Abstracting Latent Selection in Structural Causal Models

Leihao Chen[1] **Onno Zoeter**[2] **Joris M. Mooij**[1]

[1]Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Amsterdam, the Netherlands
[2]Booking.com, The Netherlands,

Abstract

Selection bias is ubiquitous in real-world data, posing a risk of yielding misleading results if not appropriately addressed. We introduce a conditioning operation on simple Structural Causal Models (SCMs), which is a more general model class than acyclic SCMs, to model latent selection from a causal perspective. We show that the conditioning operation transforms an SCM with the presence of an explicit latent selection mechanism into an SCM (without a selection mechanism) encoding as much causal semantics of the selected subpopulation according to the original SCM as possible. Graphically, in Directed Mixed Graphs we extend the semantics of bidirected edges, which originally represent only latent common causes, to also represent latent selection bias. Furthermore, we show that this conditioning operation preserves the simplicity, acyclicity, and linearity of SCMs, and commutes with marginalization and the conditioning itself. Thanks to these properties, combined with marginalization and intervention, the conditioning operation offers a valuable tool for conducting causal modeling, causal reasoning, and causal model learning tasks within causal models where latent details have been abstracted away. Through illustrative examples, we demonstrate how this abstraction process diminishes the complexity inherent in these three tasks, emphasizing both the theoretical clarity and practical utility of our proposed approach. We hope that our results can deepen the understanding of selection bias from the perspective of SCMs and be integrated into the causal modeling toolbox, ultimately helping modelers develop more reliable and trustworthy causal models.

1 INTRODUCTION

In data analysis, certain challenges persist, particularly in addressing (latent) selection bias. There are many types of selection bias and various methods to address them. In the current work, we focus on the "truncated selection bias", which arises when there is an underlying (unobserved) filtering process that selects individual samples taking some specific values and can be mathematically modeled as conditioning on an event $\{X_S \in \mathcal{S}\}$.

To understand its structural behavior, one approach is to model selection bias via a Causal Model that explicitly describes the selection mechanism, which necessitates a detailed knowledge of the selection mechanism. However, in many situations the selection mechanism is unobserved so that detailed knowledge is often unavailable, which introduces a mysterious dimension with infinitely many possibilities. The goal of the current work is to study how to model latent selection bias by effectively abstracting away its details in a Structural Causal Model (SCM).

Marginalization of causal models is a powerful tool for abstracting away latent details, which makes causal modeling more manageable and trustworthy. By marginalizing out latent variables, we use one simplified model to represent infinitely many complex models, abstracting away unnecessary latent details while preserving important causal information such as causal semantics, d-separations or σ-separations, and ancestral relationships on the observed variables. For example, the model G in Figure 1 effectively abstracts models G^i for $i = 1, \ldots, 5, \ldots$, and one has the same identification results regardless of the latent structure: $\mathrm{P}(C = c \mid \mathrm{do}(S = s)) = \sum_t \mathrm{P}(C = c \mid T = t)\mathrm{P}(T = t \mid S = s)$.

Selection bias is ubiquitous, often latent and can lead to imprecise results, therefore not taking them into account may lead to an untrustworthy model. Now the question is: (i) Is marginalization also able to deal with latent selection bias? (ii) If not, can we always find an SCM without a selection

Accepted for the 1st Workshop on Causal Abstractions and Representations (CAR) at UAI 2025 (CAR 2025).

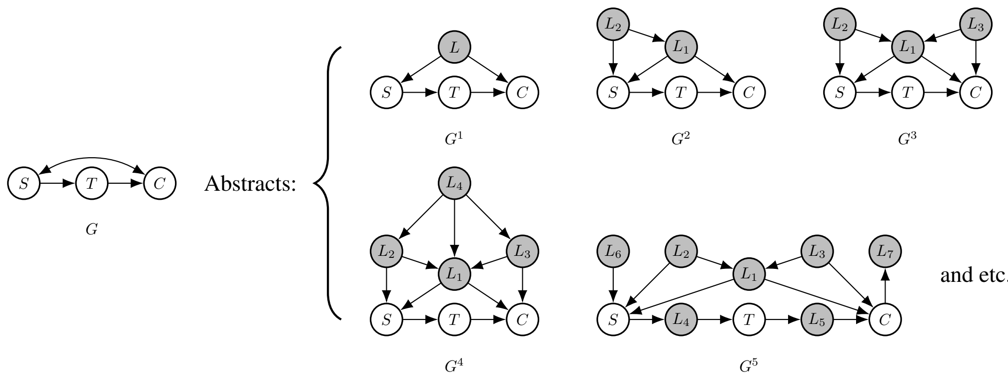

Figure 1: G effectively abstracts G^i for $i = 1, \dots, 5, \dots$

mechanism to represent an SCM with selection faithfully? (iii) If not, can we characterize which part of causal semantics of an SCM with selection mechanism can be represented via an SCM in general? (iv) Given such characterization, can we define transformations both on SCMs and causal graphs constructively? (v) What properties do these transformations have and how do the "conditioned SCMs" $M_{|X_S \in \mathcal{S}}$ and the "conditioned causal graphs" $G(M)_{|S}$ interact with each other? The interaction of marginalized SCM $M_{\setminus L}$ and marginalized causal graph $G(M)^{\setminus L}$ can be partly summarized by Figure 2 (where SCM M has endogenous variables X_V and O denote $V \setminus L$ or $V \setminus S$), can we have a similar diagram for conditioning? We shall give systematic and complete answers to these questions in our work.

To illustrate, we first discuss a toy example, demonstrating that marginalization is not sufficient for tackling selection bias and how to obtain correct results without assuming any specific details about the latent selection mechanism.

Example 1 (Car mechanic) *Cars start successfully if their battery is charged and their start engine is operational. Introduce latent binary endogenous variables B_0 ("battery"), E_0 ("start engine") and S_0 ("car starts") measured at time t_0 and observed variables B_1, E_1 and S_1 with similar meaning for the same car but measured at time t_1 with $t_1 > t_0$. We model this by the following SCM M and denote by M^* its marginalized model over observed endogenous variables.*

$$M : \begin{cases} U_B \sim \mathrm{Ber}(1-\delta), U_E \sim \mathrm{Ber}(1-\epsilon), \\ B_0 = U_B, E_0 = U_E, S_0 = B_0 \wedge E_0, \\ B_1 = B_0, E_1 = E_0, S_1 = B_1 \wedge E_1, \end{cases}$$

$$M^* : \begin{cases} U_B \sim \mathrm{Ber}(1-\delta), U_E \sim \mathrm{Ber}(1-\epsilon), \\ B_1 = U_0, E_1 = U_0, S_1 = B_1 \wedge E_1, \end{cases}$$

where U_B and U_E are latent exogenous independent Bernoulli-distributed random variables with parameters $1-\delta$ and $1-\epsilon$. Their graphs are depicted in Figure 3.

The question is whether there exists an SCM with variables B_1, E_1, S_1 encoding the casual semantics of M under the subpopulation ($S_0 = 0$)? Consider the SCM $\tilde{M}$, whose graph is depicted in Figure 3, given by

$$\tilde{M} : \begin{cases} (U_B, U_E) \sim \mathrm{P}_\theta(U_B, U_E) \\ B_1 = U_B, E_1 = U_E, S_1 = B_1 \wedge E_1 \end{cases}$$

$\mathrm{P}_\theta(U_B, U_E)$	$U_E = 0$	$U_E = 1$
$U_B = 0$	$\frac{\delta\epsilon}{\delta+(1-\delta)\epsilon}$	$\frac{\delta(1-\epsilon)}{\delta+(1-\delta)\epsilon}$
$U_B = 1$	$\frac{(1-\delta)\epsilon}{\delta+(1-\delta)\epsilon}$	0

As one can check,

$$\begin{aligned} \mathrm{P}_{\tilde{M}}(B_1, E_1, S_1) &= \mathrm{P}_M(B_1, E_1, S_1 \mid S_0 = 0) \\ &\neq \mathrm{P}_{M^*}(B_1, E_1, S_1) \\ \mathrm{P}_{\tilde{M}}(S_1 = 1 \mid \mathrm{do}(B_1 = 1)) &= \frac{\delta(1-\epsilon)}{\delta + (1-\delta)\epsilon} \\ &= \mathrm{P}_M(S_1 = 1 \mid \mathrm{do}(B_1 = 1), S_0 = 0) \\ &\neq \mathrm{P}_{M^*}(S_1 = 1 \mid \mathrm{do}(B_1 = 1)) \\ \mathrm{P}_{\tilde{M}}(S_1 = 1 \mid \mathrm{do}(E_1 = 1)) &= \frac{(1-\delta)\epsilon}{\epsilon + \delta(1-\epsilon)} \\ &= \mathrm{P}_M(S_1 = 1 \mid \mathrm{do}(E_1 = 1), S_0 = 0) \\ &\neq \mathrm{P}_{M^*}(S_1 = 1 \mid \mathrm{do}(E_1 = 1)) \end{aligned}$$

So, the car mechanic (who might not even be aware of the latent selection mechanism $S_0 = 0$) can still use an SCM as an accurate causal model to predict the effects of interventions on the subpopulation of cars that are of her concern. Note that the marginalized model does not possess the correct causal semantics of the subpopulation. Besides, graphically the graph $G(\tilde{M})$ correctly express the information that B_1 and E_1 are dependent given $\{S_0 = 0\}$ while the graph $G(M^*)$ wrongly claims that B_1 and E_1 are independent given $\{S_0 = 0\}$ via the d-separation criteria for acyclic directed mixed graphs. Therefore $\tilde{M}$ effectively abstracts away irrelevant latent modeling details: (i) the latent variables B_0, E_0 and S_0, (ii) their causal

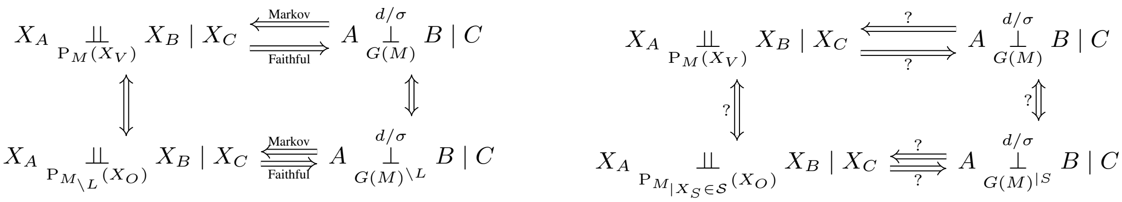

Figure 2: Commutative diagram of conditional indpendencies and graphical separations in causal models.

mechanisms, and (iii) the selection step on $S_0 = 0$. However, the marginalized model M^* does not.

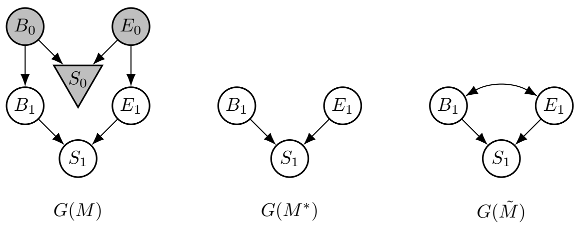

Figure 3: The causal graphs of the SCMs M, M^* and $\tilde{M}$ in Example 1. The gray nodes are latent. Conditioning on $S_0 = 0$ yields $\tilde{M}$. Marginalizing latent variables out yields M^*.

Our Contribution We provide an approach to model selection bias by *effectively abstracting away* latent selection mechanisms. To be more precise, given a Structural Causal Model M with an selection mechanism $\{X_S \in \mathcal{S}\}$ where variable X_S takes values in a measurable subset $\mathcal{S}$, we define a transformation that maps $(M, \{X_S \in \mathcal{S}\})$ to a "conditioned" SCM $M_{|X_S \in \mathcal{S}}$ without any selection mechanism, so that $M_{|X_S \in \mathcal{S}}$ is an effective abstraction of M w.r.t. the selection $\{X_S \in \mathcal{S}\}$ in the sense that:

- the conditioned SCM $M_{|X_S \in \mathcal{S}}$ correctly encodes as much *causal semantics* (observational, interventional and counterfactual) of M under the *subpopulation* as possible;
- this conditioning operation interacts well with other operations on SCMs, e.g., intervention, the conditioning operation itself, and marginalization;
- the conditioning operation preserves important model classes, e.g., linear, acyclic and simple SCMs;
- one can read off enough *qualitative causal information* about M under the selected subpopulation from the *causal graph* of $M_{|X_S \in \mathcal{S}}$ and the conditioned graph $G(M)_{|S}$.

In our work, we introduce the rigorous mathematical definition of the conditioning operation and demonstrate that it possesses all the aforementioned properties.

The significance of this conditioning operation lies in the fact that we can take $M_{|X_S \in \mathcal{S}}$ as a simplified "proxy" for M w.r.t. the selection $\{X_S \in \mathcal{S}\}$, which effectively abstracts away details about latent selection (i.e., satisfying the properties listed previously). This makes it a versatile tool for causal inference confronted with latent selection bias. Specifically:

(i) Causal Reasoning: One can directly apply all the causal inference tools on SCMs to $M_{|X_S \in \mathcal{S}}$, e.g., adjustment criterion, Pearl's do-calculus and ID-algorithm, which simplifies causal reasoning tasks without sacrificing generality.

(ii) Causal Modeling: Utilizing the marginalization and conditioning operation, we can represent infinitely many SCMs with a single marginalized conditioned SCM. This significantly streamlines causal modeling, eliminating the need to enumerate all the possibilities with different latent (un)conditioned structures. Moreover, it enhances model robustness and trustworthiness by reducing sensitivity to various causal assumptions.

(iii) Causal Model Learning: In many learning algorithm, selection bias is ruled out by assumption, which is often not a realistic setting. One can use usual algorithm to learn $M_{|X_S \in \mathcal{S}}$ instead of M, which can deal with selection bias automatically and reduces the complexity of the learning process.

Connections to Related Work In a series of papers [Bareinboim and Pearl, 2012, Bareinboim and Tian, 2015], the authors explored the 's-recoverability' problem, aiming to recover causal quantities for the whole population from selected data. This investigation operated under qualitative causal assumptions on the selection nodes, explicitly expressed in terms of causal graphs. However, such knowledge about selection nodes is not always available [Richardson and Robins, 2013, Footnote 11]. In the current work, we focus on the problem of how to *model selection bias with an SCM without explicitly modeling the selection mechanism* and draw (causal) conclusions for the selected subpopulation.

