## A Broader impact

In this work, we propose an algorithm for estimating causal effects from observational studies without relying on expert knowledge of the causal model. Our approach is particularly valuable in scenarios where conducting randomized control trials (RCTs) is challenging or unethical, such as in healthcare settings where consensus treatment protocols are often determined based on observational data. While our algorithm shows promise in providing accurate causal effect estimates, it is crucial to address the potential negative impact of incorrect results that may arise from our work.

One significant concern is the possibility of miscalculating the treatment effect due to limitations in testing power at finite sample sizes or the misidentification of certain features as direct children of the treatment variable. This introduces the risk of inaccurate estimations, which could have detrimental consequences when making decisions or establishing treatment protocols based on the conclusions derived from our algorithm. It is essential to approach the interpretation of our algorithm's results with caution and subject them to critical evaluation. It is worth noting that the potential for incorrect results is not unique to our algorithm but is inherent in most observational studies and effect estimation algorithms. Acknowledging these potential negative impacts emphasizes the need for further research to improve the reliability and accuracy of causal effect estimation in observational studies.

## B Preliminaries about ancestral graphs

In this section, we provide the definition of *partial ancestral graphs* (PAGs). PAGs are defined using *maximal ancestral graphs* (MAGs). Below, we define MAGs and PAGs based on their construction from directed acyclic graphs (DAGs).

A MAG can be obtained from a DAG as follows: if two observed nodes $x_1$ and $x_2$ cannot be d-separated conditioned on any subset of observed variables, then $(i)$ $x_1 \longrightarrow x_2$ is added in the MAG if $x_1$ is an ancestor of $x_2$ in the DAG, $(ii)$ $x_2 \longrightarrow x_1$ is added in the MAG if $x_2$ is an ancestor of $x_1$ in the DAG, and $(iii)$ $x_1 \dashleftarrow\dashrightarrow x_2$ is added in the MAG if $x_1$ and $x_2$ are not ancestrally related in the DAG. $(iv)$ After the above three operations, if both $x_1 \dashleftarrow\dashrightarrow x_2$ and $x_1 \longrightarrow x_2$ are present, we retain only the directed edge. In general, a MAG represents a collection of DAGs that share the same set of observed variables and exhibit the same independence and ancestral relations among these observed variables. It is possible for different MAGs to be Markov equivalent, meaning they represent the exact same independence model.

A PAG shares the same adjacencies as any MAG in the observational equivalence class of MAGs. An end of an edge in the PAG is marked with an arrow ($>$ or $<$) if the edge appears with the same arrow in all MAGs in the equivalence class. An end of an edge in the PAG is marked with a circle ($o$) if the edge appears as an arrow ($>$ or $<$) and a tail ($-$) in two different MAGs in the equivalence class.

## C Rules of do-calculus

In this section, we provide the do-calculus rules of Pearl [1995] that are used to prove our main results in the following sections. We build upon the definition of semi-Markovian causal model from Section 2.

For any $\mathbf{v} \in \mathcal{W}$, let $\mathcal{G}_{\overline{\mathbf{v}}}$ be the graph obtained by removing the edges going into $\mathbf{v}$ in $\mathcal{G}$, and let $\mathcal{G}_{\underline{\mathbf{v}}}$ be the graph obtained by removing the edges going out of $\mathbf{v}$ in $\mathcal{G}$.

**Theorem C.1** (Rules of do-calculus, Pearl [1995])**.** *For any disjoint subsets $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4 \subseteq \mathcal{W}$, we have the following rules.*

*Rule 1:* $\mathbb{P}(\mathbf{v}_1|do(\mathbf{v}_2), \mathbf{v}_3, \mathbf{v}_4) = \mathbb{P}(\mathbf{v}_1|do(\mathbf{v}_2), \mathbf{v}_3)$ *if* $\mathbf{v}_1 \perp\!\!\!\perp_d \mathbf{v}_4 | \mathbf{v}_2, \mathbf{v}_3$ *in* $\mathcal{G}_{\overline{\mathbf{v}_2}}$.

*Rule 2:* $\mathbb{P}(\mathbf{v}_1|do(\mathbf{v}_2), \mathbf{v}_3, do(\mathbf{v}_4)) = \mathbb{P}(\mathbf{v}_1|do(\mathbf{v}_2), \mathbf{v}_3, \mathbf{v}_4)$ *if* $\mathbf{v}_1 \perp\!\!\!\perp_d \mathbf{v}_4 | \mathbf{v}_2, \mathbf{v}_3$ *in* $\mathcal{G}_{\overline{\mathbf{v}_2}, \underline{\mathbf{v}_4}}$.

*Rule 3:* $\mathbb{P}(\mathbf{v}_1|do(\mathbf{v}_2), \mathbf{v}_3, do(\mathbf{v}_4)) = \mathbb{P}(\mathbf{v}_1|do(\mathbf{v}_2), \mathbf{v}_3)$ *if* $\mathbf{v}_1 \perp\!\!\!\perp_d \mathbf{v}_4 | \mathbf{v}_2, \mathbf{v}_3$ *in* $\mathcal{G}_{\overline{\mathbf{v}_2}, \overline{\mathbf{v}_4(\mathbf{v}_3)}}$,

*where $\mathbf{v}_4(\mathbf{v}_3)$ is the set of nodes in $\mathbf{v}_4$ that are not ancestors of any node in $\mathbf{v}_3$ in $\mathcal{G}_{\overline{\mathbf{v}_2}}$. Pearl [1995] also gave an alternative criterion for Rule 3.*

*Rule 3a:* $\mathbb{P}(\mathbf{v}_1|do(\mathbf{v}_2), \mathbf{v}_3, do(\mathbf{v}_4)) = \mathbb{P}(\mathbf{v}_1|do(\mathbf{v}_2), \mathbf{v}_3)$ *if* $\mathbf{v}_1 \perp\!\!\!\perp_d F_{\mathbf{v}_4} | \mathbf{v}_2, \mathbf{v}_3$ *in* $\mathcal{G}_{\overline{\mathbf{v}_2}}^{\mathbf{v}_4}$, *where $\mathcal{G}^{\mathbf{v}_4}$ is the graph obtained from $\mathcal{G}$ after adding $(a)$ a node $F_{\mathbf{v}_4}$ and $(b)$ edges from $F_{\mathbf{v}_4}$ to every node in $\mathbf{v}_4$.*

Also, throughout our proofs, we use the following fact.

**Fact 1.** *Consider any $\mathcal{G}'$ obtained by removing any edge(s) from $\mathcal{G}$. For any sets of variables $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \subseteq \mathcal{W}$, if $\mathbf{v}_1$ and $\mathbf{v}_2$ are d-separated by $\mathbf{v}_3$ in $\mathcal{G}$ than $\mathbf{v}_1$ and $\mathbf{v}_2$ are d-separated by $\mathbf{v}_3$ in $\mathcal{G}'$.*

# D Causal Identifiability

In this section, we derive the causal effect for the SMCM in Figure 3(top), i.e., (6), as well as prove Theorem 3.1 one by one.

## D.1 Proof of (6)

First, using the law of total probability, we have

$$\mathbb{P}(y|do(t=t)) = \sum_{z_1,z_2} \mathbb{P}(y|do(t=t), z_1=z_1, z_2=z_2)\mathbb{P}(z_1=z_1, z_2=z_2|do(t=t)). \tag{12}$$

Now, we show that the two terms in RHS of (12) can be simplified as follows

$$\mathbb{P}(y|do(t=t), z_1=z_1, z_2=z_2) = \sum_{t'} \mathbb{P}(y|z_1=z_1, z_2=z_2, t=t')\mathbb{P}(t=t'|z_1=z_1). \tag{13}$$

$$\mathbb{P}(z_1=z_1, z_2=z_2|do(t=t)) = \sum_{\boldsymbol{b}} \mathbb{P}(z_2=z_2|\mathbf{b}=\boldsymbol{b})\mathbb{P}(\mathbf{b}=\boldsymbol{b}|t=t, z_1=z_1)\mathbb{P}(z_1=z_1), \tag{14}$$

Combining (12) to (14) results in (6).

**Proof of (13):** We have

$$\mathbb{P}(y|do(t=t), z_1=z_1, z_2=z_2) \tag{15}$$

$$\stackrel{(a)}{=} \mathbb{P}(y=y|do(t=t), z_1=z_1, z_2=z_2, \mathbf{b}=\boldsymbol{b}) \tag{16}$$

$$\stackrel{(b)}{=} \mathbb{P}(y=y|do(t=t), z_1=z_1, z_2=z_2, do(\mathbf{b}=\boldsymbol{b})) \tag{17}$$

$$\stackrel{(c)}{=} \mathbb{P}(y=y|z_1=z_1, z_2=z_2, do(\mathbf{b}=\boldsymbol{b})) \tag{18}$$

$$\stackrel{(d)}{=} \sum_{t'} \mathbb{P}(y=y|z_1=z_1, z_2=z_2, do(\mathbf{b}=\boldsymbol{b}), t=t')\mathbb{P}(t=t'|z_1=z_1, z_2=z_2, do(\mathbf{b}=\boldsymbol{b})) \tag{19}$$

$$\stackrel{(e)}{=} \sum_{t'} \mathbb{P}(y=y|z_1=z_1, z_2=z_2, t=t')\mathbb{P}(t=t'|z_1=z_1, z_2=z_2, do(\mathbf{b}=\boldsymbol{b})) \tag{20}$$

$$\stackrel{(f)}{=} \sum_{t'} \mathbb{P}(y=y|z_1=z_1, z_2=z_2, t=t')\mathbb{P}(t=t'|z_1=z_1, do(\mathbf{b}=\boldsymbol{b})) \tag{21}$$

$$\stackrel{(g)}{=} \sum_{t'} \mathbb{P}(y=y|z_1=z_1, z_2=z_2, t=t')\mathbb{P}(t=t'|z_1=z_1), \tag{22}$$

where $(a)$ and $(f)$ follow from Rule 1, $(b)$ follows from Rule 2, $(c)$, $(e)$, and $(g)$ follow from Rule 3a, and $(d)$ follows from the law of total probability.

