} is caused by a latent confounder, this path can be d-connected even with X_{i+1} in the condition set. However, since $\mathbf{T}$ consists of all variables where $(X_{i-1}, X_v)_{\text{s}}$, $(X_i, X_v)_{\text{s}}$, or $(X_{i+1}, X_v)_{\text{s}}$ for $v > j$, following the previous discussions, we still have $X_{i-1} \perp\!\!\!\perp X_j | \{\mathbf{X}_{\text{Pre}(j)\setminus\{i-1,i\}}, \mathbf{T}\}$. Thus, this case is impossible.

V.xii: $\{X_{i-1} \cdots \leftrightarrow X_p \cdots X_j\}$, where $p = i$. In this case, according to assumption i, it is impossible to have $\{X_{i-1} \rightarrow X_p\}$. Thus, there must be a confounded pair $(X_p, X_{i-1})_{\text{c}}$. However, since $p = i$ and $(X_i, X_j)_{\text{c}}$, X_p cannot be caused by another latent confounder. Thus, this case is impossible.

Therefore, after Stage Three, we identify all confounded pairs and remove spurious direct relations caused by previously unidentified confounded pairs. As a result, we have shown that all selection pairs, direct relations, and confounded pairs in the ground-truth causal graph are identifiable. □

D. Additional Experimental Results

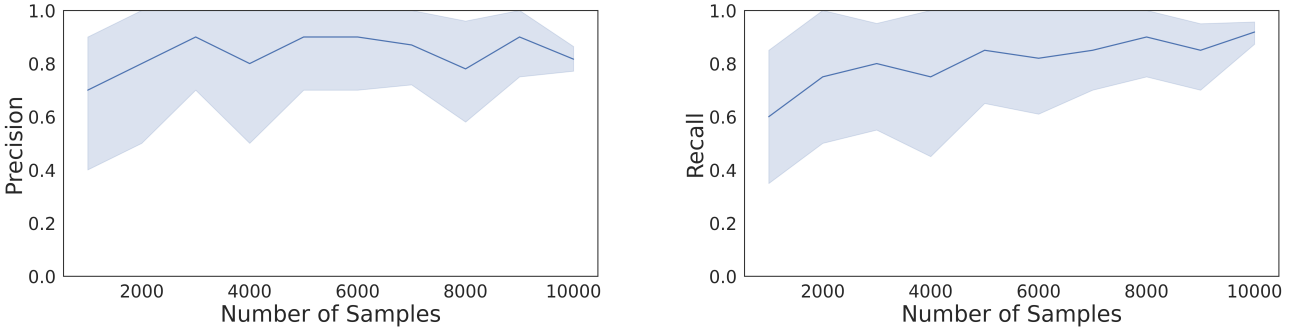


Figure 12. Precision and recall for selection pairs w.r.t. different sample sizes.

In this section, we provide additional experimental results to further explore the proposed framework. We conduct experiments on synthetic datasets with sample size in $\{1000, 2000, \ldots, 10000\}$. All results are from 10 trials with different random seeds. Although our results focus on asymptotic properties, the algorithm can still identify most selection pairs when the sample size is relatively low. As expected, the variance of both precision and recall decreases as the sample size increases.