There are graphical models with well-behaved marginalization and conditioning operations such as maximal ancestral graphs (MAGs) [Richardson and Spirtes, 2002], d-connection graphs [Hyttinen et al., 2014] and σ-connection graphs [Forré and Mooij, 2018]. Among them, MAGs were originally developed as a smallest model class containing the conditional independence models of the marginalized conditioned conditional independence models of DAGs. By summarizing the common causal features of causal DAGs represented by a MAG, one can give a causal interpretation to MAGs and call them causal MAGs. One single causal MAG can represent infinitely many SCMs with different graphs but the same conditional independences among observed variables. Interpreting a graph as a causal graph of an SCM and as a causal MAG respectively will not give the same causal conclusions in general.[1] Due to the nature of model abstraction, MAGs are well suited for causal discovery, and one can further draw *some* causal conclusions from MAGs [Spirtes et al., 1995b, Richardson, 2003, Zhang, 2008, Mooij and Claassen, 2020]. However, MAGs are not always suitable for *causal modeling* under selection bias in some cases, since: (i) it is not clear how to read off causal relationships (direct causal relations, confounding) from MAGs; (ii) there are no identification results for MAGs under selection bias yet; (iii) currently the standard theory of MAGs cannot deal with causal cycles and counterfactual reasoning. On the other hand, our conditioning operation transforms an SCM with selection mechanisms to an ordinary SCM, which carries an intuitive causal interpretation. All the theory for SCMs (causal identification, cycles, counterfactual reasoning) can be directly extended to the case with selection bias via the conditioning operation. Therefore, our results can address causal inference tasks such as fairness analysis [Kusner et al., 2017, Zhang and Bareinboim, 2018], causal modeling of dynamical systems [Bongers et al., 2022, Peters et al., 2022] and biological systems with feedback loops [Versteeg et al., 2022] under selection bias. Another subtle difference between SCM conditioning and MAG conditioning is that they consider different forms of conditioning.

Although causal graphs provide a means to differentiate selection bias from confounding due to common causes [Hernán et al., 2004, Cooper, 1995], the potential outcome community tends to amalgamate the two [Richardson and Robins, 2013, Hernán MA, 2020]. In many cases, one can be sure about the existence of "non-causal dependency", but cannot be sure whether it is induced by a latent common cause or latent selection bias or the combination of the two (see e.g., Richardson and Robins [2013, Footnote 11] and Pearl [2009, p.163]). Our conditioning operation formalizes this ambiguity within SCMs. Graphically, we employ bidirected edges to symbolize the dependence of two variables arising from either unmeasured common causes, latent selection bias, or any intricate combination of the two. Therefore, in causal modeling, our work allows the modeler to be able to represent such non-causal dependency abstractly via bidirected edges.

Some work considers the abstraction of causal models from the perspective of grouping low-level variables to high-level variables and merging values of variables [Rubenstein et al., 2017, Beckers and Halpern, 2019]. Geiger et al. [2023] study the so-called "constructive abstraction" of causal models. They show that it can be characterized as a composition of clustering sets of variables, merging values of variables, and marginalization. Our conditioning operation does not fall under the umbrella of "constructive abstraction" of Geiger et al. [2023].

2 CONDITIONING OPERATION ON SCMS

In Section 2.1, we define the conditioning operation on simple SCMs and present some discussions about the definition in Appendix. We shall derive some properties of it in Section 2.2, and the proofs can be found in Appendix. We make some important caveats on how to interpret the conditioned SCMs when modeling in Appendix E. We follow the formal setup of Bongers et al. [2021]. In the whole section, we assume:

Assumption 1 *$M = (V, W, \mathcal{X}, \mathrm{P}, f)$ is a simple SCM such that $\mathrm{P}_M(X_S \in \mathcal{S}) > 0$ for some $S \subseteq V$ and measurable subset $\mathcal{S} \subseteq \mathcal{X}_S$.*

We write $O := V \setminus S$. We use $\mathrm{P}_M(X_O \mid \mathrm{do}(X_T = x_T), X_S \in \mathcal{S}) := \mathrm{P}_{M_{\mathrm{do}(X_T=x_T)}}(X_O \mid X_S \in \mathcal{S})$ to represent the probability distribution of X_O when first intervening on $X_T = x_T$ and second conditioning on $X_S \in \mathcal{S}$.[2]

2.1 DEFINITION OF CONDITIONING OPERATION

Suppose that we condition on the event $\{X_S \in \mathcal{S}\}$. Then roughly speaking, the conditioning operation can be divided into three steps:

1. merging all the exogenous random variables that are ancestors of the selection variables;
2. updating the exogenous probability distribution to the posterior given the observation $X_S \in \mathcal{S}$;
3. marginalizing out the selection variables.

[1] For example, consider a graph consisting of $A \to B \to C$ and $A \to C$. If it is a causal graph of an SCM, then we can conclude that variable A has a direct causal effect on C according to this model and we can identify $\mathrm{P}(C = c \mid \mathrm{do}(A = a)) = \mathrm{P}(C = c \mid A = a)$ under the positivity and discreteness assumption. However, if it is a MAG, then we cannot obtain the above two conclusions.

[2] "First" and "second" here refer to the order of applying the operations on the SCM.

We give the formal definition of the conditioned SCMs specializing to the class of simple SCMs for simplicity. See Figure 4 for an intuitive graphical representation.[3]

Definition 2 (Conditioned SCM) *Assume Assumption 1. Write* $B := \mathrm{Anc}_{G^a(M)}(S)$. *Let* $g^S : \mathcal{X}_W \times \mathcal{X}_O \to \mathcal{X}_S$ *be the essentially unique solution function of* M *w.r.t.* S. *We define the* ***conditioned SCM*** $M_{|X_S \in \mathcal{S}} := \left(\hat{V}, \hat{W}, \hat{\mathcal{X}}, \hat{\mathrm{P}}, \hat{f}\right)$ *by:*

- $\hat{V} := V \setminus S$;
- $\hat{W} := (W \setminus B) \dot\cup \{\star_W\}$ *with* $\star_W := B \cap W$;
- $\hat{\mathcal{X}} := \mathcal{X}_O \times \hat{\mathcal{X}}_{\hat{W}} := \mathcal{X}_O \times \left(\mathcal{X}_{W \setminus B} \times \mathcal{X}_{\star_W}\right)$, *where* $\mathcal{X}_{\star_W} := \mathcal{X}_{W \cap B}$;
- $\hat{\mathrm{P}} := \mathrm{P}(X_{W \setminus B}) \otimes \mathrm{P}(X_{\star_W})$, *where* $\mathrm{P}(X_{\star_W}) := \mathrm{P}_M(X_{W \cap B} \mid X_S \in \mathcal{S})$;
- $\hat{f}(x_{\hat{V}}, x_{\hat{W}})$ $:= f_O(x_O, g^S(x_O, x_{W \setminus B}, x_{\star_W}), x_{W \setminus B}, x_{\star_W})$.

It is easy to check that $M_{|X_S \in \mathcal{S}}$ is indeed an SCM.

Notation 3 *We often denote* $M_{|X_S \in \mathcal{S}}$ *by* $M_{|\mathcal{S}}$ *if it is clear from the context that* $\mathcal{S}$ *is a measurable subset in which the variable* X_S *takes values.*

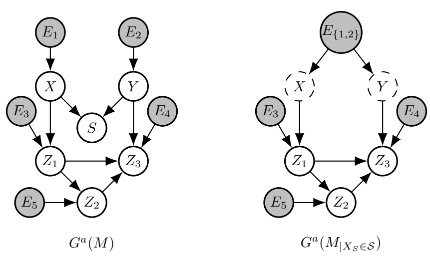

Figure 4: Graphical representation of conditioning on $X_S \in \mathcal{S}$. First merge the exogenous ancestors of S, i.e., E_1 and E_2, to get a merged node $E_{\{1,2\}}$. Then update the exogenous probability distribution $\mathrm{P}(X_{E_1}, X_{E_2})$ to the posterior $\mathrm{P}_M(X_{E_1}, X_{E_2} \mid X_S \in \mathcal{S})$. Finally, marginalize out the node S. X and Y are dashed, since we mark them as non-intervenable.

Note that the above conditioning operation is defined on simple SCMs. For causal modeling purposes, people often use causal graphs to communicate causal knowledge without referring to the underlying SCMs. To support this, we give a purely graphical conditioning operation defined on directed mixed graphs (DMGs).

[3] In the graphical representations of the conditioning operation such as Figure 4, we assume no causal effects canceled out because of marginalization or changing the underlying population.

Definition 4 (Conditioned DMG) *Let* $G = (V, E, H)$ *be a DMG consisting of nodes* V, *directed edges* E *and bidirected edges* H. *For* $S \subseteq V$, *we define the conditioned DMG as*

$$G_{|S} = \left(V_{|S}, E_{|S}, H_{|S}\right) :$$

- $V_{|S} := V \backslash S$ *with* $\mathrm{Anc}_G(S) \setminus S$ *dashed;*
- $E_{|S}$ *consists of all* $v \to u$ *with* $v, u \in V \setminus S$ *and* $v \neq u$ *for which there exists a directed walk in G:* $v \to s_1 \to \cdots \to s_n \to u$, *where all intermediate nodes* $s_1, \ldots, s_n \in S$ *(if any);*
- $H_{|S}$ *consists of all bidirected edges* $v \leftrightarrow u$ *with* $v, u \in V \setminus S$ *and* $v \neq u$, *for which there exists a bifurcation in G:* $v \leftarrow s_1 \leftarrow \cdots \leftarrow s_{k-1} \leftarrow\!\!* \; s_k \to \cdots \to s_n \to u$ *with all intermediate nodes* $s_1, \ldots, s_n \in S$ *(if any), or for which* $v, u \in \mathrm{Anc}_G(S) \cup \mathrm{Sib}_G(\mathrm{Anc}_G(S))$.[4]

We give an example in Appendix D.1. As we will show in the next subsection, the purely graphical conditioning operation interacts well with the SCM conditioning operation.

2.2 PROPERTIES OF CONDITIONING OPERATION

The conditioning operation ensures the preservation of simplicity, linearity, and acyclicity in SCMs.

Proposition 5 (Simplicity/Acyclicity/Linearity) *If* M *is a simple (resp. acyclic) SCM with conditioned SCM* $M_{|X_S \in \mathcal{S}}$, *then the conditioned SCM* $M_{|X_S \in \mathcal{S}}$ *is simple (resp. acyclic). If* M *is also linear, then so is* $M_{|X_S \in \mathcal{S}}$.

This implies that opting for simple/acyclic/linear SCMs as a model class, and performing model abstraction through the conditioning operation, will consistently maintain one within the chosen model class. This convenience proves valuable in practical applications, where adherence to specific model class is often desired.

The following lemma states that the conditioning commutes with interventions on the non-ancestors of the conditioning variables.

Lemma 6 (Conditioning & intervention) *Assume Assumption 1. Then we have* $\left(M_{\mathrm{do}(X_T = x_T)}\right)_{|X_S \in \mathcal{S}} = \left(M_{|X_S \in \mathcal{S}}\right)_{\mathrm{do}(X_T = x_T)}$ *for any* $T \subseteq O \setminus \mathrm{Anc}_{G^a(M)}(S)$ *and* $x_T \in \mathcal{X}_T$.

Remark 7 *Since* M *is simple and* $T \subseteq O \setminus \mathrm{Anc}_{G^a(M)}(S)$, *the probability* $\mathrm{P}_{M_{\mathrm{do}(X_T = x_T)}}(X_S \in \mathcal{S}) = \mathrm{P}_M(X_S \in \mathcal{S})$ *is well defined and strictly larger than zero.*

[4] $\mathrm{Sib}_G(v) := \{w \in G \mid v \leftrightarrow w \in G\}$, and $s_{k-1} \leftarrow\!\!* \; s_k$ means either $s_{k-1} \leftarrow s_k$ or $s_{k-1} \leftrightarrow s_k$.

The next presented theorem establishes that the conditioned SCM faithfully encapsulates the conditional observational distribution of every endogenous variable and the conditional causal semantics of the non-ancestors of S, in accordance with the original SCM.

Theorem 8 (Preserving causal semantics) *Assume Assumption 1. Then we have*

(1) $\mathrm{P}_{M_{|X_S\in\mathcal{S}}}(X_O) = \mathrm{P}_M(X_O \mid X_S \in \mathcal{S})$*;*

(2) for any $T \subseteq V \setminus \mathrm{Anc}_{G^a(M)}(S)$ *and* $x_T \in \mathcal{X}_T$*,*

$$\mathrm{P}_{M_{|X_S\in\mathcal{S}}}\left(X_{O\setminus T} \mid \mathrm{do}(X_T = x_T)\right) = \mathrm{P}_M\left(X_{O\setminus T} \mid \mathrm{do}(X_T = x_T), X_S \in \mathcal{S}\right);$$

(3) for any $T_1 \subseteq V \setminus \mathrm{Anc}_{G^a(M)}(S)$ *and* $x_{T_1} \in \mathcal{X}_{T_1}$*, and any* $T_2 \subseteq (V \setminus \mathrm{Anc}_{G^a(M)}(S))'$ *and* $x_{T_2} \in \mathcal{X}_{T_2}$*,*

$$\begin{aligned}&\mathrm{P}_{(M_{|X_S\in\mathcal{S}})^{\mathrm{twin}}}(X_{(O\cup O')\setminus(T_1\cup T_2)} \mid \\ &\qquad\qquad \mathrm{do}(X_{T_1} = x_{T_1}, X_{T_2} = x_{T_2})) \\ &= \mathrm{P}_{M^{\mathrm{twin}}}(X_{(O\cup O')\setminus(T_1\cup T_2)} \mid \\ &\qquad\qquad \mathrm{do}(X_{T_1} = x_{T_1}, X_{T_2} = x_{T_2}), X_S \in \mathcal{S}).\end{aligned}$$

This result assures that the simplified abstracted model retains the capacity to yield identical results as the original more intricate model, thereby solidifying the foundation for the effectiveness of the proposed abstraction process.

The following corollary implies that different orderings of iterative conditioning operations give rise to counterfactually equivalent SCMs.