**Proof of (14):** From the law of total probability, we have

$$\mathbb{P}(z_1=z_1, z_2=z_2|do(t=t)) \tag{23}$$

$$= \sum_{\boldsymbol{b}} \mathbb{P}(z_1=z_1, z_2=z_2|do(t=t), \mathbf{b}=\boldsymbol{b})\mathbb{P}(\mathbf{b}=\boldsymbol{b}|do(t=t)) \tag{24}$$

$$\stackrel{(a)}{=} \sum_{\boldsymbol{b}} \mathbb{P}(z_2=z_2|do(t=t), \mathbf{b}=\boldsymbol{b})\mathbb{P}(z_1=z_1|do(t=t), \mathbf{b}=\boldsymbol{b}, z_2=z_2)\mathbb{P}(\mathbf{b}=\boldsymbol{b}|do(t=t)) \tag{25}$$

$$\stackrel{(b)}{=} \sum_{\boldsymbol{b}} \mathbb{P}(z_2 = z_2 | \mathbf{b} = \boldsymbol{b}) \mathbb{P}(z_1 = z_1 | do(t = t), \mathbf{b} = \boldsymbol{b}, z_2 = z_2) \mathbb{P}(\mathbf{b} = \boldsymbol{b} | do(t = t)) \tag{26}$$

$$\stackrel{(c)}{=} \sum_{\boldsymbol{b}} \mathbb{P}(z_2 = z_2 | \mathbf{b} = \boldsymbol{b}) \mathbb{P}(z_1 = z_1 | do(t = t), \mathbf{b} = \boldsymbol{b}) \mathbb{P}(\mathbf{b} = \boldsymbol{b} | do(t = t)) \tag{27}$$

$$\stackrel{(d)}{=} \sum_{\boldsymbol{b}} \mathbb{P}(z_2 = z_2 | \mathbf{b} = \boldsymbol{b}) \mathbb{P}(z_1 = z_1, \mathbf{b} = \boldsymbol{b} | do(t = t)) \tag{28}$$

$$\stackrel{(e)}{=} \sum_{\boldsymbol{b}} \mathbb{P}(z_2 = z_2 | \mathbf{b} = \boldsymbol{b}) \mathbb{P}(z_1 = z_1 | do(t = t)) \mathbb{P}(\mathbf{b} = \boldsymbol{b} | do(t = t), z_1 = z_1) \tag{29}$$

$$\stackrel{(f)}{=} \sum_{\boldsymbol{b}} \mathbb{P}(z_2 = z_2 | \mathbf{b} = \boldsymbol{b}) \mathbb{P}(z_1 = z_1) \mathbb{P}(\mathbf{b} = \boldsymbol{b} | do(t = t), z_1 = z_1) \tag{30}$$

$$\stackrel{(g)}{=} \sum_{\boldsymbol{b}} \mathbb{P}(z_2 = z_2 | \mathbf{b} = \boldsymbol{b}) \mathbb{P}(z_1 = z_1) \mathbb{P}(\mathbf{b} = \boldsymbol{b} | t = t, z_1 = z_1) \tag{31}$$

where $(a)$, $(d)$, and $(e)$ follow from the definition of conditional probability, $(b)$ and $(f)$ follows from Rule 3a, $(c)$ follows from Rule 1, and $(g)$ follows from Rule 2.

### D.2 Proof of Theorem 3.1

Let $\mathrm{An}(y)$ denote the union of $y$ and the set of ancestors of $y$, and let $\mathcal{G}^{\mathrm{An}(y)}$ denote the subgraph of $\mathcal{G}$ composed only of nodes in $\mathrm{An}(y)$. First, we show that if $\mathbf{b} \perp\!\!\!\perp_d y | t, \mathbf{z}$ holds for some $\mathbf{z}$, then there is no bi-directed path between $t$ to $\mathbf{b}$ in $\mathcal{G}^{\mathrm{An}(y)}$.

**Lemma 1.** *Suppose Assumptions 1 to 3 hold. Suppose there exists a set* $\mathbf{z} \subseteq \mathcal{V} \setminus \{t, \mathbf{b}, y\}$ *such that* $\mathbf{b} \perp\!\!\!\perp_d y | t, \mathbf{z}$. *Then, there is no bi-directed path between* $t$ *and* $\mathbf{b}$ *in* $\mathcal{G}^{\mathrm{An}(y)}$.

Given this claim, Theorem 3.1 follows from Tian and Pearl [2002, Theorem 4]. It remains to prove Lemma 1.

**Proof of Lemma 1.** We prove this result by contradiction. First, from Assumptions 1 and 3, $t \in \mathrm{An}(y)$ and $\mathbf{b}_0 \in \mathrm{An}(y)$ for some $\mathbf{b}_0 \subset \mathbf{b}$. Assume there exists a bi-directed path between $t$ and some $b \in \mathbf{b}_0$ in $\mathcal{G}^{\mathrm{An}(y)}$. Let $\mathcal{P}(t, b)$ denote the shortest of these paths. This path is of the form $t \dashleftarrow\!\!\dashrightarrow v_1 \dashleftarrow\!\!\dashrightarrow \cdots \dashleftarrow\!\!\dashrightarrow v_r \dashleftarrow\!\!\dashrightarrow b$ for some $r \geq 0$ where $v_q \in \mathcal{G}^{\mathrm{An}(y)}$ for every $q \in [r]$. We have the following two cases depending on the value of $r$.

(i) $r = 0$: In this case, consider the path $\mathcal{P}(y, b) \supset \mathcal{P}(t, b)$ in $\mathcal{G}$ of the form: $y \dashleftarrow\!\!\dashrightarrow t \dashleftarrow\!\!\dashrightarrow b$ in $\mathcal{G}$ (such a path exists because of Assumption 2). The path $\mathcal{P}(y, b)$ is unblocked when $t$ and $\mathbf{z}$ are conditioned on contradicting $\mathbf{b} \perp\!\!\!\perp_d y | t, \mathbf{z}$.

(ii) $r \geq 1$: In this case, consider the path $\mathcal{P}(y, b) \supset \mathcal{P}(t, b)$ in $\mathcal{G}$ of the form: $y \dashleftarrow\!\!\dashrightarrow t \dashleftarrow\!\!\dashrightarrow v_1 \dashleftarrow\!\!\dashrightarrow \cdots \dashleftarrow\!\!\dashrightarrow v_r \dashleftarrow\!\!\dashrightarrow b$ (such a path exists because of Assumption 2). We have the following two scenarios depending on whether the path $\mathcal{P}(y, b)$ is unblocked or blocked when $t$ and $\mathbf{z}$ are conditioned on. Suppose we condition on $t$ and $\mathbf{z}$.

   (a) The path $\mathcal{P}(y, b)$ is unblocked: In this case, by assumption, $\mathbf{b} \perp\!\!\!\perp_d y | t, \mathbf{z}$ is contradicted.

   (b) The path $\mathcal{P}(y, b)$ is blocked: We create a set $\mathbf{w}$ such that for any $w \in \mathbf{w}$ the following are true: $(a)$ $w = v_q$ for some $q \in [r]$, $(b)$ $w \notin \mathbf{z}$, $(b)$ there is no descendant path $\mathcal{P}(w, z)$ between $w$ and some $z \in \mathbf{z}$, and $(c)$ there is no descendant path $\mathcal{P}(w, t)$ between $w$ and $t$.
   In this scenario, $\mathbf{w} \neq \emptyset$ because $\mathcal{P}(y, b)$ is blocked. Let $w_c \in \mathbf{w}$ be that node which is closest to $b$ in the path $\mathcal{P}(y, b)$. By the choice of $w_c$, the path $\mathcal{P}(w_c, b) \subset \mathcal{P}(y, b)$ is unblocked (when $t$ and $\mathbf{z}$ are conditioned on). Furthermore, by the definition of $\mathbf{w}$, $(a)$ $w_c \in \mathcal{G}^{\mathrm{An}(y)}$ (because $w_c = v_q$ for some $q \in [r]$) and $(b)$ there exists a descendant path $\mathcal{P}(w_c, y)$ between $w_c$ and $y$ such that $t \notin \mathcal{P}(w_c, y)$ as well as $z \notin \mathcal{P}(w_c, y)$ for every $z \in \mathbf{z}$. Therefore, the path $\mathcal{P}(w_c, y)$ is unblocked (when $t$ and $\mathbf{z}$ are conditioned on).
   Consider the path $\mathcal{P}'(y, b)$ obtained after concatenating $\mathcal{P}(w_c, y)$ and $\mathcal{P}(w_c, b)$ at $w_c$. This path is unblocked (when $t$ and $\mathbf{z}$ are conditioned on) because: $(a)$ $\mathcal{P}(w_c, b)$ is unblocked, $(b)$ $\mathcal{P}(w_c, y)$ is unblocked, and $(c)$ there is no collider at $w_c$ in this path (because $\mathcal{P}(w_c, y)$ is a descendant path to $y$). However, this contradicts $\mathbf{b} \perp\!\!\!\perp_d y | t, \mathbf{z}$.

## E A generalized front-door condition

In this section, we prove Theorem 3.2. We begin by stating a few d-separation statements used in this proof. See Appendix F for a proof.

**Lemma 2.** *Suppose Assumptions 1 to 3 and d-separation criteria in Theorem 3.2, i.e.,* (7) *and* (8)*, hold. Then,*

*(a)* $\mathsf{y} \perp\!\!\!\perp_d F_{\mathsf{t}} | \mathbf{z}, \mathbf{b}$ *in* $\mathcal{G}^{\mathsf{t}}_{\underline{\mathbf{b}}}$ *and* $\mathsf{y} \perp\!\!\!\perp_d F_{\mathsf{t}} | \mathbf{z}^{(i)}, \mathbf{b}$ *in* $\mathcal{G}^{\mathsf{t}}_{\underline{\mathbf{b}}}$*,*

*(b)* $\mathsf{t} \perp\!\!\!\perp_d \mathbf{b}$ *in* $\mathcal{G}_{\underline{\mathsf{t}}}$*,*

*(c)* $\mathsf{t} \perp\!\!\!\perp_d \mathbf{z}^{(i)} | \mathbf{b}$ *in* $\mathcal{G}_{\underline{\mathsf{t}}}$*,*

*(d)* $\mathsf{t} \perp\!\!\!\perp_d F_{\mathbf{b}} | \mathbf{z}^{(i)}$ *in* $\mathcal{G}^{\mathbf{b}}$*, and*

*(e)* $\mathsf{y} \perp\!\!\!\perp_d \mathbf{b} | \mathsf{t}, \mathbf{z}^{(i)}$ *in* $\mathcal{G}_{\underline{\mathbf{b}}}$*.*

Now, we proceed with the proof in two parts. In the first part, we prove (9), and in the second part, we prove (10).

### E.1 Proof of (9)

First, using the law of total probability, we have

$$\mathbb{P}(\mathsf{y} = y | do(\mathsf{t} = t)) = \sum_{\boldsymbol{z}} \mathbb{P}(\mathsf{y} = y | do(\mathsf{t} = t), \mathbf{z} = \boldsymbol{z}) \mathbb{P}(\mathbf{z} = \boldsymbol{z} | do(\mathsf{t} = t)). \tag{32}$$

Now, we show that the two terms in RHS of (32) can be simplified as follows

$$\mathbb{P}(\mathsf{y} = y | do(\mathsf{t} = t), \mathbf{z} = \boldsymbol{z}) = \sum_{t'} \mathbb{P}(\mathsf{y} = y | \mathbf{z} = \boldsymbol{z}, \mathsf{t} = t') \mathbb{P}(\mathsf{t} = t'). \tag{33}$$

$$\mathbb{P}(\mathbf{z} = \boldsymbol{z} | do(\mathsf{t} = t)) = \mathbb{P}(\mathbf{z} = \boldsymbol{z} | \mathsf{t} = t), \tag{34}$$

Combining (33) and (34) completes the proof of (9).

**Proof of (33):** We have

$$\begin{aligned}
\mathbb{P}(\mathsf{y} = y | do(\mathsf{t} = t), \mathbf{z} = \boldsymbol{z}) &\overset{(a)}{=} \mathbb{P}(\mathsf{y} = y | do(\mathsf{t} = t), \mathbf{z} = \boldsymbol{z}, \mathbf{b} = \boldsymbol{b}) && (35)\\
&\overset{(b)}{=} \mathbb{P}(\mathsf{y} = y | do(\mathsf{t} = t), \mathbf{z} = \boldsymbol{z}, do(\mathbf{b} = \boldsymbol{b})) && (36)\\
&\overset{(c)}{=} \mathbb{P}(\mathsf{y} = y | \mathbf{z} = \boldsymbol{z}, do(\mathbf{b} = \boldsymbol{b})) && (37)\\
&\overset{(d)}{=} \sum_{t'} \mathbb{P}(\mathsf{y} = y | \mathbf{z} = \boldsymbol{z}, do(\mathbf{b} = \boldsymbol{b}), \mathsf{t} = t') \mathbb{P}(\mathsf{t} = t' | \mathbf{z} = \boldsymbol{z}, do(\mathbf{b} = \boldsymbol{b})), && (38)
\end{aligned}$$

where $(a)$ follows from Rule 1, (7), and Fact 1, $(b)$ follows from Rule 2, (7), and Fact 1, and $(c)$ follows from Rule 3a and Lemma 2(a), and $(d)$ follows from the law of total probability.

Now, we simplify the first term in (38) as follows:

$$\begin{aligned}
\mathbb{P}(\mathsf{y} = y | \mathbf{z} = \boldsymbol{z}, do(\mathbf{b} = \boldsymbol{b}), \mathsf{t} = t') &\overset{(a)}{=} \mathbb{P}(\mathsf{y} = y | \mathbf{z} = \boldsymbol{z}, \mathbf{b} = \boldsymbol{b}, \mathsf{t} = t') && (39)\\
&\overset{(7)}{=} \mathbb{P}(\mathsf{y} = y | \mathbf{z} = \boldsymbol{z}, \mathsf{t} = t'), && (40)
\end{aligned}$$

where $(a)$ follows from Rule 2, (7), and Fact 1. Likewise, we simplify the second term in (38) as follows:

$$\begin{aligned}
\mathbb{P}(\mathsf{t} = t' | do(\mathbf{b} = \boldsymbol{b}), \mathbf{z} = \boldsymbol{z}) \overset{(a)}{=} \mathbb{P}(\mathsf{t} = t' | do(\mathbf{b} = \boldsymbol{b}), \mathbf{z}^{(i)} = \boldsymbol{z}^{(i)}) &\overset{(b)}{=} \mathbb{P}(\mathsf{t} = t' | \mathbf{z}^{(i)} = \boldsymbol{z}^{(i)}) && (41)\\
&\overset{(c)}{=} \mathbb{P}(\mathsf{t} = t'), && (42)
\end{aligned}$$

where $(a)$ follows from Rule 1, (8), and Fact 1, $(b)$ follows from Rule 3a and Lemma 2(d), and $(c)$ follows (8).

Putting together (38), (40), and (42) results in (33).

**Proof of (34):** From the law of total probability, we have

$$\mathbb{P}(\mathbf{z} = \boldsymbol{z} | do(\mathsf{t} = t)) = \sum_{\boldsymbol{b}} \mathbb{P}(\mathbf{z} = \boldsymbol{z} | do(\mathsf{t} = t), \mathbf{b} = \boldsymbol{b}) \mathbb{P}(\mathbf{b} = \boldsymbol{b} | do(\mathsf{t} = t)). \tag{43}$$

Now, we simplify the first term in (43) as follows:

$$\mathbb{P}(\mathbf{z} = \boldsymbol{z} | do(\mathsf{t} = t), \mathbf{b} = \boldsymbol{b}) \tag{44}$$
$$\stackrel{(a)}{=} \mathbb{P}(\mathbf{z}^{(i)} = \boldsymbol{z}^{(i)} | do(\mathsf{t} = t), \mathbf{b} = \boldsymbol{b}) \cdot \mathbb{P}(\mathbf{z}^{(o)} = \boldsymbol{z}^{(o)} | do(\mathsf{t} = t), \mathbf{b} = \boldsymbol{b}, \mathbf{z}^{(i)} = \boldsymbol{z}^{(i)}) \tag{45}$$
$$\stackrel{(b)}{=} \mathbb{P}(\mathbf{z}^{(i)} = \boldsymbol{z}^{(i)} | do(\mathsf{t} = t), \mathbf{b} = \boldsymbol{b}) \cdot \mathbb{P}(\mathbf{z}^{(o)} = \boldsymbol{z}^{(o)} | \mathsf{t} = t, \mathbf{b} = \boldsymbol{b}, \mathbf{z}^{(i)} = \boldsymbol{z}^{(i)}) \tag{46}$$
$$\stackrel{(c)}{=} \mathbb{P}(\mathbf{z}^{(i)} = \boldsymbol{z}^{(i)} | \mathsf{t} = t, \mathbf{b} = \boldsymbol{b}) \cdot \mathbb{P}(\mathbf{z}^{(o)} = \boldsymbol{z}^{(o)} | \mathsf{t} = t, \mathbf{b} = \boldsymbol{b}, \mathbf{z}^{(i)} = \boldsymbol{z}^{(i)}) \tag{47}$$
$$\stackrel{(d)}{=} \mathbb{P}(\mathbf{z} = \boldsymbol{z} | \mathsf{t} = t, \mathbf{b} = \boldsymbol{b}), \tag{48}$$

where $(a)$ and $(d)$ follow from the definition of conditional probability, $(b)$ follows from Rule 2, (8), and Fact 1, and $(c)$ follows from Rule 2 and Lemma 2(c). Likewise, we simplify the second term in (43) as follows:

$$\mathbb{P}(\mathbf{b} = \boldsymbol{b} | do(\mathsf{t} = t)) \stackrel{(a)}{=} \mathbb{P}(\mathbf{b} = \boldsymbol{b} | \mathsf{t} = t), \tag{49}$$

where $(a)$ follows from Rule 2 and Lemma 2(b).

Putting together (43), (48), and (49), results in (34) as follows:

$$\mathbb{P}(\mathbf{z} = \boldsymbol{z} | do(\mathsf{t} = t)) = \sum_{\boldsymbol{b}} \mathbb{P}(\mathbf{z} = \boldsymbol{z} | \mathsf{t} = t, \mathbf{b} = \boldsymbol{b}) \mathbb{P}(\mathbf{b} = \boldsymbol{b} | \mathsf{t} = t) \stackrel{(a)}{=} \mathbb{P}(\mathbf{z} = \boldsymbol{z} | \mathsf{t} = t), \tag{50}$$

where $(a)$ follows from the law of total probability.