Corollary 9 *Assume Assumption 1 with* $S = S_1 \cup S_2$ *and* $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2$ *with* $\mathcal{S}_1 \subseteq \mathcal{X}_{S_1}$ *and* $\mathcal{S}_2 \subseteq \mathcal{X}_{S_2}$ *both measurable. Then* $(M_{|\mathcal{S}_1})_{|\mathcal{S}_2}$*,* $(M_{|\mathcal{S}_2})_{|\mathcal{S}_1}$*, and* $M_{|\mathcal{S}_1\times\mathcal{S}_2}$ *are counterfactually equivalent w.r.t.* $V \setminus \mathrm{Anc}_{G^a(M)}(S_1 \cup S_2)$*.*[5]

The subsequent result establishes the commutativity of conditioning and marginalization.

Proposition 10 (Conditioning & marginalization) *Assume Assumption 1 and let* $L \subseteq V \setminus S$*. Then we have that* $(M_{\setminus L})_{|\mathcal{S}}$ *and* $(M_{|\mathcal{S}})_{\setminus L}$ *are counterfactually equivalent.*

Recall that marginalization commutes with itself and that the order of conditioning operations does not matter up to counterfactual equivalence. Given a set of latent unconditioned variables and latent conditioned variables, irrespective of the intermediate steps taken, one consistently arrives at counterfactually equivalent models in practice by using marginalization and the conditioning operation to abstract away latent details. This underscores the robustness and reliability of the overall procedure.

[5] See Definition 25 or Bongers et al. [2021, Definition 4.5] for the definition of counterfactual equivalence.

The following proposition states that the purely graphical conditioning operation is compatible with the SCM conditioning operation.

Proposition 11 (Conditioned SCM & DMG) *Let* M *be a simple SCM with conditioned SCM* $M_{|X_S\in\mathcal{S}}$*. Then* $G(M_{|X_S\in\mathcal{S}})$ *is a subgraph of* $G(M)_{|S}$*.*

Remark 12 *Note that* $G(M_{|X_S\in\mathcal{S}})$ *can be a strict subgraph of* $G(M)_{|S}$*. This means that* $G(M)_{|S}$ *is generally a more conservative representation of the underlying SCM with less causal information due to the nature of abstraction.*

Markov properties connect the causal graph and the induced distribution of M in the sense that they enable one to read off conditional independence relations from the graph via the d- or σ-separation criterion (see Appendix A or Bongers et al. [2021]). Obviously, $\mathrm{P}_{M_{|X_S\in\mathcal{S}}}(X_O)$ satisfies this property relative to $G(M_{|X_S\in\mathcal{S}})$. Notably, $\mathrm{P}_{M_{|X_S\in\mathcal{S}}}(X_O)$ also satisfies this property relative to $G(M)_{|S}$, illustrating the role of the conditioned graph $G(M)_{|S}$ as an effective graphical abstraction.

Corollary 13 (Markov property) *If* M *is simple, then* $\mathrm{P}_{M_{|X_S\in\mathcal{S}}}(X_O)$ *satisfies the generalized directed global Markov property relative to* $G(M)_{|S}$*. If* M *is acyclic, then* $\mathrm{P}_{M_{|X_S\in\mathcal{S}}}(X_O)$ *satisfies the directed global Markov property relative to* $G(M)_{|S}$*.*

3 SOME APPLICATIONS

The conditioning operation has a wide spectrum of applications. All the classical results for SCMs, such as identification results (the back-door adjustment, do-calculus), can be applied to conditioned SCMs $M_{|X_S\in\mathcal{S}}$ immediately. Using the properties of the conditioning operation, we can then translate these conclusions on $M_{|X_S\in\mathcal{S}}$ back to $(M, X_S \in \mathcal{S})$. Combining with marginalization, the conditioning operation also provides a way to interpret a DMG as a causal graph that compactly encodes causal assumptions, where latent details of both latent common causes and latent selection have been abstracted away.

For illustration, we briefly discuss several examples in this section. They form a cohesive sequence navigating us from philosophical implications of the conditioning operation ("generalized Reichenbach's principle"), to the versatility of applications of classical results to conditioned SCMs (back-door criterion, instrumental variables, ID-algorithm, mediation analysis), and finally concrete practical application of conditioned SCMs in modeling real-world problems (COVID example). For the sake of space, some examples are given in Appendix.

Example 2 (Reichenbach's principle) *Reichenbach's Principle of Common Cause [Reichenbach, 1956] is often stated in this way: if two variables are dependent, then one must cause the other, or the variables must have a common cause (or any combination of these three possibilities). Note that this conclusion holds only when latent selection bias is ruled out, an assumption that is often left implicit.*

With our conditioning operation, we can generalize it in the following way. Assume that M is a simple SCM that has two observed endogenous variables X and Y. By the Markov property, if X and Y are dependent, then $X \rightarrow Y$, $X \leftarrow Y$, or $X \leftrightarrow Y$ (or any combination of these three possibilities) are in the graph $G(M)$. There exist infinitely many SCMs M^i, $i \in I$ with an infinite index set I, such that $(M^i_{\setminus L_i})_{|\mathcal{S}_i} = M$ where L_i is a set of latent variables of M^i and $X_{S_i} \in \mathcal{S}_i$ is the latent selection in M^i. Hence, it implies that if two variables are dependent, then one causes the other, or the variables have a common cause, or are subject to latent selection (or any combination of these four possibilities).

This provides one possible explanation for some real-world scenario in which one can exclude the possibilities of causal effects and common causes between two variables but can still observe the stochastic dependency between them.

Example 3 (Back-door theorem) *Let M^1 and M^2 be two SCMs with three variables T ("treatment"), X ("covariates"), and Y ("outcome") whose causal graphs are shown in Figure 5. Under some assumptions, Pearl's Back-Door Theorem [Pearl, 2009] gives, for $i = 1, 2$, the identification result:*[6]

$$\begin{aligned} &\mathrm{P}_{M^i}(Y \mid \mathrm{do}(T = t)) \\ &= \int \mathrm{P}_{M^i}(Y \mid X = x, T = t)\mathrm{P}_{M^i}(X \in \mathrm{d}x). \end{aligned} \tag{1}$$

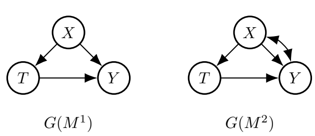

Figure 5: Causal graphs of SCMs M^1 and M^2 in Example 3.

Thanks to marginalization and the conditioning operation, we can see M^1 and M^2 as abstractions of other SCMs, i.e., $M^i = (\tilde{M}^i_{\setminus L^i})_{|\mathcal{S}^i}$, for SCMs $\tilde{M}^i$, latent variables $L^i = \{L^i_1, \ldots, L^i_n\}$, and latent selection variables $S^i = \{S^i_1, \ldots, S^i_m\}$ taking values in measurable sets $\mathcal{S}^i$ with $i = 1, 2$. For both M^1 and M^2, we present two examples $\tilde{M}^i_{(j)}$ for $j = 1, 2$, respectively, out of the infinite possibilities in Figure 6.

We can write (1) *as*

$$\begin{aligned} &\mathrm{P}_{\tilde{M}^i}(Y \mid \mathrm{do}(T = t), S^i \in \mathcal{S}^i) \\ &= \int \mathrm{P}_{\tilde{M}^i}(Y \mid X = x, T = t, S^i \in \mathcal{S}^i)\mathrm{P}_{\tilde{M}^i}(X \in \mathrm{d}x \mid S^i \in \mathcal{S}^i). \end{aligned} \tag{2}$$

Thus, the back-door theorem can be applied directly to the conditioned SCM, which is useful especially if the specific latent structure of the SCM is unknown.

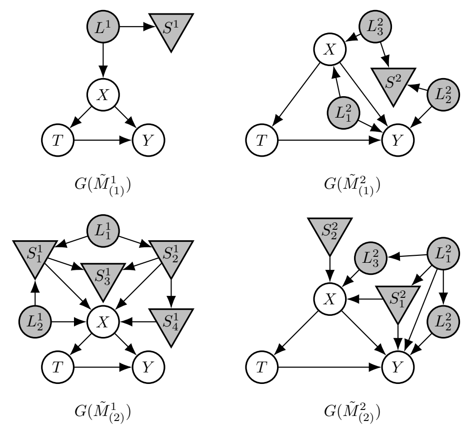

Figure 6: Some possible causal graphs of SCMs $\tilde{M}^i$ in Example 3.

One can generalize other identification results similarly.

Example 4 (ID-algorithm) *Pearl's do-calculus is proved to be sound and complete (under some conditions) for identifying interventional distributions given a causal graph [Pearl, 1995a, Huang and Valtorta, 2006]. Using a causal graph and observational distribution as inputs, the ID-algorithm, as a sound and complete algorithm, systematically outputs a functional of the observational distribution to the target interventional distribution if it is identifiable and outputs FAIL if it is not [Tian and Pearl, 2002, Shpitser and Pearl, 2006, Huang and Valtorta, 2008]. Various variants of the ID-algorithm exist, each with different targets and inputs (see e.g., Kivva et al. [2023] and the references therein).*

One such variant, the s-ID-algorithm, is a sound and complete algorithm for the s-identification problem*, whose goal is to identify interventional distributions on a subpopulation (*$\mathrm{P}(X_A \mid \mathrm{do}(X_T = x_T), X_S = 1)$*) given a causal graph with selection mechanism (*G^s*) and selected observational distribution (*$\mathrm{P}(X_V \mid X_S = 1)$*) [Abouei et al., 2024a, Theorem 1, Corollary 2].*[7] *As we shall see, the conditioning*

[6] For simplicity, here we ignore the measure-theoretic subtlety. Indeed, we need to assume $\mathrm{P}_{M^i}(X) \otimes \mathrm{P}_{M^i}(T) \ll \mathrm{P}_{M^i}(X, T)$ and then the identity holds $\mathrm{P}_{M^i}(T)$-a.s.

[7] Note that in the usual c-ID-algorithm for conditional interventional distribution, the input is $\mathrm{P}(X_V)$ but not $\mathrm{P}(X_V \mid X_S = 1)$.

operation can help simplify some parts of the original proof.

For simplicity, we only consider the single-variable case via the conditioning operation. In the setting of Abouei et al. [2024a], there are no latent variables. Therefore, if $T = \{t\} \cap \mathrm{Anc}_{G^s}(S) = \emptyset$*, then there are no bidirected edges connecting to* t *in* $G^s_{|S}$*, which implies that* $\mathrm{P}(X_A \mid \mathrm{do}(X_T = x_T), X_S = 1)$ *is identifiable by Tian and Pearl [2002, Theorem 1]. Now, assuming that* $A \underset{G^s_{\underline{X}}}{\overset{d}{\perp}} T \mid S$*, the second rule of Pearl's do-calculus provides the identification result. Combining these two gives a sound and complete algorithm for the s-identification problem [Abouei et al., 2024a, Theorem 1].*

Besides, if $T \cap \mathrm{Anc}_{G^s}(S) = \emptyset$*, then one can also consider identifying the conditional causal effect on the subpopulation* $\mathrm{P}(X_A \mid \mathrm{do}(X_T = x_T), X_B, X_S = 1)$ *from a graph with latent variables and selected observation distribution* $\mathrm{P}(X_V \mid X_S = 1)$*, by first applying the conditioning operation on* G^s *to get* $G^s_{|S}$ *and then applying the classical ID-algorithm on* $G^s_{|S}$ *for conditional causal effect with latent variables.*[8] *This result seems to be new in the literature to our knowledge.*[9] *Similar generalizations can be made for other variants of the ID-algorithm, by first applying the conditioning operation on the graph and then applying the corresponding version of the ID-algorithm to the conditioned graph.*

However, one should note that applying the ID-algorithm to the conditioned graph alone can hardly give a complete algorithm in general, due to the abstraction nature of the conditioning operation. For example, in the case of the s-ID-algorithm, we can use the conditioning operation to handle cases where $T \cap \mathrm{Anc}_{G^s}(S) = \emptyset$*, but a complete algorithm should also be able to address cases where* $T \cap \mathrm{Anc}_{G^s}(S) \neq \emptyset$ *or* $T \underset{G^s_{\underline{X}}}{\overset{d}{\perp}} A \mid S$ *(see Abouei et al. [2024a, Theorem 1]), which can be tackled by combining with the second rule of Pearl's do-calculus.*

Example 5 (Causal discovery) *Many causal discovery algorithms address unobserved common causes, exclude selection bias, and output a single graph. In fact, we can interpret the output of such algorithms as* $G((M_{\setminus L})_{|\mathcal{S}})$ *where* M *is a simple (or acyclic) SCM with latent nodes* L*, selection mechanism* $X_S \in \mathcal{S}$*, and* $L \cap S = \emptyset$*. This can give a certain causal interpretation to the output of these algorithms under selection bias even if they exlcude selection bias in their original formulations.*

For one instance, Wang and Drton [2023] explored recovering causal graphs uniquely from data generated by an acyclic linear non-Gaussian SCM with a bow-free graph (i.e., no simultaneous bidirected and directed edges between two variables) and rule out selection bias. Assume that the data are generated from an acyclic linear SCM M *and there is no latent common cause or selection bias between any two variables that have a direct causal effect according to* M*. Then, according to the properties of marginalization and the conditioning operation,* $(M_L)_{|\mathcal{S}}$ *is an acyclic linear SCM with a bow-free graph (see Proposition 5 and Bongers et al. [2021, Proposition 5.11, C.5]). If the exogenous distribution of* $(M_L)_{|\mathcal{S}}$ *is non-Gaussian, then we can use the algorithm* BANG *in Wang and Drton [2023] to recover the graph of* $(M_{\setminus L})_{|\mathcal{S}}$*.*

If we know from data or prior knowledge that a node t *is not an ancestor of* S*, then we can give a causal interpretation of* X_t *in the discovered graph and apply causal identification results to identify* $\mathrm{P}_M(X_O \mid \mathrm{do}(X_t = x_t), X_S \in \mathcal{S})$ *with* $O := V \setminus (L \cup S)$*. For example, if the data are selected by* $X_S = x_S$*, we can sometimes read off whether* $t \notin \mathrm{Anc}_{G(M)}(S)$ *from a PAG (Partial Ancestral Graphs) or a MAG [Spirtes et al., 1995a, Richardson and Spirtes, 2002].*[10]

Example 6 (Causal modeling) *In this example, we show how the conditioning operation can help with causal modeling under selection bias.*[11] *The high-level idea is from Example 2 that even if there are no causal effects and no common causes between two variables there could still be dependency between them caused by selection bias.*

To state the example, recall that one possible workflow of causal inference is:

(i) *asking causal queries;*

(ii) ***building a causal model*** *from prior knowledge and data;*

(iii) *determining the target causal quantity and identifying the estimand in terms of available observational and interventional distributions;*

(iv) *using data to estimate the estimand.*

[8] If $T \cap \mathrm{Anc}_{G^s}(S) \neq \emptyset$, one can still apply the corresponding ID-algorithm to $G^s_{|S}$, but the algorithm would output an expression for $\mathrm{P}(X_A(x_T) \mid X_S = 1)$ instead of $\mathrm{P}(X_A \mid \mathrm{do}(X_T = x_T), X_S = 1)$.

[9] When we were writing this manuscript, we found that an s-ID-algorithm under latent variables was proposed in Abouei et al. [2024b]. However, they only consider identification for the unconditional interventional distribution $\mathrm{P}(X_A \mid \mathrm{do}(X_T = x_T), X_S = 1)$ not for the conditional interventional distribution $\mathrm{P}(X_A \mid \mathrm{do}(X_T = x_T), X_B, X_S = 1)$.

[10] Note that if $t \in \mathrm{Anc}_{G(M)}(S)$, we can still apply the identification result to the interventional distribution given $\mathrm{do}(X_t = x_t)$ in $M_{|X_S \in \mathcal{S}}$, but the causal identification results will output a formula for $\mathrm{P}_M(X_O(x_t) \mid X_S \in \mathcal{S})$ instead of $\mathrm{P}_M(X_O \mid \mathrm{do}(X_t = x_t), X_S \in \mathcal{S})$.

[11] How to perform causal modeling under selection bias is the original motivation for this work.

As concise encodings of causal assumptions, causal graphs can be used to decide the estimand for addressing causal queries, and therefore incorrect graphs may generate wrong results. For example, to understand the causal effect of treatment strategies from different countries on the fatality rate of the COVID-19 case, Von Kügelgen et al. [2021] analyzed data from the initial virus outbreaks in 2020 in China and Italy, and assumed the causal graph G shown in Figure 7. For COVID-19 infected people, age (A), country of residence (C) at the time of infection and fatality rate (F) are recorded.

The data suggest that C and A are dependent. In the traditional understanding of bidirected edges, assuming that C and A do not share a latent common cause, one has to draw a directed edge between C and A so that the hypothesized graph is compatible with the observation. However, drawing a directed edge from C to A is not a reasonable causal assumption. It assumes that if we conduct a randomized trail to assign people to different countries, then immediately (A and C are measured almost the same time) the resulting age distribution will differ depending on the assigned country. Similarly, $A \rightarrow C$ would also be an unreasonable assumption.

However, the conditioning operation tells us that bidirected edges do not have to represent latent common causes only, but can also represent latent selection bias. Therefore, we can draw a bidirected edge $C \leftrightarrow A$ as shown in $\tilde{G}$ to explain the statistical association between C and A, which could represent different latent selection mechanisms or latent common causes or combinations of the two between C and A.[12] *First, the age distribution may differ between two countries already before the outbreak of the virus (latent selection on 'person was alive ($S' = 1$) in early 2020', as in G^1). Second, since only* infected *patients were registered and both the country and the age may influence the risk of getting infected, selection of the infection status ($S = 1$) can also lead to $C \leftrightarrow A$ (as in G^2). The combinations of both selection mechanisms (such as in G^3 or G^4) also lead to $C \leftrightarrow A$. With the conditioning operation, we do not need to list (potentially infinitely many) all the possible causal graphs in detail, including* all *relevant latent variables that model the selection mechanism. We only need to consider DMGs on these three observed variables, which is a much smaller (finite) model space.*

Thanks to properties of the conditioning operation, we can answer causal queries like "what would be the effect on fatality of changing from China to Italy". It allows us to compute the total causal effect $\mathrm{TCE}(Y; c' \to c) := \mathbb{E}[F \mid \mathrm{do}(C = c)] - \mathbb{E}[F \mid \mathrm{do}(C = c')]$ *via the abstracted (conditioned) model $\tilde{G}$ (e.g., by adjusting on age) without fully knowing all the latent details. Note that the results based on G and $\tilde{G}$ are clearly different. In fact, for an SCM with graph G, one has:*

$$\mathrm{TCE}(Y; c' \to c) := \mathbb{E}[F \mid \mathrm{do}(C = c)] - \mathbb{E}[F \mid \mathrm{do}(C = c')].$$

On the other hand, for an SCM with graph $\tilde{G}$, one has:

$$\begin{aligned}\mathrm{TCE}(Y; c' \to c) &:= \mathbb{E}[F \mid \mathrm{do}(C = c)] - \mathbb{E}[F \mid \mathrm{do}(C = c')] \\ &\neq \sum_a (\mathbb{E}[F \mid C = c, A = a]\mathrm{P}(A = a \mid C = c) \\ &\quad - \mathbb{E}[F \mid C = c', A = a]\mathrm{P}(A = a \mid C = c')) \\ &= \mathbb{E}[F \mid C = c] - \mathbb{E}[F \mid C = c'],\end{aligned}$$

where in the second equality we use the back-door theorem allowed by the graph $\tilde{G}$.

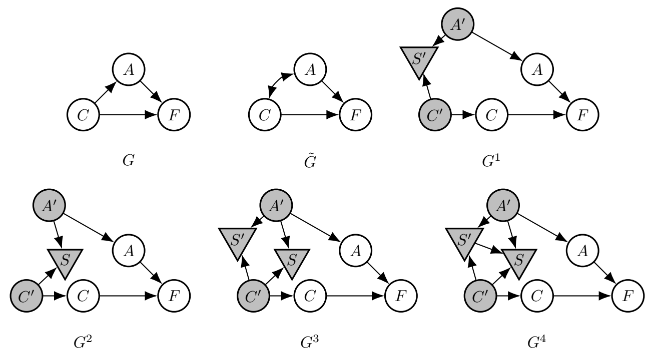

Figure 7: Causal graphs for COVID-19 data. Note that after applying the conditioning operation to selection variables and marginalizing out remaining latent variables, we reduce G^i to $\tilde{G}$ for $i = 1, 2, 3, 4$.

4 CONCLUSIONS

While marginalization plays a role of abstracting away unnecessary *unconditioned* latent details of causal models, we need another operation in case of latent selection mechanisms. We gave a formal definition of a conditioning operation on SCMs to take care of latent selection. The conditioning operation preserves a large part of the causal information, preserves important model classes and interacts well with other operations on SCMs. We generalized the interpretation of bidirected edges in directed mixed graphs to represent both latent common causes and selection on a latent event. Combined with marginalization and intervention, the conditioning operation provides a powerful tool for causal model abstraction and helps with many causal inference tasks such as prediction of interventions, identification and model selection.

Acknowledgements

This work is supported by Booking.com.

[12] Note that the difference of common cause and selection bias does not matter for the current task, which shows the power of model abstraction.

Abstracting Latent Selection in Structural Causal Models (Supplementary Material)

Leihao Chen[1] **Onno Zoeter**[2] **Joris M. Mooij**[1]

[1]Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Amsterdam, the Netherlands
[2]Booking.com, The Netherlands,

A SCM PRELIMINARIES

To be as self-contained as possible, we include the relevant SCM preliminaries. We follow the formal definitions of Bongers et al. [2021].

Definition 14 (Structural Causal Model) *A **Structural Causal Model (SCM)** is a tuple $M = (V, W, \mathcal{X}, \mathrm{P}, f)$ such that*

- *V, W are disjoint finite sets of labels for the **endogenous variables** and the **exogenous random variables**, respectively;*
- *the **state space** $\mathcal{X} = \prod_{i \in V \,\dot\cup\, W} \mathcal{X}_i$ is a product of standard measurable spaces $\mathcal{X}_i$;*
- *the **exogenous distribution** P is a probability distribution on $\mathcal{X}_W$ that factorizes as a product $\mathrm{P} = \bigotimes_{w \in W} \mathrm{P}(X_w)$ of probability distributions $\mathrm{P}(X_w)$ on $\mathcal{X}_w$;*
- *the **causal mechanism** is specified by the measurable mapping $f : \mathcal{X} \to \mathcal{X}_V$.*

Definition 15 (Hard intervention) *Given an SCM M, an intervention target $T \subseteq V$ and an intervention value $x_T \in \mathcal{X}_T$, we define the **intervened SCM***

$$M_{\mathrm{do}(X_T = x_T)} := (V, W, \mathcal{X}, \mathrm{P}, (f_{V \setminus T}, x_T)).$$

This replaces the targeted endogenous variables by specified values. In this work, we do not assume that all the endogenous variables in an SCM can be intervened on, which deviates from the standard modeling assumption. One can define other types of interventions like soft or probabilistic ones, and the results in the following also hold replacing hard interventions by other types of interventions.

Besides interventional semantics, one can also describe counterfactual semantics of an SCM by performing interventions in its twin SCM (see Definition 21).

Given an SCM M, one can define its causal graph $G(M)$ and augmented causal graph $G^a(M)$ to give an intuitive and compact graphical representation of the causal model (see Definition 23). One can read off useful causal information purely from the causal graphs without knowing the details of the underlying SCMs.

Notation 16 *In all the causal graphs, we use gray nodes to represent latent variables. We assume that latent variables are non-intervenable. Dashed nodes represent observable but non-intervenable variables, and solid nodes represent observable and intervenable variables. Exogenous variables are assumed latent.*

Definition 17 (Solution function of an SCM) *Given an SCM M, we call a measurable mapping $g^S : \mathcal{X}_{V \setminus S} \times \mathcal{X}_W \to \mathcal{X}_S$ a **solution function of** M **w.r.t.** $S \subseteq V$ if for $\mathrm{P}(X_W)$-a.a. $x_W \in \mathcal{X}_W$ and for all $x_{V \setminus S} \in \mathcal{X}_{V \setminus S}$, one has that $g^S(x_{V \setminus S}, x_W)$ satisfies the structural equations for S, i.e.,*

$$g^S(x_{V \setminus S}, x_W) = f_S(x_{V \setminus S}, g^S(x_{V \setminus S}, x_W), x_W).$$

*When $S = V$, we denote g^V by g, and call g a **solution function of** M.*

Accepted for the 1st Workshop on Causal Abstractions and Representations (CAR) at UAI 2025 (CAR 2025).

Definition 18 (Unique solvability) *An SCM M is called* ***uniquely solvable w.r.t.*** *$S \subseteq V$ if it has a solution function w.r.t. S that is* ***essentially unique*** *in the sense that if g^S and $\tilde{g}^S$ both satisfy the structural equations for S, then for $\mathrm{P}(X_W)$-a.a. $x_W \in \mathcal{X}_W$ and for all $x_{V \setminus S} \in \mathcal{X}_{V \setminus S}$, one has $g^S(x_{V \setminus S}, x_W) = \tilde{g}^S(x_{V \setminus S}, x_W)$. If M has an essentially unique solution function w.r.t. V, we call it* ***uniquely solvable****.*

Note that a subset S does not inherit unique solvability from unique solvability of any of its supersets in general [Bongers et al., 2021, Appendix B.2].

Definition 19 (Simple SCM) *An SCM M is called a* ***simple SCM*** *if it is uniquely solvable w.r.t. each subset $S \subseteq V$.*

Simple SCMs form a class of SCMs that preserves most convenient properties of acyclic SCMs but allows for weak cycles (in particular acyclic SCMs are simple). We focus on simple SCMs in this work so that we can avoid many mathematical technicalities and focus on conceptual issues. We use $\mathrm{P}_M(X_V, X_W)$ to denote the unique probability distribution of (X_V, X_W) induced by a simple SCM M.

For a simple SCM, we can plug the solution function of one component into other parts of the model so that we can get a simple SCM that "marginalizes" it out while preserving causal semantics of the remaining variables [Bongers et al., 2021].

Definition 20 (Marginalization) *Let M be a simple SCM and $L \subseteq V$. Then we call $M_{\setminus L} = (V \setminus L, W, \mathcal{X}_{V \setminus L} \times \mathcal{X}_W, \mathrm{P}, \tilde{f})$ with*

$$\tilde{f}(x_{V \setminus L}, x_W) = f_{V \setminus L}(x_{V \setminus L}, g^L(x_{V \setminus L}, x_W), x_W)$$

a ***marginalization*** *of M over $V \setminus L$.*

Definition 21 (Twin SCM) *[Bongers et al., 2021, Definition 2.17] Let $M = (V, W, \mathcal{X}, \mathrm{P}, f)$ be an SCM. The twinning operation maps M to the* ***twin structural causal model (twin SCM)***

$$M^{\text{twin}} := \left(V \cup V', W, \mathcal{X}_V \times \mathcal{X}_{V'} \times \mathcal{X}_W, \mathrm{P}, \tilde{f}\right),$$

where $V' = \{v' : v \in V\}$ is a disjoint copy of V and the causal mechanism $\tilde{f} : \mathcal{X}_V \times \mathcal{X}_{V'} \times \mathcal{X}_W \to \mathcal{X}_V \times \mathcal{X}_{V'}$ is the measurable mapping given by $\tilde{f}\,(x_V, x_{V'}, x_W) = (f(x_V, x_W), f\,(x_{V'}, x_W))$.