### E.2 Proof of (10)

First, using the law of total probability, we have

$$\mathbb{P}(\mathsf{y} = y | do(\mathsf{t} = t)) = \sum_{\boldsymbol{b}} \mathbb{P}(\mathsf{y} = y | do(\mathsf{t} = t), \mathbf{b} = \boldsymbol{b}) \mathbb{P}(\mathbf{b} = \boldsymbol{b} | do(\mathsf{t} = t)). \tag{51}$$

Now, we show that the first term in RHS of (51) can be simplified as follows

$$\mathbb{P}(\mathsf{y} = y | do(\mathsf{t} = t), \mathbf{b} = \boldsymbol{b}) \tag{52}$$
$$= \sum_{\boldsymbol{z}^{(i)}} \Big( \sum_{t'} \mathbb{P}(\mathsf{y} = y | \mathbf{s} = \boldsymbol{s}, \mathsf{t} = t') \mathbb{P}(\mathsf{t} = t') \Big) \mathbb{P}(\mathbf{z}^{(i)} = \boldsymbol{z}^{(i)} | \mathbf{b} = \boldsymbol{b}, \mathsf{t} = t). \tag{53}$$

where $\mathbf{s} \triangleq (\mathbf{b}, \mathbf{z}^{(i)})$. Using (49) and (53) in (51), completes the proof of (10) as follows:

$$\mathbb{P}(\mathsf{y} = y | do(\mathsf{t} = t)) \tag{54}$$
$$= \sum_{\boldsymbol{s}} \Big( \sum_{t'} \mathbb{P}(\mathsf{y} = y | \mathbf{s} = \boldsymbol{s}, \mathsf{t} = t') \mathbb{P}(\mathsf{t} = t') \Big) \mathbb{P}(\mathbf{z}^{(i)} = \boldsymbol{z}^{(i)} | \mathbf{b} = \boldsymbol{b}, \mathsf{t} = t) \mathbb{P}(\mathbf{b} = \boldsymbol{b} | \mathsf{t} = t) \tag{55}$$
$$\stackrel{(a)}{=} \sum_{\boldsymbol{s}} \Big( \sum_{t'} \mathbb{P}(\mathsf{y} = y | \mathbf{s} = \boldsymbol{s}, \mathsf{t} = t') \mathbb{P}(\mathsf{t} = t') \Big) \mathbb{P}(\mathbf{s} = \boldsymbol{s} | \mathsf{t} = t), \tag{56}$$

where $(a)$ follows from the definition of conditional probability.

**Proof of (53):** From the law of total probability, we have

$$\mathbb{P}(y = y | do(t = t), \mathbf{b} = \boldsymbol{b}) \tag{57}$$
$$= \sum_{\boldsymbol{z}^{(i)}} \mathbb{P}(y = y | \mathbf{b} = \boldsymbol{b}, \mathbf{z}^{(i)} = \boldsymbol{z}^{(i)}, do(t = t)) \mathbb{P}(\mathbf{z}^{(i)} = \boldsymbol{z}^{(i)} | \mathbf{b} = \boldsymbol{b}, do(t = t)). \tag{58}$$

Now, we simplify the first term in (58) as follows:

$$\mathbb{P}(y = y | \mathbf{b} = \boldsymbol{b}, \mathbf{z}^{(i)} = \boldsymbol{z}^{(i)}, do(t = t)) \tag{59}$$
$$\overset{(a)}{=} \mathbb{P}(y = y | do(\mathbf{b} = \boldsymbol{b}), \mathbf{z}^{(i)} = \boldsymbol{z}^{(i)}, do(t = t)) \tag{60}$$
$$\overset{(b)}{=} \mathbb{P}(y = y | do(\mathbf{b} = \boldsymbol{b}), \mathbf{z}^{(i)} = \boldsymbol{z}^{(i)}) \tag{61}$$
$$\overset{(c)}{=} \sum_{t'} \mathbb{P}(y = y | do(\mathbf{b} = \boldsymbol{b}), \mathbf{z}^{(i)} = \boldsymbol{z}^{(i)}, t = t') \mathbb{P}(t = t' | do(\mathbf{b} = \boldsymbol{b}), \mathbf{z}^{(i)} = \boldsymbol{z}^{(i)}), \tag{62}$$

where $(a)$ follows from Rule 2, Lemma 2(e), and Fact 1, $(b)$ follows from Rule 3a and Lemma 2(a), and $(c)$ follows from the law of total probability. We further simplify the first term in (62) as follows:

$$\mathbb{P}(y = y | do(\mathbf{b} = \boldsymbol{b}), \mathbf{z}^{(i)} = \boldsymbol{z}^{(i)}, t = t') \overset{(a)}{=} \mathbb{P}(y = y | \mathbf{b} = \boldsymbol{b}, \mathbf{z}^{(i)} = \boldsymbol{z}^{(i)}, t = t'), \tag{63}$$

where $(a)$ follows from Rule 2 and Lemma 2(e). Using (63) and (42) in (62), we have

$$\mathbb{P}(y = y | \mathbf{b} = \boldsymbol{b}, \mathbf{z}^{(i)} = \boldsymbol{z}^{(i)}, do(t = t)) = \sum_{t'} \mathbb{P}(y = y | \mathbf{b} = \boldsymbol{b}, \mathbf{z}^{(i)} = \boldsymbol{z}^{(i)}, t = t') \mathbb{P}(t = t'). \tag{64}$$

Now, we simplify the second term in (58) as follows:

$$\mathbb{P}(\mathbf{z}^{(i)} = \boldsymbol{z}^{(i)} | \mathbf{b} = \boldsymbol{b}, do(t = t)) \overset{(a)}{=} \mathbb{P}(\mathbf{z}^{(i)} = \boldsymbol{z}^{(i)} | \mathbf{b} = \boldsymbol{b}, t = t), \tag{65}$$

where $(a)$ follows from Rule 2 and Lemma 2(c).

Putting together (58), (64), and (65) results in (53).

### E.3 Necessity of Assumption 2

In this section, we provide an example to signify the importance of Assumption 2 to Theorem 3.2. Consider the semi-Markovian causal model in Figure 7 where Assumptions 1 and 3 hold but Assumption 2 does not hold.

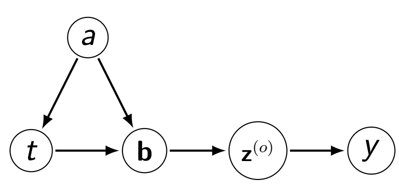


Figure 7: An SMCM signifying the importance of Assumption 2

While $\mathbf{z} = (\mathbf{z}^{(i)}, \mathbf{z}^{(o)})$ satisfies (7) and (8) where $\mathbf{z}^{(i)} = \emptyset$, the causal effect is not equal to the formulae in (9) or (10). To see this, we note that the set $\{a\}$ is a back-door set in Figure 7 implying

$$\mathbb{P}(y | do(t = t)) = \sum_{a} \mathbb{P}(y | a, t) \mathbb{P}(a). \tag{66}$$

Now, we simplify the right hand side of (66) to show explicitly that it is not equivalent to (9). From the law of total probability, we have

$$\mathbb{P}(y | a, t) = \sum_{\mathbf{z}} \mathbb{P}(y | \mathbf{z}, a, t) \mathbb{P}(\mathbf{z} | a, t) \tag{67}$$

$$\overset{(a)}{=} \sum_{\mathbf{z}} \Big( \sum_{t'} \mathbb{P}(y|\mathbf{z}, a, t)\mathbb{P}(t') \Big) \mathbb{P}(\mathbf{z}|a, t) \overset{(b)}{=} \sum_{\mathbf{z}} \Big( \sum_{t'} \mathbb{P}(y|\mathbf{z}, t')\mathbb{P}(t') \Big) \mathbb{P}(\mathbf{z}|a, t), \quad (68)$$

where $(a)$ follows because $\sum_{t'} \mathbb{P}(t') = 1$ and $(b)$ follows $y$ is independent of every other variable conditioned on $\mathbf{z}$. Plugging (68) in (66), we have

$$\mathbb{P}(y|do(t = t)) = \sum_{\mathbf{z}} \Big( \sum_{t'} \mathbb{P}(y|\mathbf{z}, t')\mathbb{P}(t') \Big) \Big( \sum_{a} \mathbb{P}(\mathbf{z}|a, t)\mathbb{P}(a) \Big). \quad (69)$$

Lastly, using the law of total probability, (9) can be rewritten as

$$\mathbb{P}(y|do(t = t)) = \sum_{\mathbf{z}} \Big( \sum_{t'} \mathbb{P}(y|\mathbf{z}, t')\mathbb{P}(t') \Big) \Big( \sum_{a} \mathbb{P}(\mathbf{z}|a, t)\mathbb{P}(a|t) \Big). \quad (70)$$

Therefore, the variables $a$ and $t$ could be such that (69) is different from (70). We note that similar steps can be used to show that (66) is not equivalent to (10). In conclusion, Assumption 2 is crucial for the formulae in (9) and (10) to hold.

## F Proof of Lemma 2

First, we state the following d-separation criterion used to prove Lemma 2(b) and Lemma 2(d). See Appendix F.1 for a proof.

**Lemma 3.** *Suppose Assumptions 1 to 3 hold. Then,* $t \perp\!\!\!\perp_d \mathbf{b}|\mathbf{z}^{(i)}$ *in* $\mathcal{G}_{\underline{t}}$.

Now, we prove each part of Lemma 2 one-by-one.