Definition 22 (Parent) *[Bongers et al., 2021, Definition 2.6] Let $M = (V, W, \mathcal{X}, \mathrm{P}, f)$ be an SCM. We call $k \in V \cup W$ a* ***parent*** *of $v \in V$ if and only if there does not exist a measurable mapping $\tilde{f}_v : \mathcal{X}_{V \setminus k} \times \mathcal{X}_{W \setminus k} \to \mathcal{X}_v$ such that for $\mathrm{P}(X_W)$-almost every $x_W \in \mathcal{X}_W$, for all $x_V \in \mathcal{X}_V$,*

$$x_v = f_v(x_V, x_W) \quad \Longleftrightarrow \quad x_v = \tilde{f}_v\left(x_{V \setminus k}, x_{W \setminus k}\right).$$

Definition 23 (Graph and augmented graph) *[Bongers et al., 2021, Definition 2.7] Let $M = (V, W, \mathcal{X}, \mathrm{P}, f)$ be an SCM. We define:*

(1) the augmented graph $G^a(M)$ as the directed graph with nodes $V \cup W$ and directed edges $u \to v$ if and only if $u \in V \cup W$ is a parent of $v \in V$;

(2) the graph $G(M)$ as the directed mixed graph with nodes V, directed edges $u \to v$ if and only if $u \in V$ is a parent of $v \in V$ and bidirected edges $u \leftrightarrow v$ if and only if there exists a $w \in W$ that is a parent of both $u \in V$ and $v \in V$.

Example 7 *Consider the SCM*

$$M : \begin{cases} U \sim \mathrm{Ber}(1-\xi), U_B \sim \mathrm{Ber}(1-\delta), U_E \sim \mathrm{Ber}(1-\varepsilon), \\ B_0 = U, E_0 = U, S_0 = B_0 \wedge E_0, \\ B_1 = B_0 \wedge U_B, E_1 = E_0 \wedge U_E, S_1 = B_1 \wedge E_1. \end{cases}$$

Then we have the (augmented) causal graphs of M shown in Figure 8.

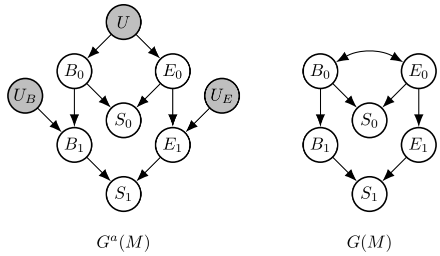

Figure 8: The (augmented) causal graphs of the SCM M in Example 7.

Definition 24 (Equivalence) *[Bongers et al., 2021, Definition 2.5] An SCM $M = (V, W, \mathcal{X}, \mathrm{P}, f)$ is **equivalent** to an SCM $\tilde{M} = (V, W, \mathcal{X}, \mathrm{P}, \tilde{f})$ if for all $v \in V$, for P-a.a. $x_W \in \mathcal{X}_W$ and for all $x_V \in \mathcal{X}_V$,*

$$x_v = f_v(x_V, x_W) \quad \iff \quad x_v = \tilde{f}_v(x_V, x_W).$$

Definition 25 (Counterfactual equivalence) *[Bongers et al., 2021, Definition 4.5] An SCM $M = (V, W, \mathcal{X}, \mathrm{P}, f)$ is **counterfactually equivalent** to an SCM $\tilde{M} = (\tilde{V}, \tilde{W}, \tilde{\mathcal{X}}, \tilde{\mathrm{P}}, \tilde{f})$ w.r.t. $O \subseteq V \cap \tilde{V}$ if for any $T_1 \subseteq O$ and $x_{T_1} \in \mathcal{X}_{T_1}$, and any $T_2 \subseteq O'$ and $x_{T_2} \in \mathcal{X}_{T_2}$,*

$$\begin{aligned}&\mathrm{P}_{M^{\mathrm{twin}}}(X_{(O \cup O') \setminus (T_1 \cup T_2)} \mid \mathrm{do}(X_{T_1} = x_{T_1}, X_{T_2} = x_{T_2})) \\ &= \mathrm{P}_{\tilde{M}^{\mathrm{twin}}}(X_{(O \cup O') \setminus (T_1 \cup T_2)} \mid \mathrm{do}(X_{T_1} = x_{T_1}, X_{T_2} = x_{T_2})).\end{aligned}$$

Definition 26 (Directed global Markov property) *Let $G = (V, E, H)$ be a DMG and $\mathrm{P}(X_V)$ a probability distribution on $\mathcal{X}_V = \prod_{v \in V} \mathcal{X}_v$ for standard measurable spaces X_v. We say that the probability distribution $\mathrm{P}(X_V)$ satisfies the **directed global Markov property relative to** G if for subsets $A, B, C \subseteq V$ the set A being d-separated from B given C implies that the random variable X_A is conditional independent of X_B given X_C.*

Theorem 27 (Directed Markov property for SCMs [Forré and Mooij, 2017]) *Let M be a uniquely solvable SCM that satisfies at least one of the following three conditions:*

(1) M is acyclic;

(2) all endogenous state spaces $\mathcal{X}_v$ are discrete and M is ancestrally uniquely solvable;

(3) M is linear and each of its causal mechanisms $\{f_v\}_{v \in V}$ has a nontrivial dependence on at least one exogenous variable, and $\mathrm{P}(X_W)$ has a density w.r.t. the Lebesgue measure on $\mathbb{R}^W$.

Then its observational distribution $\mathrm{P}(X_V)$ exists, is unique and satisfies the directed global Markov property relative to $G(M)$.

Definition 28 (Generalized directed global Markov property Forré and Mooij [2017]) *Let $G = (V, E, H)$ be a DMG and $\mathrm{P}(X_V)$ a probability distribution on $\mathcal{X}_V = \prod_{v \in V} \mathcal{X}_v$ for standard measurable spaces X_v. We say that the probability distribution $\mathrm{P}(X_V)$ satisfies the **generalized directed global Markov property relative to** G if for subsets $A, B, C \subseteq V$ the set A being σ-separated from B given C implies that the random variable X_A is conditional independent of X_B given X_C.*

Theorem 29 (Generalized directed Markov property for SCMs [Forré and Mooij, 2017, Bongers et al., 2021]) *Let M be an SCM that is simple. Then its observational distribution $\mathrm{P}(X_V)$ exists, is unique and it satisfies the general directed global Markov property relative to $G(M)$*

B MORE EXAMPLES

Example 8 (ID-algorithm) *Pearl's do-calculus is proved to be sound and complete (under some conditions) for identifying interventional distributions given a causal graph [Pearl, 1995a, Huang and Valtorta, 2006]. Using a causal graph and observational distribution as inputs, the ID-algorithm, as a sound and complete algorithm, systematically outputs a* *functional of the observational distribution to the target interventional distribution if it is identifiable and outputs FAIL if it is not [Tian and Pearl, 2002, Shpitser and Pearl, 2006, Huang and Valtorta, 2008]. Various variants of the ID-algorithm exist, each with different targets and inputs (see e.g., Kivva et al. [2023] and the references therein).*

One such variant, the s-ID-algorithm, is a sound and complete algorithm for the s-identification problem, *whose goal is to identify interventional distributions on a subpopulation (*$\mathrm{P}(X_A \mid \mathrm{do}(X_T = x_T), X_S = 1)$*) given a causal graph with selection mechanism (*G^s*) and selected observational distribution (*$\mathrm{P}(X_V \mid X_S = 1)$*) [Abouei et al., 2024a, Theorem 1, Corollary 2].*[1] *As we shall see, the conditioning operation can help simplify some parts of the original proof.*

For simplicity, we only consider the single-variable case via the conditioning operation. In the setting of Abouei et al. [2024a], there are no latent variables. Therefore, if $T = \{t\} \cap \mathrm{Anc}_{G^s}(S) = \emptyset$*, then there are no bidirected edges connecting to* t *in* $G^s_{|S}$*, which implies that* $\mathrm{P}(X_A \mid \mathrm{do}(X_T = x_T), X_S = 1)$ *is identifiable by Tian and Pearl [2002, Theorem 1]. Now, assuming that* $A \underset{G^s_{\underline{X}}}{\overset{d}{\perp}} T \mid S$*, the second rule of Pearl's do-calculus provides the identification result. Combining these two gives a sound and complete algorithm for the s-identification problem [Abouei et al., 2024a, Theorem 1].*

Besides, if $T \cap \mathrm{Anc}_{G^s}(S) = \emptyset$*, then one can also consider identifying the conditional causal effect on the subpopulation* $\mathrm{P}(X_A \mid \mathrm{do}(X_T = x_T), X_B, X_S = 1)$ *from a graph with latent variables and selected observation distribution* $\mathrm{P}(X_V \mid X_S = 1)$*, by first applying the conditioning operation on* G^s *to get* $G^s_{|S}$ *and then applying the classical ID-algorithm on* $G^s_{|S}$ *for conditional causal effect with latent variables.*[2] *This result seems to be new in the literature to our knowledge.*[3] *Similar generalizations can be made for other variants of the ID-algorithm, by first applying the conditioning operation on the graph and then applying the corresponding version of the ID-algorithm to the conditioned graph.*

However, one should note that applying the ID-algorithm to the conditioned graph alone can hardly give a complete algorithm in general, due to the abstraction nature of the conditioning operation. For example, in the case of the s-ID-algorithm, we can use the conditioning operation to handle cases where $T \cap \mathrm{Anc}_{G^s}(S) = \emptyset$*, but a complete algorithm should also be able to address cases where* $T \cap \mathrm{Anc}_{G^s}(S) \neq \emptyset$ *or* $T \underset{G^s_{\underline{X}}}{\overset{d}{\perp}} A \mid S$ *(see Abouei et al. [2024a, Theorem 1]), which can be tackled by combining with the second rule of Pearl's do-calculus.*

Example 9 (Instrumental variables) *In some situations, we cannot achieve point identification results, but we can derive informative bounds for target causal effects. A well-known example is the instrumental inequality [Pearl, 2009, Balke and Pearl, 1994, 1997, Pearl, 1995b]. More recent advances include, e.g., showing that the instrumental inequality is sharp for finite discrete variables under certain constraints on the cardinality of the variables [Van Himbeeck et al., 2019], and extending the bounds to continuous outcomes [Zhang and Bareinboim, 2021]. Not only can the original instrumental inequality for binary variables be extended to the case with certain selection bias immediately via the conditioning operation, but also the results we mentioned above.*

The inequality was derived for the SCMs with the graph $G(M)$ *shown in Figure 9. Similarly to Example 3, if we know that for an SCM* $\tilde{M}$ *with latent variables* L *and latent selection* $S \in \mathcal{S}$*, the causal graph* $G\left(\left(\tilde{M}_{\setminus L}\right)_{|\mathcal{S}}\right)$ *takes the form shown in Figure 9, then we can conclude that the same form of inequality also holds for* $\tilde{M}$ *under the subpopulation.*

If we further assume a continuous linear model $Y = \beta X + f(U)$ *in* M*, then the parameter* β *is identifiable when* $\mathrm{Cov}(X, Y) \neq 0$ *and is estimated as* $\frac{\mathrm{Cov}_M(T,Y)}{\mathrm{Cov}_M(X,Y)}$*, where selection bias is implicitly ruled out [Imbens et al., 2000]. With the conditioning operation, we can see that the parameter remains identifiable from the selected conditional distribution* $\mathrm{P}_{\tilde{M}}(T, X, Y \mid S \in \mathcal{S})$ *with the same formula* $\frac{\mathrm{Cov}_{\tilde{M}}(T,Y|S\in\mathcal{S})}{\mathrm{Cov}_{\tilde{M}}(X,Y|S\in\mathcal{S})}$ *even under certain forms of selection bias. Therefore, we have extended the identification result to include a certain form of selection bias.*

Example 10 (Causal discovery) *Many causal discovery algorithms address unobserved common causes, exclude selection bias, and output a single graph. In fact, we can interpret the output of such algorithms as* $G((M_{\setminus L})_{|\mathcal{S}})$ *where* M *is a*

[1] Note that in the usual c-ID-algorithm for conditional interventional distribution, the input is $\mathrm{P}(X_V)$ but not $\mathrm{P}(X_V \mid X_S = 1)$.

[2] If $T \cap \mathrm{Anc}_{G^s}(S) \neq \emptyset$, one can still apply the corresponding ID-algorithm to $G^s_{|S}$, but the algorithm would output an expression for $\mathrm{P}(X_A(x_T) \mid X_S = 1)$ instead of $\mathrm{P}(X_A \mid \mathrm{do}(X_T = x_T), X_S = 1)$.

[3] When we were writing this manuscript, we found that an s-ID-algorithm under latent variables was proposed in Abouei et al. [2024b]. However, they only consider identification for the unconditional interventional distribution $\mathrm{P}(X_A \mid \mathrm{do}(X_T = x_T), X_S = 1)$ not for the conditional interventional distribution $\mathrm{P}(X_A \mid \mathrm{do}(X_T = x_T), X_B, X_S = 1)$.

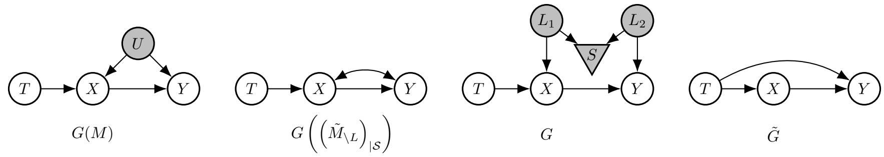

Figure 9: $G(M)$ and $G\left(\left(\tilde{M}_{\setminus L}\right)_{|\mathcal{S}}\right)$ are graphs for the instrumental variables model. G is the graph of a model with selection bias whose marginalized and conditioned graph is $G\left(\left(\tilde{M}_{\setminus L}\right)_{|\mathcal{S}}\right)$ while $\tilde{G}$ is its MAG representation.

simple (or acyclic) SCM with latent nodes L, selection mechanism $X_S \in \mathcal{S}$, and $L \cap S = \emptyset$. This can give a certain causal interpretation to the output of these algorithms under selection bias even if they exlcude selection bias in their original formulations.

For one instance, Wang and Drton [2023] explored recovering causal graphs uniquely from data generated by an acyclic linear non-Gaussian SCM with a bow-free graph (i.e., no simultaneous bidirected and directed edges between two variables) and rule out selection bias. Assume that the data are generated from an acyclic linear SCM M and there is no latent common cause or selection bias between any two variables that have a direct causal effect according to M. Then, according to the properties of marginalization and the conditioning operation, $(M_L)_{|\mathcal{S}}$ is an acyclic linear SCM with a bow-free graph (see Proposition 5 and Bongers et al. [2021, Proposition 5.11, C.5]). If the exogenous distribution of $(M_L)_{|\mathcal{S}}$ is non-Gaussian, then we can use the algorithm BANG *in Wang and Drton [2023] to recover the graph of $(M_{\setminus L})_{|\mathcal{S}}$.*

If we know from data or prior knowledge that a node t is not an ancestor of S, then we can give a causal interpretation of X_t in the discovered graph and apply causal identification results to identify $\mathrm{P}_M(X_O \mid \mathrm{do}(X_t = x_t), X_S \in \mathcal{S})$ with $O := V \setminus (L \cup S)$. For example, if the data are selected by $X_S = x_S$, we can sometimes read off whether $t \notin \mathrm{Anc}_{G(M)}(S)$ from a PAG (Partial Ancestral Graphs) or a MAG [Spirtes et al., 1995a, Richardson and Spirtes, 2002].[4]

In addition to the causal discovery algorithms mentioned above, some causal model selection methods, such as the inflation technique [Wolfe et al., 2019], can also be generalized to deal with selection bias via the conditioning operation.

Example 11 (Mediation analysis and fairness) *Mediation analysis is crucial in many fields such as epidemiology, natural science, and policy making, where understanding "path-specific" causal effects is often necessary [Pearl, 2001, 2014, 2009, Robins and Greenland, 1992].*

Traditional methods relying on linear regression, but linear SCMs have been proven problematic due to potential nonlinear interactions among variables, latent common causes, and selection bias in real-world problems [Shpitser, 2013]. With the help of potential outcomes and causal graphs of SCMs, Pearl [2014] and Shpitser [2013] study methods to perform mediation analysis when there are nonlinear functional dependencies and unobserved common causes. By extending the interpretation of bidirected edges to also represent selection bias*, we can extend these results to account for selection bias immediately, similarly to the approach in previous examples.*

For another example on how the conditioning operation is helpful, suppose that one is interested in the effect of, e.g., A (obesity) on Y (mortality) while conditioning a mediating variable on the path between them to a specific value (e.g., $S = 1$: having heart disease) [Smith, 2020]. The graph G is shown in Figure 10. Applying the graphical conditioning operation gives $G_{|S}$. This shows that we can obtain the direct causal effect given such conditioning via back-door adjustment on L.

This extension indicates that the conditioning operation can also play a significant role in fairness analysis [Nabi and Shpitser, 2018, Chiappa, 2019, Kusner et al., 2017, Zhang and Bareinboim, 2018]. Thus, we can adapt existing results to address selection bias, ensuring more robust and reliable causal inferences in the presence of such biases.

[4]Note that if $t \in \mathrm{Anc}_{G(M)}(S)$, we can still apply the identification result to the interventional distribution given $\mathrm{do}(X_t = x_t)$ in $M_{|X_S \in \mathcal{S}}$, but the causal identification results will output a formula for $\mathrm{P}_M(X_O(x_t) \mid X_S \in \mathcal{S})$ instead of $\mathrm{P}_M(X_O \mid \mathrm{do}(X_t = x_t), X_S \in \mathcal{S})$.

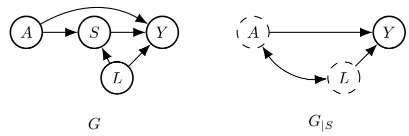

Figure 10: Graph G for mediation analysis conditioning on one mediator and its conditioned graph $G_{|S}$.

C PROOFS

C.1 PROOF OF PROPOSITION 5

If M is a simple (resp. acyclic) SCM with conditioned SCM $M_{|X_S \in \mathcal{S}}$, then the conditioned SCM $M_{|X_S \in \mathcal{S}}$ is simple (resp. acyclic). If M is also linear, then so is $M_{|X_S \in \mathcal{S}}$.

Proof We first show that the conditioning operation preserves simplicity of SCMs. First note that marginalization preserves simplicity [Bongers et al., 2021, Proposition 8.2]. Also note that both merging exogenous random variables and changing the exogenous probability distribution preserve simplicity. Hence, the conditioning operation preserves simplicity.

We give the proof of the fact that the conditioning operation preserves acyclicity of SCMs. First note that merging exogenous random variables and updating the exogenous probabilistic distribution preserve acyclicity, since exogenous random variables do not have parents. Then since marginalization preserves acyclicity [Bongers et al., 2021, Proposition 5.11], we get that the conditioning operation preserves acyclicity.

We now show that the conditioning operation preserves linearity of SCMs. Merging exogenous random variables and changing the exogenous probability distribution preserve linearity. Marginalization also preserves linearity [Bongers et al., 2021, Proposition C.5]. Combining these ingredients, we can conclude that the conditioning operation also preserves linearity of SCMs. ■

C.2 PROOF OF LEMMA 6

Assume Assumption 1. Then we have $\left(M_{\mathrm{do}(X_T=x_T)}\right)_{|X_S \in \mathcal{S}} = \left(M_{|X_S \in \mathcal{S}}\right)_{\mathrm{do}(X_T=x_T)}$ for any $T \subseteq O \setminus \mathrm{Anc}_{G^a(M)}(S)$ and $x_T \in \mathcal{X}_T$.

Proof In the proof, we set $B := \mathrm{Anc}_{G^a(M)}(S)$ and $O := V \setminus S$. We check the definition one by one. For $\left(M_{|\mathcal{S}}\right)_{\mathrm{do}(X_T=x_T)} := (\hat{V}, \hat{W}, \hat{\mathcal{X}}, \hat{\mathrm{P}}, \hat{f})$, we have:

- $\hat{V} = V \setminus S$;
- $\hat{W} = (W \setminus B) \,\dot{\cup}\, \{\star_W\}$ with $\star_W = B \cap W$;
- $\hat{\mathcal{X}} = \mathcal{X}_{S^c \setminus (B \cap W)} \times \mathcal{X}_{\star_W}$;
- $\hat{\mathrm{P}} = \hat{\mathrm{P}}\left(X_{W \setminus B}\right) \otimes \hat{\mathrm{P}}(X_{\star_W}) = \mathrm{P}\left(X_{W \setminus B}\right) \otimes \mathrm{P}_M(X_{W \cap B} \mid X_S \in \mathcal{S})$;
- $\hat{f}\left(x_{\hat{V}}, x_{\hat{W}}\right) = \left(f_{O \setminus T}\left(x_O, g^S\left(x_O, x_{W \setminus B}, x_{\star_W}\right), x_{W \setminus B}, x_{\star_W}\right), x_T\right)$.

We write $\tilde{B} := \mathrm{Anc}_{G^a(M_{\mathrm{do}(X_T=x_T)})}(S)$. Note that since $T \cap B = \emptyset$, it follows that $\tilde{B} = B$. Since $T \cap B = T \cap \tilde{B} = \emptyset$, we have $\mathrm{P}_M(X_B) = \mathrm{P}_{M_{\mathrm{do}(X_T=x_T)}}(X_{\tilde{B}})$. Hence, we can conclude that

$$\mathrm{P}_M(X_{W \cap B} \mid X_S \in \mathcal{S}) = \mathrm{P}_{M_{\mathrm{do}(X_T=x_T)}}(X_{W \cap \tilde{B}} \mid X_S \in \mathcal{S}).$$

Combining all the above ingredients, we have for $\left(M_{\mathrm{do}(X_T=x_T)}\right)_{|\mathcal{S}} := (\hat{\hat{V}}, \hat{\hat{W}}, \hat{\hat{\mathcal{X}}}, \hat{\hat{\mathrm{P}}}, \hat{\hat{f}})$:

- $\hat{\hat{V}} = V \setminus S$;
- $\hat{\hat{W}} = \left(W \setminus \tilde{B}\right) \dot{\cup} \{\star_W\} = (W \setminus B) \,\dot{\cup}\, \{\star_W\}$ with $\star_W = \tilde{B} \cap W = B \cap W$;

- $\hat{\hat{\mathcal{X}}} = \mathcal{X}_{S^c \setminus (\tilde{B} \cap W)} \times \mathcal{X}_{\star_W} = \mathcal{X}_{S^c \setminus (B \cap W)} \times \mathcal{X}_{\star_W}$;
- $\hat{\hat{\mathrm{P}}} = \mathrm{P}(X_{W \setminus \tilde{B}}) \otimes \hat{\hat{\mathrm{P}}}(X_{\star_W}) = \mathrm{P}(X_{W \setminus \tilde{B}}) \otimes \mathrm{P}_{M_{\mathrm{do}(X_T = x_T)}}(X_{W \cap \tilde{B}} \mid X_S \in \mathcal{S}) = \mathrm{P}(X_{W \setminus B}) \otimes \mathrm{P}_M(X_{W \cap B} \mid X_S \in \mathcal{S})$;
- For the causal mechanism, we have

$$\begin{aligned}\hat{\hat{f}}\left(x_{\hat{\hat{V}}}, x_{\hat{\hat{W}}}\right) &= \tilde{f}_O\left(x_O, \tilde{g}^S\left(x_O, x_{W \setminus \tilde{B}}, x_{\star_W}\right), x_{W \setminus \tilde{B}}, x_{\star_W}\right) \\ &= \left(f_{O \setminus T}\left(x_O, \tilde{g}^S\left(x_O, x_{W \setminus B}, x_{\star_W}\right), x_{W \setminus B}, x_{\star_W}\right), x_T\right),\end{aligned}$$

where $\tilde{f}$ is the causal mechanism of $M_{\mathrm{do}(X_T = x_T)}$ and $\tilde{g}^S$ is the (essentially unique) solution function of $M_{\mathrm{do}(X_T = x_T)}$ w.r.t. S. Note that $g^S = \tilde{g}^S$ as $T \cap B = \emptyset$. Overall, it is then easy to see that $\left(M_{\mathrm{do}(X_T = x_T)}\right)_{|\mathcal{S}} = \left(M_{|\mathcal{S}}\right)_{\mathrm{do}(X_T = x_T)}$. ■

C.3 PROOF OF THEOREM 8

Assume Assumption 1. Then we have

(1) $\mathrm{P}_{M_{|X_S \in \mathcal{S}}}(X_O) = \mathrm{P}_M(X_O \mid X_S \in \mathcal{S})$;

(2) for any $T \subseteq V \setminus \mathrm{Anc}_{G^a(M)}(S)$ and $x_T \in \mathcal{X}_T$,

$$\mathrm{P}_{M_{|X_S \in \mathcal{S}}}\left(X_{O \setminus T} \mid \mathrm{do}(X_T = x_T)\right) = \mathrm{P}_M\left(X_{O \setminus T} \mid \mathrm{do}(X_T = x_T), X_S \in \mathcal{S}\right);$$

(3) for any $T_1 \subseteq V \setminus \mathrm{Anc}_{G^a(M)}(S)$ and $x_{T_1} \in \mathcal{X}_{T_1}$, and any $T_2 \subseteq (V \setminus \mathrm{Anc}_{G^a(M)}(S))'$ and $x_{T_2} \in \mathcal{X}_{T_2}$,

$$\begin{aligned}&\mathrm{P}_{(M_{|X_S \in \mathcal{S}})^{\mathrm{twin}}}(X_{(O \cup O') \setminus (T_1 \cup T_2)} \mid \mathrm{do}(X_{T_1} = x_{T_1}, X_{T_2} = x_{T_2})) \\ &= \mathrm{P}_{M^{\mathrm{twin}}}(X_{(O \cup O') \setminus (T_1 \cup T_2)} \mid \mathrm{do}(X_{T_1} = x_{T_1}, X_{T_2} = x_{T_2}), X_S \in \mathcal{S}).\end{aligned}$$

Proof We first prove (1) of Theorem 8. Let $g : \mathcal{X}_W \to \mathcal{X}_V$ be the essentially unique solution function of M. Write $O := V \setminus S$ and $B := \mathrm{Anc}_{G^a(M)}(S)$ and $\star_W = B \cap W$. First note that the function $\hat{g} : \mathcal{X}_{W \setminus B} \times \mathcal{X}_{\star_W} \to \mathcal{X}_{V \setminus S}$ with

$$\hat{g}(x_{W \setminus B}, x_{\star_W}) := g_O(x_{W \setminus B}, x_{\star_W})$$

is the essentially unique solution function of $M_{|\mathcal{S}}$. In fact, for $\mathrm{P}(X_W)$-a.a. $x_W \in \mathcal{X}_W$ and all $x_V \in \mathcal{X}_V$

$$\begin{cases} x_S = g_S(x_W) \\ x_O = g_O(x_W) \end{cases} \Leftrightarrow \begin{cases} x_S = f_S(x_V, x_W) \\ x_O = f_O(x_V, x_W) \end{cases} \Leftrightarrow \begin{cases} x_S = g^S(x_O, x_W) \\ x_O = f_O(x_V, x_W) \end{cases} \Leftrightarrow \begin{cases} x_S = g^S(x_O, x_W) \\ x_O = f_O\left(x_O, g^S(x_O, x_W), x_W\right). \end{cases}$$

Let $\hat{\mathrm{P}}$ denote the exogenous probability distribution of $M_{|\mathcal{S}}$, that is, $\hat{\mathrm{P}} := \mathrm{P}(X_{W \setminus B}) \otimes \hat{\mathrm{P}}(X_{\star_W})$, where $\hat{\mathrm{P}}(X_{\star_W}) = \mathrm{P}_M(X_{W \cap B} \mid X_S \in \mathcal{S})$. Recall that we have

$$\mathrm{P}_M(X_V) = g_*\left(\mathrm{P}(X_W)\right)(X_V) \text{ i.e. } \mathrm{P}_M(X_V \in A) = \mathrm{P}(X_W \in g^{-1}(A))$$

for any measurable subset $A \subseteq \mathcal{X}_V$. Then we have for any measurable subset $A \subseteq \mathcal{X}_V$

$$\begin{aligned}\mathrm{P}_{M_{|\mathcal{S}}}(X_O \in A) &= \hat{\mathrm{P}}\left(X_{\hat{W}} \in \hat{g}^{-1}(A)\right) \\ &= \hat{\mathrm{P}}\left(X_{\hat{W}} \in g_O^{-1}(A)\right) \\ &= \mathrm{P}_M(X_W \in g_O^{-1}(A) \mid X_S \in \mathcal{S}) \\ &= \mathrm{P}_M\left(X_O \in A \mid X_S \in \mathcal{S}\right).\end{aligned}$$

We then show (2) of Theorem 8. Lemma 6 gives that $\left(M_{|\mathcal{S}}\right)_{\mathrm{do}(X_T = x_T)} = \left(M_{\mathrm{do}(X_T = x_T)}\right)_{|\mathcal{S}}$ for any $T \subseteq V \setminus B$ and $x_T \in \mathcal{X}_T$. We then have for $T \subseteq V \setminus B$ and $x_T \in \mathcal{X}_T$

$$\begin{aligned}\mathrm{P}_{M_{|\mathcal{S}}}\left(X_{O \setminus T} \mid \mathrm{do}(X_T = x_T)\right) &= \mathrm{P}_{(M_{|\mathcal{S}})_{\mathrm{do}(X_T = x_T)}}\left(X_{O \setminus T}\right) \\ &= \mathrm{P}_{(M_{\mathrm{do}(X_T = x_T)})_{|\mathcal{S}}}\left(X_{O \setminus T}\right) \\ &= \mathrm{P}_{M_{\mathrm{do}(X_T = x_T)}}(X_{O \setminus T} \mid \tilde{g}_S(X_W) \in \mathcal{S}) \\ &= \mathrm{P}_M(X_{O \setminus T} \mid \mathrm{do}(X_T = x_T), g_S(X_W) \in \mathcal{S}) \\ &= \mathrm{P}_M(X_{O \setminus T} \mid \mathrm{do}(X_T = x_T), X_S \in \mathcal{S}),\end{aligned}$$

where $\tilde{g}$ is the essentially unique solution function of $M_{\mathrm{do}(X_T=x_T)}$, which satisfies $\tilde{g}_S(x_W) = g_S(x_W)$ for $\mathrm{P}(X_W)$-a.a. $x_W \in \mathcal{X}_W$.