**Proof of Lemma 2(a)** In $\mathcal{G}^{t}_{\overline{\mathbf{b}}}$, all edges going into $\mathbf{b}$ are removed. Under Assumption 3, this implies that all edges going out of $t$ are removed. Now, consider any path $\mathcal{P}(F_t, y)$ between $F_t$ and $y$ in $\mathcal{G}^{t}_{\overline{\mathbf{b}}}$. This path takes one of the following two forms: $(a)$ $F_t \longrightarrow t \longleftarrow \cdots y$ or $(b)$ $F_t \longrightarrow t \dashleftarrow\!\dashrightarrow \cdots y$. In either case, there is a collider at $t$ in $\mathcal{P}(F_t, y)$. This collider is blocked when $\mathbf{z}$ and $\mathbf{b}$ are conditioned on because $t \notin \mathbf{z}$, $t \notin \mathbf{b}$, and $t$ does not have any descendants in $\mathcal{G}^{t}_{\overline{\mathbf{b}}}$. Therefore, $y \perp\!\!\!\perp_d F_t|\mathbf{z}, \mathbf{b}$ in $\mathcal{G}^{t}_{\overline{\mathbf{b}}}$. Similarly, the collider is blocked when $\mathbf{z}^{(i)}$ and $\mathbf{b}$ are conditioned on because $t \notin \mathbf{z}^{(i)}$, $t \notin \mathbf{b}$, and $t$ does not have any descendants in $\mathcal{G}^{t}_{\overline{\mathbf{b}}}$. Therefore, $y \perp\!\!\!\perp_d F_t|\mathbf{z}^{(i)}, \mathbf{b}$ in $\mathcal{G}^{t}_{\overline{\mathbf{b}}}$.

**Proof of Lemma 2(b)** We prove this by contradiction. Assume there exists at least one unblocked path between $t$ and some $b \in \mathbf{b}$ in $\mathcal{G}_{\underline{t}}$. Let $\mathcal{P}(t, b)$ denote any such unblocked path.

Suppose we condition on $\mathbf{z}^{(i)}$. From Lemma 3, $\mathcal{P}(t, b)$ is blocked in $\mathcal{G}_{\underline{t}}$ when $\mathbf{z}^{(i)}$ is conditioned on. Let $v$ be any node at which $\mathcal{P}(t, b)$ is blocked in $\mathcal{G}_{\underline{t}}$ when $\mathbf{z}^{(i)}$ is conditioned on. We must have that $v \in \mathcal{P}(t, b) \setminus \{t, b\}$ and $v \in \mathbf{z}^{(i)}$. Then, the path $\mathcal{P}(t, v) \subset \mathcal{P}(t, b)$ is unblocked in $\mathcal{G}_{\underline{t}}$ when $\mathbf{z}^{(i)}$ is unconditioned on. However, this contradicts $t \perp\!\!\!\perp_d \mathbf{z}^{(i)}$ in $\mathcal{G}_{\underline{t}}$ (which follows from (8)$(i)$ and Fact 1).

**Proof of Lemma 2(c)** We prove this by contradiction. Assume there exists at least one unblocked path between $t$ and some $z^{(i)} \in \mathbf{z}^{(i)}$ in $\mathcal{G}_{\underline{t}}$ when $\mathbf{b}$ is conditioned on. Let $\mathcal{P}(t, z^{(i)})$ denote any such unblocked path.

Suppose, we uncondition on $\mathbf{b}$. From (8)(i) and Fact 1, we have $t \perp\!\!\!\perp_d \mathbf{z}^{(i)}$ in $\mathcal{G}_{\underline{t}}$. Therefore, $\mathcal{P}(t, z^{(i)})$ is blocked in $\mathcal{G}_{\underline{t}}$ when $\mathbf{b}$ is unconditioned on. Now, we create a set $\mathbf{v}$ consisting of all the nodes at which $\mathcal{P}(t, z^{(i)})$ is blocked in $\mathcal{G}_{\underline{t}}$ when $\mathbf{b}$ is unconditioned on. Define the set $\mathbf{v}$ such that for any $v \in \mathbf{v}$, the following are true: $(a)$ $v \in \mathcal{P}(t, z^{(i)}) \setminus \{t, z^{(i)}\}$, $(b)$ $\mathcal{P}(t, z^{(i)})$ contains a collider at $v$ in $\mathcal{G}_{\underline{t}}$, and $(c)$ there exists an unblocked descendant path from $v$ to some $b \in \mathbf{b}$ in $\mathcal{G}_{\underline{t}}$.

Now, we must have $\mathbf{v} \neq \emptyset$, since $\mathcal{P}(t, z^{(i)})$ is blocked in $\mathcal{G}_{\underline{t}}$ when $\mathbf{b}$ is unconditioned on. Let $v_c \in \mathbf{v}$ be that node which is closest to $t$ in the path $\mathcal{P}(t, z^{(i)})$, and let $\mathcal{P}(v_c, b)$ be an unblocked descendant path from $v$ to some $b \in \mathbf{b}$ in $\mathcal{G}_{\underline{t}}$ (there must be one from the definition of the set $\mathbf{v}$). Consider the path $\mathcal{P}(t, b)$ obtained after concatenating $\mathcal{P}(t, v_c) \subset \mathcal{P}(t, z^{(i)})$ and $\mathcal{P}(v_c, b)$. By the definition of $\mathbf{v}$ and the choice of $v_c$, $\mathcal{P}(t, b)$ is unblocked in $\mathcal{G}_{\underline{t}}$ since $(a)$ $\mathcal{P}(t, v_c)$ is unblocked in $\mathcal{G}_{\underline{t}}$, $(b)$ $\mathcal{P}(v_c, b)$ is unblocked in $\mathcal{G}_{\underline{t}}$, and $(c)$ there is no collider at $v_c$ in $\mathcal{P}(t, b)$. However, this contradicts $t \perp\!\!\!\perp_d \mathbf{b}$ in $\mathcal{G}_{\underline{t}}$ (which follows from Lemma 2(b)).

**Proof of Lemma 2(d)** We prove this by contradiction. Assume there exists at least one unblocked path between $t$ and $F_{\mathbf{b}}$ in $\mathcal{G}^{\mathbf{b}}$ when $\mathbf{z}^{(i)}$ is conditioned on. Let $\mathcal{P}(t, F_{\mathbf{b}})$ denote the shortest of these unblocked path. By definition of $\mathcal{G}^{\mathbf{b}}$, this path has to be of the form: $t \cdots, b \longleftarrow F_{\mathbf{b}}$ for some $b \in \mathbf{b}$. Now, we have the following three cases:

(i) $\mathcal{P}(t, F_{\mathbf{b}})$ contains $t \longrightarrow b$: In this case, because a path is a sequence of distinct nodes, $\mathcal{P}(t, F_{\mathbf{b}})$ has to be $t \longrightarrow b \longleftarrow F_{\mathbf{b}}$. By assumption, $\mathcal{P}(t, F_{\mathbf{b}})$ is unblocked when $\mathbf{z}^{(i)}$ is conditioned on. Since there is a collider at $b$ in $\mathcal{P}(t, F_{\mathbf{b}})$, there exists at least one unblocked descendant path from $b$ to $\mathbf{z}^{(i)}$ when $\mathbf{z}^{(i)}$ is conditioned on. Let $\mathcal{P}(b, z^{(i)})$ denote the shortest of these paths from $b$ to some $z^{(i)} \in \mathbf{z}^{(i)}$ in $\mathcal{G}^{\mathbf{b}}$. We note that this path also exists in $\mathcal{G}$ and is of the form $b \longrightarrow \cdots \longrightarrow z^{(i)}$

Suppose we uncondition on $\mathbf{z}^{(i)}$. Consider the path $\mathcal{P}(t, z^{(i)}) \supset \mathcal{P}(b, z^{(i)})$ between $t$ and $z^{(i)}$ of the form $t \longrightarrow b \longrightarrow \cdots \longrightarrow z^{(i)}$ in $\mathcal{G}$. This path remains unblocked even when $\mathbf{z}^{(i)}$ is unconditioned on as it does not have any colliders. This contradicts $z^{(i)} \perp\!\!\!\perp_d t$ (which follows from (8)).

(ii) $\mathcal{P}(t, F_{\mathbf{b}})$ contains $t \longrightarrow b_1$ for some $b_1 \in \mathbf{b}$ such that $b_1 \neq b$: In this case, the path $\mathcal{P}(t, F_{\mathbf{b}})$ has to be of the form $t \longrightarrow b_1 \cdots b \longleftarrow F_{\mathbf{b}}$. Therefore, there exists at least one collider on the path $\mathcal{P}(t, F_{\mathbf{b}})$. Let $v \in \mathcal{P}(t, F_{\mathbf{b}}) \setminus \{t, F_{\mathbf{b}}\}$ be the collider on the path $\mathcal{P}(t, F_{\mathbf{b}})$ that is closest to $b_1$. Consider the path $\mathcal{P}(t, v) \subset \mathcal{P}(t, F_{\mathbf{b}})$. We note that this path also exists in $\mathcal{G}$ and is of the form $t \longrightarrow b_1 \longrightarrow \cdots \longrightarrow v$.

By assumption, $\mathcal{P}(t, F_{\mathbf{b}})$ is unblocked when $\mathbf{z}^{(i)}$ is conditioned on. Since there is a collider at $v$ in $\mathcal{P}(t, F_{\mathbf{b}})$, there exists at least one unblocked descendant path from $v$ to $\mathbf{z}^{(i)}$ when $\mathbf{z}^{(i)}$ is conditioned on. Let $\mathcal{P}(v, z^{(i)})$ denote the shortest of these paths from $v$ to some $z^{(i)} \in \mathbf{z}^{(i)}$ in $\mathcal{G}^{\mathbf{b}}$. We note that this path also exists in $\mathcal{G}$ and is of the form $v \longrightarrow \cdots \longrightarrow z^{(i)}$.

Suppose we uncondition on $\mathbf{z}^{(i)}$. Consider the path $\mathcal{P}(t, z^{(i)})$ between $t$ and $z^{(i)}$ in $\mathcal{G}$ obtained after concatenating $\mathcal{P}(t, v) \subset \mathcal{P}(t, F_{\mathbf{b}})$ and $\mathcal{P}(v, z^{(i)})$. This path, of the form $t \longrightarrow b_1 \longrightarrow \cdots \longrightarrow v \longrightarrow \cdots \longrightarrow z^{(i)}$, remains unblocked even when $\mathbf{z}^{(i)}$ is unconditioned on as it does not have any colliders. This contradicts $z^{(i)} \perp\!\!\!\perp_d t$ (which follows from (8)).