We finally show (3) of Theorem 8. Firs note that $\mathrm{Anc}_{G^a(M^{\mathrm{twin}})}(S) = \mathrm{Anc}_{G^a(M^{\mathrm{twin}})}(S')$ and

$$\mathrm{P}_{M^{\mathrm{twin}}}(X_{\mathrm{Anc}_{G^a(M^{\mathrm{twin}})}(S)\cap W} \mid X_S \in \mathcal{S}) = \mathrm{P}_{M^{\mathrm{twin}}}(X_{\mathrm{Anc}_{G^a(M^{\mathrm{twin}})}(S')\cap W} \mid X_{S'} \in \mathcal{S}').$$

By the definition of conditioning operation and twinning operation, we have $\left((M^{\mathrm{twin}})_{|\mathcal{S}}\right)_{|\mathcal{S}'} = \left((M^{\mathrm{twin}})_{|\mathcal{S}}\right)_{\setminus S'} = (M_{|\mathcal{S}})^{\mathrm{twin}}$, where $\mathcal{S}' \subseteq \mathcal{X}_{S'}$ is such that $\mathcal{S}' = \mathcal{S}$ and S' is the copy of S. We have from (2) of Theorem 8 that for any $T_1 \subseteq V \setminus \mathrm{Anc}_{G^a(M)}(S)$ and $x_{T_1} \in \mathcal{X}_{T_1}$, and for any $T_2 \subseteq (V \setminus \mathrm{Anc}_{G^a(M)}(S))'$ and $x_{T_2} \in \mathcal{X}_{T_2}$,

$$\begin{aligned}
&\mathrm{P}_{(M_{|\mathcal{S}})^{\mathrm{twin}}}\left(X_{(O\cup O')\setminus(T_1\cup T_2)} \mid \mathrm{do}(X_{T_1} = x_{T_1}, X_{T_2} = x_{T_2})\right) \\
&= \mathrm{P}_{\left((M^{\mathrm{twin}})_{|\mathcal{S}}\right)_{\setminus S'}}\left(X_{(O\cup O')\setminus(T_1\cup T_2)} \mid \mathrm{do}(X_{T_1} = x_{T_1}, X_{T_2} = x_{T_2})\right) \\
&= \mathrm{P}_{(M^{\mathrm{twin}})_{|\mathcal{S}}}\left(X_{(O\cup O')\setminus(T_1\cup T_2)} \mid \mathrm{do}(X_{T_1} = x_{T_1}, X_{T_2} = x_{T_2})\right) \\
&= \mathrm{P}_{M^{\mathrm{twin}}}(X_{(O\cup O')\setminus(T_1\cup T_2)} \mid \mathrm{do}(X_{T_1} = x_{T_1}, X_{T_2} = x_{T_2}), X_S \in \mathcal{S}).
\end{aligned}$$

■

C.4 PROOF OF COROLLARY 9

Assume Assumption 1 with $S = S_1 \cup S_2$ and $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2$ with $\mathcal{S}_1 \subseteq \mathcal{X}_{S_1}$ and $\mathcal{S}_2 \subseteq \mathcal{X}_{S_2}$ both measurable. Then $(M_{|\mathcal{S}_1})_{|\mathcal{S}_2}$, $(M_{|\mathcal{S}_2})_{|\mathcal{S}_1}$, and $M_{|\mathcal{S}_1\times\mathcal{S}_2}$ are counterfactually equivalent w.r.t. $V \setminus \mathrm{Anc}_{G^a(M)}(S_1 \cup S_2)$.

Proof

Write $O := V \setminus (S_1 \cup S_2)$. From (3) of Theorem 8, it is easy to see that for any $T_1 \subseteq V \setminus \mathrm{Anc}_{G^a(M)}(S_1 \cup S_2)$ and $x_{T_1} \in \mathcal{X}_{T_1}$, and any $T_2 \subseteq (V \setminus \mathrm{Anc}_{G^a(M)}(S_1 \cup S_2))'$ and $x_{T_2} \in \mathcal{X}_{T_2}$,

$$\begin{aligned}
&\mathrm{P}_{\left((M_{|\mathcal{S}_1})_{|\mathcal{S}_2}\right)^{\mathrm{twin}}}\left(X_{(O\cup O')\setminus(T_1\cup T_2)} \mid \mathrm{do}(X_{T_1} = x_{T_1}, X_{T_2} = x_{T_2})\right) \\
&= \mathrm{P}_{(M_{|\mathcal{S}_1})^{\mathrm{twin}}}\left(X_{(O\cup O')\setminus(T_1\cup T_2)} \mid \mathrm{do}(X_{T_1} = x_{T_1}, X_{T_2} = x_{T_2}), X_{S_2} \in \mathcal{S}_2\right) \\
&= \mathrm{P}_{M^{\mathrm{twin}}}(X_{(O\cup O')\setminus(T_1\cup T_2)} \mid \mathrm{do}(X_{T_1} = x_{T_1}, X_{T_2} = x_{T_2}), X_{S_1} \in \mathcal{S}_1, X_{S_2} \in \mathcal{S}_2) \\
&= \mathrm{P}_{(M_{|\mathcal{S}_2})^{\mathrm{twin}}}\left(X_{(O\cup O')\setminus(T_1\cup T_2)} \mid \mathrm{do}(X_{T_1} = x_{T_1}, X_{T_2} = x_{T_2}), X_{S_1} \in \mathcal{S}_1\right) \\
&= \mathrm{P}_{\left((M_{|\mathcal{S}_2})_{|\mathcal{S}_1}\right)^{\mathrm{twin}}}\left(X_{(O\cup O')\setminus(T_1\cup T_2)} \mid \mathrm{do}(X_{T_1} = x_{T_1}, X_{T_2} = x_{T_2})\right).
\end{aligned}$$

Also note that

$$\begin{aligned}
&\mathrm{P}_{(M_{|\mathcal{S}_1\times\mathcal{S}_2})^{\mathrm{twin}}}\left(X_{(O\cup O')\setminus(T_1\cup T_2)} \mid \mathrm{do}(X_{T_1} = x_{T_1}, X_{T_2} = x_{T_2})\right) \\
&= \mathrm{P}_{M^{\mathrm{twin}}}(X_{(O\cup O')\setminus(T_1\cup T_2)} \mid \mathrm{do}(X_{T_1} = x_{T_1}, X_{T_2} = x_{T_2}), X_{S_1} \in \mathcal{S}_1, X_{S_2} \in \mathcal{S}_2).
\end{aligned}$$

■

C.5 PROOF OF PROPOSITION 10

Assume Assumption 1 and let $L \subseteq V \setminus S$. Then we have $(M_{\setminus L})_{|\mathcal{S}}$ and $(M_{|\mathcal{S}})_{\setminus L}$ are counterfactually equivalent.

Proof Since $(M_{\setminus L})_{|\mathcal{S}}$ and $(M_{|\mathcal{S}})_{\setminus L}$ have the same exogenous probability distribution, which equals to $\mathrm{P}_M(X_W \mid X_S \in \mathcal{S})$. Also note that $(M_{\setminus L})_{\setminus S} = (M_{\setminus S})_{\setminus L}$, since $S \cap L = \emptyset$ [Bongers et al., 2021, Proposition 5.4]. Combining these implies that $(M_{\setminus L})_{|\mathcal{S}}$ and $(M_{|\mathcal{S}})_{\setminus L}$ are counterfactually equivalent. ■

C.6 PROOF OF PROPOSITION 11

Let M be a simple SCM with conditioned SCM $M_{|X_S\in\mathcal{S}}$. Then $G(M_{|X_S\in\mathcal{S}})$ is a subgraph of $G(M)_{|S}$.

Proof Recall that we call a subset A of V ancestral if $\mathrm{Anc}_{G(M)}(A) = A$ and call an SCM M ancestrally uniquely solvable if for every ancestral subset A of V the SCM M is essentially uniquely solvable w.r.t. A. Since simple SCM is ancestrally uniquely solvable, we have that $G(M_{\setminus S})$ is a subgraph of $G(M)_{\setminus S}$ by Bongers et al. [2021, Proposition 5.11]. Note that in $G^a(M_{|X_S\in\mathcal{S}})$ all the exogenous ancestors of S are merged, so in $G(M_{|X_S\in\mathcal{S}})$ there might be bidirected edges between pairs of nodes that are ancestors or siblings of S (the presence of a bidirected edge between a pair depends on the functional relationships in the SCM $M_{|X_S\in\mathcal{S}}$) but no other new bidirected edge in $G(M_{|X_S\in\mathcal{S}})$ compared to $G(M)$. From the definition of the conditioned DMG, we can conclude that $G(M_{|X_S\in\mathcal{S}})$ is a subgraph of $G(M)_{|S}$ ■

C.7 PROOF OF COROLLARY 13

If M is simple, then $\mathrm{P}_{M_{|X_S\in\mathcal{S}}}(X_O)$ satisfies the generalized directed global Markov property relative to $G(M)_{|S}$. If M is acyclic, then $\mathrm{P}_{M_{|X_S\in\mathcal{S}}}(X_O)$ satisfies the directed global Markov property relative to $G(M)_{|S}$.

Proof If M is a simple SCM, Proposition 5 implies that $M_{|X_S\in\mathcal{S}}$ is a simple SCM. Since $M_{|X_S\in\mathcal{S}}$ is a simple SCM, the observational distribution $\mathrm{P}_{M_{|X_S\in\mathcal{S}}}(X_O)$ satisfies the generalized directed global Markov property relative to $G(M_{|X_S\in\mathcal{S}})$ by Theorem 29. From Proposition 11, we know that $G(M_{|X_S\in\mathcal{S}})$ is a subgraph of $G(M)_{|S}$. Hence, $\mathrm{P}_{M_{|X_S\in\mathcal{S}}}(X_O)$ also satisfies the generalized directed global Markov property relative to $G(M)_{|S}$.

If M is an acyclic SCM, then by Proposition 5 $M_{|X_S\in\mathcal{S}}$ is an acyclic SCM. Therefore the observational distribution $\mathrm{P}_{M_{|X_S\in\mathcal{S}}}(X_O)$ satisfies the generalized directed global Markov property relative to $G(M_{|X_S\in\mathcal{S}})$ by Theorem 27. Again Proposition 11 implies that $G(M_{|X_S\in\mathcal{S}})$ is a subgraph of $G(M)_{|S}$. We can therefore conclude that $\mathrm{P}_{M_{|X_S\in\mathcal{S}}}(X_O)$ satisfies the directed global Markov property relative to $G(M)_{|S}$. ■

Remark 30 *In other words, conditioning operation on SCM will not generate new directed causal path from the graph of the original SCM. If one starts from a structurally minimal SCM M (we call an SCM structurally minimal if for every causal mechanism f_v the number of the variables that f_v depends on cannot be reduced, see Bongers et al. [2021, Definition 2.10]), then the conditioned SCM $M_{|(X_S\in\mathcal{S})}$ may not be structurally minimal (similar things happen to marginalization). Reflecting it in the level of causal graphs means that if one starts with a causal graph in which directed edges are present if and only if there are some causal effects, then she may end up with a graph in which directed edges only indicate possible causal effects and no directed edges mean no direct effects surely. A causal effect exists in some subpopulation must exist in the whole population but not the other way around.*

D MORE EXAMPLES

D.1 EXAMPLE OF DEFINITION 4

Here we show an example of the purely graphical conditioning operation, i.e., Definition 4. Assume we are given a graph G as shown in Figure 11. Then conditioning on the node V_5 gives the graph $G_{|V_5}$ shown in Figure 11.

D.2 MORE DETAILS ABOUT EXAMPLE 6

We explain why one has two different answers to the same question in Example 6 based on G and $\tilde{G}$, respectively. For an SCM with graph G, one has:

$$\mathrm{TCE}(Y; c' \to c) := \mathbb{E}[F \mid \mathrm{do}(C = c)] - \mathbb{E}[F \mid \mathrm{do}(C = c')] = \mathbb{E}[F \mid C = c] - \mathbb{E}[F \mid C = c'].$$

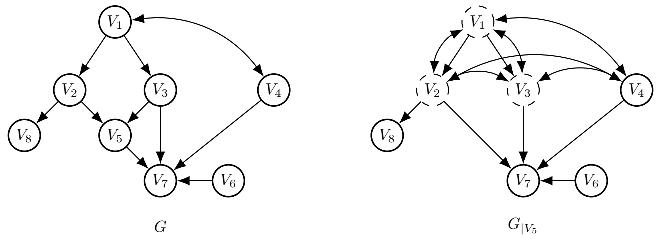

Figure 11: DMG G and its Conditioned DMG $G_{|V_5}$.