(ii) $\mathcal{P}(t, F_{\mathbf{b}})$ does not contain $t \longrightarrow b_1$ for every $b_1 \in \mathbf{b}$: By assumption, $\mathcal{P}(t, F_{\mathbf{b}})$ is unblocked in $\mathcal{G}^{\mathbf{b}}$ when $\mathbf{z}^{(i)}$ is conditioned on. Therefore, if $\mathcal{P}(t, F_{\mathbf{b}})$ does not contain the edge $t \longrightarrow b_1$ for any $b_1 \in \mathbf{b}$, there exists a path $\mathcal{P}(t, b)$ between $t$ to $b$ in $\mathcal{G}$ that is unblocked when $\mathbf{z}^{(i)}$ is conditioned on, and takes one of the following two forms: $(a)$ $t \longleftarrow \cdots b$ or $(b)$ $t \dashleftarrow\!\!\dashrightarrow \cdots b$. Then, it is easy to see that the path $\mathcal{P}(t, b)$ also remains unblocked in $\mathcal{G}_{\underline{t}}$ while $\mathbf{z}^{(i)}$ is conditioned on. However, this contradicts $t \perp\!\!\!\perp_d \mathbf{b} | \mathbf{z}^{(i)}$ in $\mathcal{G}_{\underline{t}}$ (which follows from Lemma 3).

**Proof of Lemma 2(e)** We prove this by contradiction. Assume there exists at least one unblocked path between $y$ and some $b \in \mathbf{b}$ in $\mathcal{G}_{\underline{\mathbf{b}}}$ when $t$ and $\mathbf{z}^{(i)}$ are conditioned on. Let $\mathcal{P}(b, y)$ denote the shortest of these unblocked path. Therefore, no $b_1 \in \mathbf{b}$, such that $b_1 \neq b$, is on the path $\mathcal{P}(b, y)$, i.e., $b_1 \notin \mathcal{P}(b, y)$. Further, $\mathcal{P}(b, y)$ takes one of the following two forms because all the edges going out of $\mathbf{b}$ are removed in $\mathcal{G}_{\underline{\mathbf{b}}}$: $(a)$ $b \longleftarrow \cdots y$ or $(b)$ $b \dashleftarrow\!\!\dashrightarrow \cdots y$.

Suppose we condition on $\mathbf{z}^{(o)}$ (while $t$ and $\mathbf{z}^{(i)}$ are still conditioned on). From (7) and Fact 1, we have $y \perp\!\!\!\perp_d b | t, \mathbf{z}$ in $\mathcal{G}_{\underline{\mathbf{b}}}$. Therefore, the path $\mathcal{P}(b, y)$ is blocked in $\mathcal{G}_{\underline{\mathbf{b}}}$ when $\mathbf{z}^{(o)}$ is conditioned on (while $t$ and $\mathbf{z}^{(i)}$ are still conditioned on). Let $v$ be any node at which $\mathcal{P}(b, y)$ is blocked in $\mathcal{G}_{\underline{\mathbf{b}}}$ when $\mathbf{z}^{(o)}$ is conditioned on (while $t$ and $\mathbf{z}^{(i)}$ are still conditioned on). We must have that $v \in \mathcal{P}(b, y) \setminus \{y, b\}$ and $v \in \mathbf{z}^{(o)}$. Suppose we uncondition on $\mathbf{z}^{(o)}$ (while $t$ and $\mathbf{z}^{(i)}$ are still conditioned on). Then, the path $\mathcal{P}(b, v) \subset \mathcal{P}(b, y)$ is unblocked in $\mathcal{G}_{\underline{\mathbf{b}}}$.

We consider the following two scenarios depending on whether or not $\mathcal{P}(b, v)$ contains $t$. In both scenarios, we show that there is an unblocked path between $t$ and $v$ in $\mathcal{G}_{\underline{\mathbf{b}}}$ when we condition on $\mathbf{b}$ (while $t$ and $\mathbf{z}^{(i)}$ are still conditioned on).

(i) $\mathcal{P}(b, v)$ contains $t$: Consider the path $\mathcal{P}(t, v) \subset \mathcal{P}(b, v)$ which is unblocked in $\mathcal{G}_{\underline{\mathbf{b}}}$ when $t$ and $\mathbf{z}^{(i)}$ are conditioned on. Further, by the choice of $\mathcal{P}(b, y)$, no $b_1 \in \mathbf{b}$ is on the path

$\mathcal{P}(t, v)$. Therefore, the path $\mathcal{P}(t, v)$ in $\mathcal{G}_{\underline{\mathbf{b}}}$ remains unblocked when we condition on $\mathbf{b}$ (while $t$ and $\mathbf{z}^{(i)}$ are still conditioned on).

(ii) $\mathcal{P}(b, v)$ does not contain $t$: Consider the path $\mathcal{P}(t, v) \supset \mathcal{P}(b, v)$ (by including the extra edge $t \to b$) which takes one of the following two forms: $(a)$ $t \longrightarrow b \longleftarrow \cdots v$ or $(b)$ $t \longrightarrow b \dashleftarrow\dashrightarrow \cdots v$. Further, by the choice of $\mathcal{P}(b, y)$, no $b_1 \in \mathbf{b}$ ($b_1 \neq b$) is on the path $\mathcal{P}(t, v)$. Suppose we condition on $\mathbf{b}$ (while $t$ and $\mathbf{z}^{(i)}$ are still conditioned on). Then, the path $\mathcal{P}(t, v)$ in $\mathcal{G}_{\underline{\mathbf{b}}}$ is unblocked because $(a)$ the collider at $b$ is unblocked when $\mathbf{b}$ is conditioned on and $(b)$ the path $\mathcal{P}(b, v)$ in $\mathcal{G}_{\underline{\mathbf{b}}}$ remains unblocked when $\mathbf{b}$ is conditioned on (while $t$ and $\mathbf{z}^{(i)}$ are still conditioned on).

Now, suppose we uncondition on $t$ (while $\mathbf{b}$ and $\mathbf{z}^{(i)}$ are still conditioned on). We have the following two scenarios depending on whether or not $\mathcal{P}(t, v)$ in $\mathcal{G}_{\underline{\mathbf{b}}}$ remains unblocked. In both scenarios, we show that there is an unblocked path between $t$ and $v$ in $\mathcal{G}_{\underline{\mathbf{b}}}$ when we uncondition on $t$ (while $\mathbf{b}$ and $\mathbf{z}^{(i)}$ are still conditioned on).

1. If $\mathcal{P}(t, v)$ remains unblocked: In this case, $\mathcal{P}(t, v)$ in $\mathcal{G}_{\underline{\mathbf{b}}}$ is an unblocked path between $t$ and $v$ when $\mathbf{z}^{(i)}$ and $\mathbf{b}$ are conditioned on, as desired.
2. If $\mathcal{P}(t, v)$ does not remain unblocked: In this case, it is the unconditioning on $t$ (while $\mathbf{b}$ and $\mathbf{z}^{(i)}$ are still conditioned on) that blocks $\mathcal{P}(t, v)$. Now, we create a set $\mathbf{w}$ consisting of all the nodes at which $\mathcal{P}(t, v)$ is blocked in $\mathcal{G}_{\underline{\mathbf{b}}}$ when $t$ is unconditioned on (while $\mathbf{b}$ and $\mathbf{z}^{(i)}$ are still conditioned on). Define the set $\mathbf{w}$ such that for any $w \in \mathbf{w}$, the following are true: $(a)$ $w \in \mathcal{P}(t, v) \setminus \{t, v\}$, $(b)$ $\mathcal{P}(t, v)$ contains a collider at $w$ in $\mathcal{G}_{\underline{\mathbf{b}}}$, and $(c)$ there exists an unblocked descendant path from $w$ to $t$ in $\mathcal{G}_{\underline{\mathbf{b}}}$.
   Now, we must have $\mathbf{w} \neq \emptyset$, since $\mathcal{P}(t, v)$ is blocked in $\mathcal{G}_{\underline{\mathbf{b}}}$ when $t$ is unconditioned on (while $\mathbf{b}$ and $\mathbf{z}^{(i)}$ are still conditioned on). Let $w_c \in \mathbf{w}$ be that node which is closest to $v$ in the path $\mathcal{P}(t, v)$, and let $\mathcal{P}(w_c, t)$ be an unblocked descendant path from $w_c$ to $t$ in $\mathcal{G}_{\underline{\mathbf{b}}}$ (there must be one from the definition of the set $\mathbf{w}$). Consider the path $\mathcal{P}'(v, t)$ obtained after concatenating $\mathcal{P}(v, w_c) \subset \mathcal{P}(t, v)$ and $\mathcal{P}(w_c, t)$. By the definition of $\mathbf{w}$ and the choice of $w_c$, $\mathcal{P}'(v, t)$ is unblocked in $\mathcal{G}_{\underline{\mathbf{b}}}$ when $t$ is unconditioned on (while $\mathbf{b}$ and $\mathbf{z}^{(i)}$ are still conditioned on) since $(a)$ $\mathcal{P}(v, w_c)$ is unblocked, $(b)$ $\mathcal{P}(w_c, t)$ is unblocked, and $(c)$ there is no collider at $w_c$ in $\mathcal{P}'(v, t)$. Therefore, we have an unblocked path between $t$ and $v$ in $\mathcal{G}_{\underline{\mathbf{b}}}$ when $\mathbf{z}^{(i)}$ and $\mathbf{b}$ are conditioned on, as desired.

To conclude the proof, we note that the existence of an unblocked path between $t$ and $v \in \mathbf{z}^{(o)}$ in $\mathcal{G}_{\underline{\mathbf{b}}}$ when $\mathbf{z}^{(i)}$ and $\mathbf{b}$ are conditioned on contradicts $\mathbf{z}^{(o)} \perp\!\!\!\perp_d t | \mathbf{b}, \mathbf{z}^{(i)}$ in $\mathcal{G}_{\underline{\mathbf{b}}}$ (which follows from (8) and Fact 1).

### F.1 Proof of Lemma 3

First, we claim $t \perp\!\!\!\perp_d \mathbf{b} | \mathbf{z}$ in $\mathcal{G}_{\underline{t}}$. We assume this claim and proceed to prove the statement in the Lemma by contradiction. Assume there exists at least one unblocked path between $t$ and some $b \in \mathbf{b}$ in $\mathcal{G}_{\underline{t}}$ when $\mathbf{z}^{(i)}$ is conditioned on. Let $\mathcal{P}(t, b)$ denote the shortest of these unblocked path. Therefore, no $b_1 \in \mathbf{b}$ such that $b_1 \neq b$ is not on the path $\mathcal{P}(t, b)$, i.e., $b_1 \notin \mathcal{P}(t, b)$.

Suppose we condition on $\mathbf{z}^{(o)}$ (while $\mathbf{z}^{(i)}$ is still conditioned on). From the claim, $\mathcal{P}(t, b)$ is blocked in $\mathcal{G}_{\underline{t}}$ when $\mathbf{z}^{(o)}$ is conditioned on (while $\mathbf{z}^{(i)}$ is still conditioned on). Let $v$ be any node at which $\mathcal{P}(t, b)$ is blocked in $\mathcal{G}_{\underline{t}}$ when $\mathbf{z}^{(o)}$ is conditioned on (while $\mathbf{z}^{(i)}$ is still conditioned on). We must have that $v \in \mathcal{P}(t, b) \setminus \{t, b\}$ and $v \in \mathbf{z}^{(o)}$. Then, the path $\mathcal{P}(t, v) \subset \mathcal{P}(t, b)$ is unblocked in $\mathcal{G}_{\underline{t}}$ when $\mathbf{z}^{(o)}$ is unconditioned on (while $\mathbf{z}^{(i)}$ is still conditioned on). Further, no $b \in \mathbf{b}$ is on the path $\mathcal{P}(t, v)$. As a result, the path $\mathcal{P}(t, v)$ remains unblocked when $\mathbf{b}$ is conditioned on (while $\mathbf{z}^{(i)}$ is still conditioned on). However, this contradicts $t \perp\!\!\!\perp_d \mathbf{z}^{(o)} | b, \mathbf{z}^{(i)}$ in $\mathcal{G}_{\underline{t}}$ (which follows from (8)$(ii)$ and Fact 1).

*Proof of Claim - $t \perp\!\!\!\perp_d \mathbf{b} | \mathbf{z}$ in $\mathcal{G}_{\underline{t}}$:* It remains to prove the claim $t \perp\!\!\!\perp_d \mathbf{b} | \mathbf{z}$ in $\mathcal{G}_{\underline{t}}$. We prove this by contradiction. Assume there exists at least one unblocked path between $t$ and some $b \in \mathbf{b}$ in $\mathcal{G}_{\underline{t}}$ when $\mathbf{z}$ is conditioned on. Let $\mathcal{P}(t, b)$ denote any such unblocked path. This path takes one of the following two forms: $(a)$ $t \longleftarrow \cdots b$ or $(b)$ $t \dashleftarrow\dashrightarrow \cdots b$ because all edges going out of $t$ are removed in $\mathcal{G}_{\underline{t}}$.

Suppose we condition on $t$ (while $\mathbf{z}$ is still conditioned on). The path $\mathcal{P}(t, b)$ remains unblocked because $t \notin \mathcal{P}(t, b)$ (a path is a sequence of distinct nodes). Then, the path $\mathcal{P}(y, b) \supset \mathcal{P}(t, b)$ of the form $(a)$ $y \dashleftarrow\dashrightarrow t \longleftarrow \cdots b$ or $(b)$ $y \dashleftarrow\dashrightarrow t \dashleftarrow\dashrightarrow \cdots b$ is unblocked because the additional conditioning on $t$ (while $\mathbf{z}$ is still conditioned on) unblocks the collider at $t$. However, this contradicts $\mathbf{b} \perp\!\!\!\perp_d y | t, \mathbf{z}$ in $\mathcal{G}_{\underline{t}}$ (which follows from (7) and Fact 1).

# G Experimental Results

In this section, we provide additional experimental results. First, we provide more details regarding the numerical example in Section 3.1. Next, we demonstrate the applicability of our method on a class of graphs slightly different from the one in Section 4.1. Then, we provide the 6 random graphs from Section 4.2 as well as ATE estimation results on specific choices of SMCMs including the one in Figure 2. Finally, we provide histograms analogous to Figure 6 for the second choice of $\mathbf{b}$ on German credit dataset as well as details about our analysis with Adult dataset.

## G.1 Numerical example in Section 3.1

The observed variables for this example also follow the structural equation model in (11). Also, to generate the true ATE, we intervene on the generation model in (11) by setting $t = 0$ and $t = 1$.

## G.2 Applicability to a class of random graphs

As in Section 4.1, we create a class of random SMCMs, sample 100 SMCMs from this class, and check if (7) and (8) hold by checking for corresponding d-separations in the SMCMs. The class of random graphs considered here is analogous to the class of random graphs considered in Section 4.1 expect for the choice of $t$. Here, we choose any variable that is ancestor of $y$ but not its parent as $t$. This is in contrast to Section 4.1 where we choose any variable that is ancestor of $y$ but not its parent or grandparent as $t$. We compare the success rate of the same two approaches: $(i)$ exhaustive search for $\mathbf{z}$ satisfying (7) and (8) and $(ii)$ search for a $\mathbf{z}$ of size at-most 5 satisfying (7) and (8). We provide the number of successes of these approaches as a tuple in Table 2 for various $p$, $d$, and $q$. As before, we see that the two approaches have comparable performances and the IDP algorithm gives 0 successes across various $p$, $d$, and $q$ even though it is supplied with the true PAG. Also, as expected the number of successes for this class of graphs is much lower than the class considered in Section 4.1.

Table 2: Number of successes out of 100 random graphs for our methods shown as a tuple. The first method searches a $\mathbf{z}$ exhaustively and the second method searches a $\mathbf{z}$ with size at-most 5.

| | $p = 10$ | | | $p = 15$ | | |
|---|---|---|---|---|---|---|
| | $d = 2$ | $d = 3$ | $d = 4$ | $d = 2$ | $d = 3$ | $d = 4$ |
| $q = 0.0$ | $(6, 6)$ | $(3, 3)$ | $(1, 1)$ | $(11, 11)$ | $(2, 2)$ | $(1, 1)$ |
| $q = 0.5$ | $(3, 3)$ | $(0, 0)$ | $(0, 0)$ | $(5, 5)$ | $(2, 2)$ | $(1, 1)$ |
| $q = 1.0$ | $(1, 1)$ | $(0, 0)$ | $(0, 0)$ | $(1, 1)$ | $(0, 0)$ | $(0, 0)$ |

## G.3 ATE estimation

We also conduct ATE estimation experiments on four specific SMCMs. The first SMCM is the graph $\mathcal{G}^{toy}$ in Figure 2. The remaining graphs, named $\mathcal{G}_i^{toy}$, $i \in \{1, 2, 3\}$, are shown in Figure 8, and are obtained by adding additional edges and modifying $\mathcal{G}^{toy}$. These SMCMs are designed in a way such that there exists $\mathbf{z} = (\mathbf{z}^{(i)}, \mathbf{z}^{(o)})$ satisfying the conditional independence statements in Theorem 3.2.

We follow a data generation procedure similar to the one in Section 4.2. In contrast, we show the performance of our approach for a fixed $n$ but different thresholds of p-value $p_v$. We average the ATE error over 50 runs where in each run we set $n = 50000$. As we see in Figure 9, both the ATE estimates returned by Algorithm 1 are far superior compared to the naive front-door adjustment using $\mathbf{b}$.

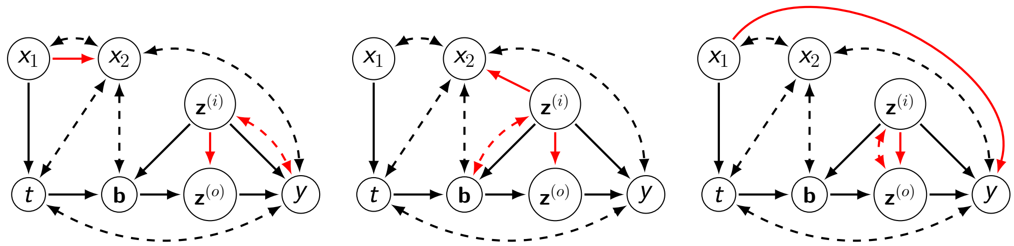


Figure 8: The causal graphs used to further validate our theoretical results. These are obtained by adding additional edges (shown in red) to $\mathcal{G}^{toy}$ in Figure 2. We denote these graphs (from left to right) by $\mathcal{G}_1^{toy}$, $\mathcal{G}_2^{toy}$, and $\mathcal{G}_3^{toy}$, respectively.