On the other hand, for an SCM with graph $\tilde{G}$, one has:

$$\begin{aligned}
\mathrm{TCE}(Y; c' \to c) &:= \mathbb{E}[F \mid \mathrm{do}(C=c)] - \mathbb{E}[F \mid \mathrm{do}(C=c')] \\
&= \sum_a \left(\mathbb{E}[F \mid C=c, A=a] - \mathbb{E}[F \mid C=c', A=a]\right) \mathrm{P}(A=a) \\
&\neq \sum_a \left(\mathbb{E}[F \mid C=c, A=a]\mathrm{P}(A=a \mid C=c) - \mathbb{E}[F \mid C=c', A=a]\mathrm{P}(A=a \mid C=c')\right) \quad \text{(in general)} \\
&= \mathbb{E}[F \mid C=c] - \mathbb{E}[F \mid C=c'],
\end{aligned}$$

where in the second equality we use the Back-door theorem allowed by the structure of the graph $\tilde{G}$.

E SOME IMPORTANT CAVEATS ON MODELING INTERPRETATION

In Section 2, we presented the conditioning operation as a purely mathematical operation and derived some mathematical properties of it. In this subsection, we shall make some remarks on how to interpret the conditioned SCMs appropriately to avoid confusion in modeling applications.

The subtleties are about intervening on ancestors of selection nodes. In this case, conditioning and interventions are not commutative as we showed before. Therefore, one should be careful about the order of these two operations. On the one hand, if we first intervene and second condition on descendants of intervened variables, then the selected subpopulation will also change according to the intervention. On the other hand, first conditioning and second intervening on ancestors of selection nodes has a "counterfactual flavor". Suppose that an SCM M with three variables T ("treatment"), Y ("outcome") and S ("selection") has causal graph $T \to Y \to S$. Intuitively, "first-conditioning-second-intervening" indicates that we first observe the results of the treatment and select units with specific values (say $S = s$) and fix this subpopulation. After that, we go back to then perform an intervention (say $\mathrm{do}(T = t)$) on this *fixed selected subpopulation* instead of the total population. Mathematically, we have

$$\begin{aligned}
\mathrm{P}_{\left((M_{|S=s})_{\mathrm{do}(T=t)}\right)}(Y) &= \mathrm{P}_{M_{|S=s}}(Y \mid \mathrm{do}(T=t)) \\
&= \mathrm{P}_M(Y_t \mid S=s) \\
&= \mathrm{P}_{M^{\mathrm{twin}}}(Y' \mid \mathrm{do}(T'=t), S=s) \\
&\neq \mathrm{P}(Y \mid \mathrm{do}(T=t), S=s), \quad \text{(in general)} \\
&= \mathrm{P}_{\left((M_{\mathrm{do}(T=t)})_{|S=s}\right)}(Y)
\end{aligned}$$

where we used the language of potential outcomes. In Pearl's terminology, this mixes different rungs: a rung-two query in the conditioned SCM is equivalent to a rung-three query in the original SCM.

As far as we know, there are two possible ways to use the conditioning operation for modeling without introducing confusion:

- before (or after) performing the conditioning operation, marginalizing out all the ancestors of the selection nodes, so that one can no longer intervene on the ancestors of the selection nodes;
- specifying in the conditioned SCM and its graph which variables are ancestors of the selection nodes in the original SCM, and marking them as non-intervenable (e.g., making them dashed).[5]

[5]This means that we obtain a graph with mixed interpretation in the sense that some part of the graph is causal and some part is non-causal (purely probabilistic).

Remark 31 *When the selection variables do not have any intervenable ancestors (e.g., all the ancestors of the selection nodes are latent), one can safely apply the conditioning operation without any extra steps.*

As the above discussions showed, there are some relations between conditioned SCMs and counterfactual reasoning. It is an interesting future work to explore the relation further. There are many important notions defined via nested counterfactual quantity such as various notions of fairness Kusner et al. [2017], Zhang and Bareinboim [2018]. Although the conditioned SCM is only able to express unnested counterfactual quantity, we can rewrite nested counterfactual quantity as unnested one by the Counterfactual Unnesting Theorem [Correa et al., 2021, Theorem 1].

References

Amir Mohammad Abouei, Ehsan Mokhtarian, and Negar Kiyavash. s-id: Causal effect identification in a sub-population. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(18):20302–20310, Mar. 2024a. doi: 10.1609/aaai.v38i18.30011. URL https://ojs.aaai.org/index.php/AAAI/article/view/30011.

Amir Mohammad Abouei, Ehsan Mokhtarian, Negar Kiyavash, and Matthias Grossglauser. Causal effect identification in a sub-population with latent variables. *arXiv.org preprint, arXiv:2405.14547 [cs.LG]*, 2024b.

Alexander Balke and Judea Pearl. Counterfactual probabilities: computational methods, bounds and applications. In *Proceedings of the 10th International Conference on Uncertainty in Artificial Intelligence*, UAI'94, page 46–54, San Francisco, CA, USA, 1994. Morgan Kaufmann Publishers Inc. ISBN 1558603328.

Alexander Balke and Judea Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1171–1176, 1997. doi: 10.1080/01621459.1997.10474074. URL https://doi.org/10.1080/01621459.1997.10474074.

Elias Bareinboim and Judea Pearl. Controlling selection bias in causal inference. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, volume 22 of *Proceedings of Machine Learning Research*, pages 100–108, La Palma, Canary Islands, 21–23 Apr 2012. PMLR. URL https://proceedings.mlr.press/v22/bareinboim12.html.

Elias Bareinboim and Jin Tian. Recovering causal effects from selection bias. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, page 2410–2416, 2015.

Sander Beckers and Joseph Y. Halpern. Abstracting causal models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):2678–2685, Jul. 2019. doi: 10.1609/aaai.v33i01.33012678. URL https://ojs.aaai.org/index.php/AAAI/article/view/4117.

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

Stephan Bongers, Tineke Blom, and Joris M. Mooij. Causal modeling of dynamical systems. *arXiv.org preprint*, arXiv:1803.08784v4 [cs.AI], 2022. URL https://arxiv.org/abs/1803.08784v4.

Silvia Chiappa. Path-specific counterfactual fairness. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):7801–7808, Jul. 2019. doi: 10.1609/aaai.v33i01.33017801. URL https://ojs.aaai.org/index.php/AAAI/article/view/4777.

Gregory F. Cooper. Causal discovery from data in the presence of selection bias. In Doug Fisher and Hans-Joachim Lenz, editors, *Pre-proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics*, volume R0 of *Proceedings of Machine Learning Research*, pages 140–150. PMLR, 04–07 Jan 1995. URL https://proceedings.mlr.press/r0/cooper95a.html. Reissued by PMLR on 01 May 2022.

Juan Correa, Sanghack Lee, and Elias Bareinboim. Nested counterfactual identification from arbitrary surrogate experiments. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 6856–6867. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/36bedb6eb7152f39b16328448942822b-Paper.pdf.

Patrick Forré and Joris M. Mooij. Markov properties for graphical models with cycles and latent variables. *arXiv.org preprint*, arXiv:1710.08775 [math.ST], 2017. URL `https://arxiv.org/abs/1710.08775`.

Patrick Forré and Joris M Mooij. Constraint-based causal discovery for non-linear structural causal models with cycles and latent confounders. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence*, pages 269–278, 2018.

Atticus Geiger, Chris Potts, and Thomas F. Icard. Causal abstraction for faithful model interpretation. *arXiv.org preprint*, arxiv:2301.04709 [cs.AI], 2023.

Miguel A. Hernán, Sonia Hernández-Díaz, and James M. Robins. A structural approach to selection bias. *Epidemiology*, 15 (5):615–625, 2004.

Robins JM Hernán MA. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020.

Yimin Huang and Marco Valtorta. Pearl's calculus of intervention is complete. In *Proceedings of the 22ed Conference on Uncertainty in Artificial Intelligence*, page 217–224, 2006.

Yimin Huang and Marco G. Valtorta. On the completeness of an identifiability algorithm for semi-markovian models. *Annals of Mathematics and Artificial Intelligence*, 54:363–408, 2008. URL `https://api.semanticscholar.org/CorpusID:52818662`.

Antti Hyttinen, Frederick Eberhardt, and Matti Järvisalo. Constraint-based causal discovery: Conflict resolution with answer set programming. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence*, page 340–349, 2014.

Guido Imbens, Joshua Angrist, and Kathryn Graddy. The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *Review of Economic Studies*, 67, July:499–527, 2000.

Yaroslav Kivva, Jalal Etesami, and Negar Kiyavash. On identifiability of conditional causal effects. In *Proceedings of the 39th Conference on Uncertainty in Artificial Intelligence*, 2023.

Matt Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. Counterfactual fairness. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 4069–4079, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Joris M. Mooij and Tom Claassen. Constraint-based causal discovery using partial ancestral graphs in the presence of cycles. In Jonas Peters and David Sontag, editors, *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI-20)*, volume 124, pages 1159–1168. PMLR, 8 2020. URL `http://proceedings.mlr.press/v124/m-mooij20a/m-mooij20a-supp.pdf`.

Razieh Nabi and Ilya Shpitser. Fair inference on outcomes. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32 (1), Apr. 2018. doi: 10.1609/aaai.v32i1.11553. URL `https://ojs.aaai.org/index.php/AAAI/article/view/11553`.

Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995a. ISSN 00063444. URL `http://www.jstor.org/stable/2337329`.

Judea Pearl. On the testability of causal models with latent and instrumental variables. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, page 435–443, 1995b.

Judea Pearl. Direct and indirect effects. In *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, pages 411–420. PMLR, 2001.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition, 2009. doi: 10.1017/CBO9780511803161.

Judea Pearl. Interpretation and identification of causal mediation. *Psychological methods*, 19(4):459—481, December 2014. ISSN 1082-989X. doi: 10.1037/a0036434. URL `https://doi.org/10.1037/a0036434`.

Jonas Peters, Stefan Bauer, and Niklas Pfister. Causal models for dynamical systems. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, pages 671–690. 2022.

Hans Reichenbach. *The Direction of Time*. Dover Publications, Mineola, N.Y., 1956.

Thomas Richardson. Markov properties for acyclic directed mixed graphs. *Scandinavian Journal of Statistics*, 30(1): 145–157, 2003. doi: https://doi.org/10.1111/1467-9469.00323. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9469.00323`.

Thomas S. Richardson and James M. Robins. Single world intervention graphs: a primer. `https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=07bbcb458109d2663acc0d098e8913892389a2a7`, 2013. URL `https://api.semanticscholar.org/CorpusID:13292974`.

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

James M. Robins and Sander Greenland. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3: 143–155, 1992. URL `https://api.semanticscholar.org/CorpusID:10757981`.

Paul K. Rubenstein, Sebastian Weichwald, Stephan Bongers, Joris M. Mooij, Dominik Janzing, Moritz Grosse-Wentrup, and Bernhard Schölkopf. Causal consistency of structural equation models. In *Proceedings of the 33rd Annual Conference on Uncertainty in Artificial Intelligence (UAI-17)*, 2017.

Ilya Shpitser. Counterfactual graphical models for longitudinal mediation analysis with unobserved confounding. *Cognitive science*, 37(6):1011–1035, 2013.

Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-markovian causal models. In *Proceedings of the 21st National Conference on Artificial Intelligence and the 18th Innovative Applications of Artificial Intelligence Conference, AAAI-06/IAAI-06*, Proceedings of the National Conference on Artificial Intelligence, pages 1219–1226, November 2006. ISBN 1577352815. 21st National Conference on Artificial Intelligence and the 18th Innovative Applications of Artificial Intelligence Conference, AAAI-06/IAAI-06 ; Conference date: 16-07-2006 Through 20-07-2006.

Louisa H Smith. Selection mechanisms and their consequences: understanding and addressing selection bias. *Current Epidemiology Reports*, 7:179–189, 2020.

Peter Spirtes, Christopher Meek, and Thomas Richardson. Causal inference in the presence of latent variables and selection bias. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, page 499–506, 1995a.

Peter Spirtes, Christopher Meek, and Thomas Richardson. Causal inference in the presence of latent variables and selection bias. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, UAI'95, page 499–506, 1995b.

Jin Tian and Judea Pearl. A general identification condition for causal effects. In *Eighteenth National Conference on Artificial Intelligence*, page 567–573, USA, 2002. American Association for Artificial Intelligence. ISBN 0262511290.

Thomas Van Himbeeck, Jonatan Bohr Brask, Stefano Pironio, Ravishankar Ramanathan, Ana Belén Sainz, and Elie Wolfe. Quantum violations in the Instrumental scenario and their relations to the Bell scenario. *Quantum*, 3:186, September 2019. ISSN 2521-327X. doi: 10.22331/q-2019-09-16-186. URL `https://doi.org/10.22331/q-2019-09-16-186`.

Philip Versteeg, Joris Mooij, and Cheng Zhang. Local constraint-based causal discovery under selection bias. In Bernhard Schölkopf, Caroline Uhler, and Kun Zhang, editors, *Proceedings of the First Conference on Causal Learning and Reasoning*, volume 177 of *Proceedings of Machine Learning Research*, pages 840–860. PMLR, 11–13 Apr 2022. URL `https://proceedings.mlr.press/v177/versteeg22a.html`.

Julius Von Kügelgen, Luigi Gresele, and Bernhard Schölkopf. Simpson's paradox in COVID-19 case fatality rates: a mediation analysis of age-related causal effects. *IEEE Transactions on Artificial Intelligence*, 2(1):18–27, 2021.

Y. Samuel Wang and Mathias Drton. Causal discovery with unobserved confounding and non-gaussian data. *Journal of Machine Learning Research*, 24(271):1–61, 2023. URL `http://jmlr.org/papers/v24/21-1329.html`.

Elie Wolfe, Robert W. Spekkens, and Tobias Fritz. The inflation technique for causal inference with latent variables. *Journal of Causal Inference*, 7(2):20170020, 2019. doi: doi:10.1515/jci-2017-0020. URL `https://doi.org/10.1515/jci-2017-0020`.

Jiji Zhang. Causal reasoning with ancestral graphs. *Journal of Machine Learning Research*, 9:1437–1474, 2008.

Junzhe Zhang and Elias Bareinboim. Fairness in decision-making—the causal explanation formula. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Junzhe Zhang and Elias Bareinboim. Bounding causal effects on continuous outcome. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(13):12207–12215, May 2021. doi: 10.1609/aaai.v35i13.17449. URL https://ojs.aaai.org/index.php/AAAI/article/view/17449.