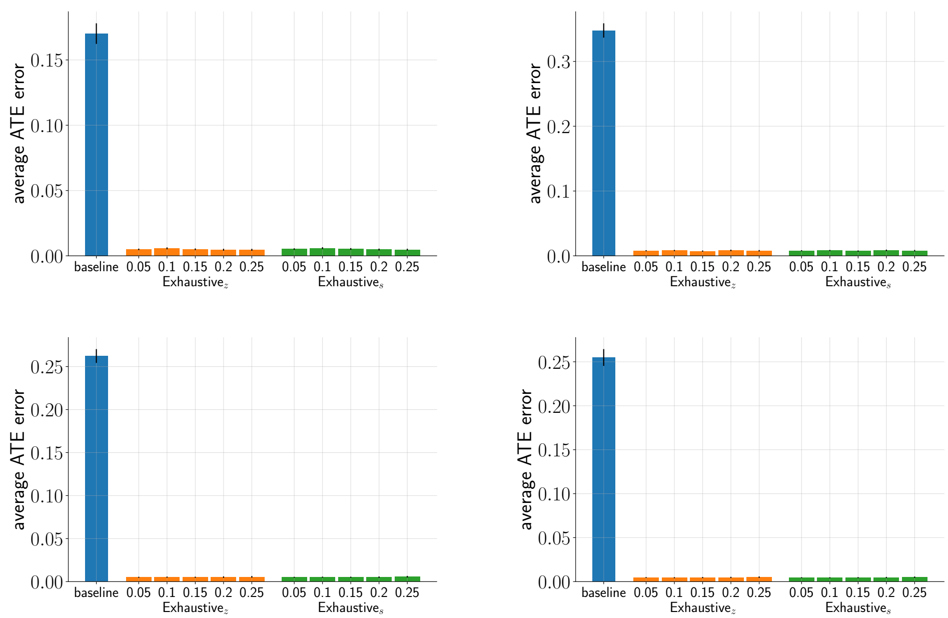


Figure 9: Performance of Algorithm 1 for different p-value thresholds $p_v$ on $\mathcal{G}^{toy}$ in Figure 2 on top left, on $\mathcal{G}_1^{toy}$ from Figure 8 on top right, on $\mathcal{G}_2^{toy}$ in Figure 8 on bottom left, and on $\mathcal{G}_3^{toy}$ from Figure 8 on bottom right

In Figure 11, we provide the 6 random SMCMs used in Section 4.2. As mentioned in Section 4.1, we choose the last variable in the causal ordering as $y$ and a variable that is ancestor of $y$ but not its parent or grandparent as $t$. We also show the corresponding $\mathbf{z} = (\mathbf{z}^{(i)}, \mathbf{z}^{(o)})$ satisfying (7) and (8).

### G.4 German Credit dataset

As in Section 4.3, we assess the conditional independence associated with the selected $\mathbf{z}$ for the choice of $\mathbf{b} = \{$# of people financially dependent on the applicant, applicant's savings$\}$, Algorithm 1 results in $\mathbf{z}^{(i)} = \{$purpose for which the credit was needed, applicant's checking account status with the bank$\}$ via 100 random bootstraps. We show the corresponding p-values for these bootstraps in a histogram in Figure 10 below. As expected, we observe the p-values to be spread out.

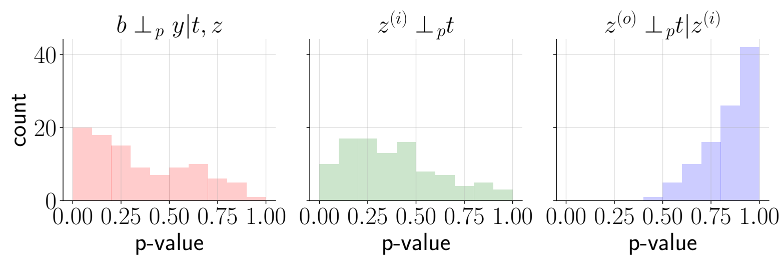


Figure 10: Histograms of p-values of the conditional independencies in (9) and (10) over 100 bootstrap runs for $\mathbf{b} = \{$# of people financially dependent on the applicant, applicant's savings$\}$, Algorithm 1 results in $\mathbf{z}^{(i)} = \{$purpose for which the credit was needed, applicant's checking account status with the bank$\}$.

### G.5 Adult dataset

The Adult dataset [Kohavi and Becker, 1996] is used for income analysis where the goal is to predict whether an individual's income is more than \$50,000 using 14 demographic and socio-economic features. The sensitive attribute $t$ is the individual's sex, either male or female. Further, the categorical attributes are one-hot encoded. As with German Credit dataset, we apply Algorithm 1 with $n_r = 100$ and $p_v = 0.1$ where we search for a set $\mathbf{z} = (\mathbf{z}^{(o)}, \mathbf{z}^{(i)})$ of size at most 3 under the following two assumptions on the set of all children $\mathbf{b}$ of $t$: (1) $\mathbf{b} = \{$# individual's relationship status (which includes wife/husband)$\}$ and (2) $\mathbf{b} = \{$# individual's relationship status (which includes wife/husband), individual's occupation$\}$. In either case, Algorithm 1 was unable to find a suitable $\mathbf{z}$ satisfying $\mathbf{b} \perp_p y|\mathbf{z}, t$. This suggests that in this dataset, there may not be any non-child descendants of the sensitive attribute, which is required for our criterion to hold.

### G.6 Licenses

In this work, we used a workstation with an AMD Ryzen Threadripper 3990X 64-Core Processor (128 threads in total) with 256 GB RAM and 2x Nvidia RTX 3090 GPUs. However, our simulations only used the CPU resources of the workstation.

We mainly relied on the following Python repositories — (a) networkx (`https://networkx.org`), (b) causal-learn (`https://causal-learn.readthedocs.io/en/latest/`), (c) RCoT [Strobl et al., 2019b] and (d) ridgeCV, (`https://github.com/scikit-learn/scikit-learn/tree/15a949460/sklearn/linear_model/_ridge.py`). We did not modify any of the code under licenses; we only installed these repositories as packages.

In addition to these, we used two public datasets (a) German Credit dataset (`https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data)`) and (b) Adult dataset (`https://archive.ics.uci.edu/ml/datasets/adult`). These datasets are commonly used benchmark datasets for causal fairness, which is why we chose them for our comparisons.

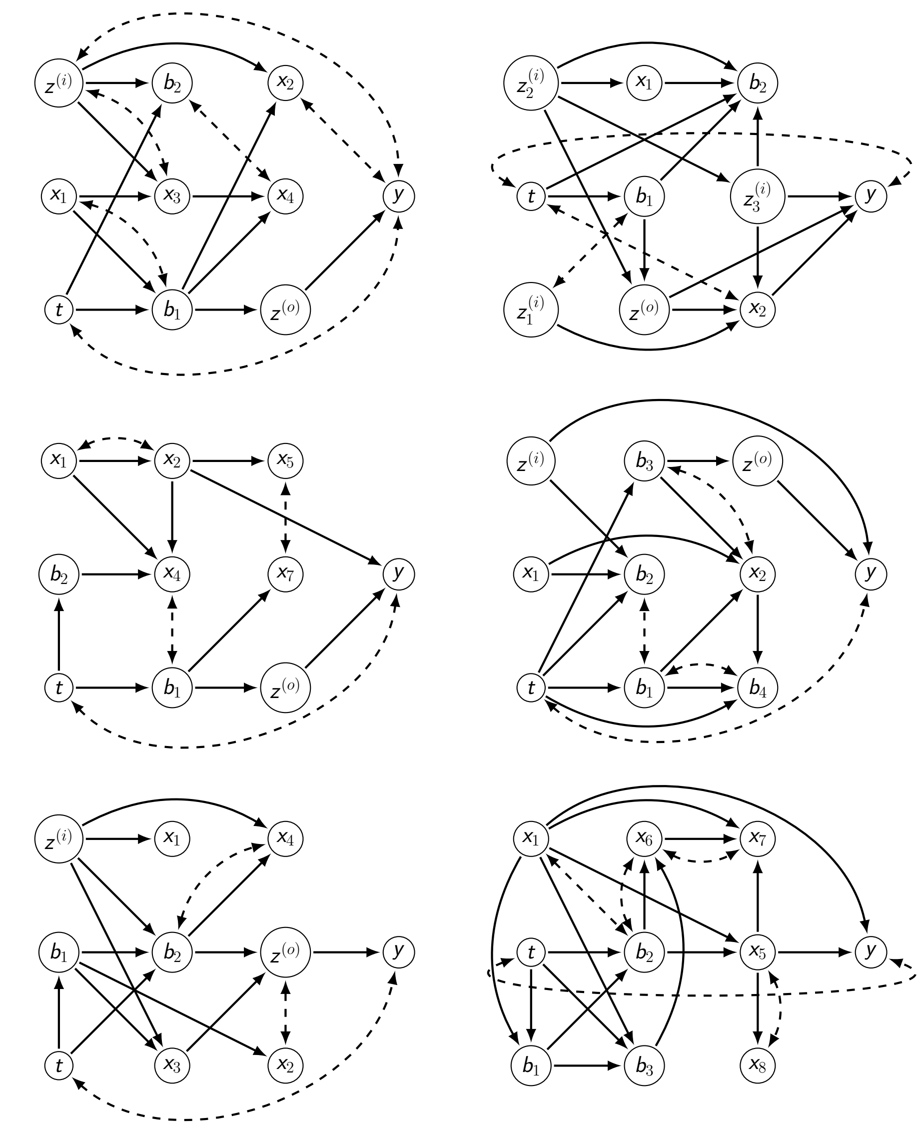


Figure 11: The SMCMs used in Section 4.2 to compare Algorithm 1 with the `Baseline` that uses **b** for front-door adjustment. These are the 6 out of the 100 random graphs in Section 4.1 for $p = 10$, $d = 2$, and $q = 1.0$ where our approach was successful indicating existence of $\mathbf{z} = (\mathbf{z}^{(i)}, \mathbf{z}^{(o)})$ such that the conditional independence statements in Theorem 3.2 hold.