TECHNICAL REPORT
R-34-L
July 2018

Non-Parametric Path Analysis in Structural Causal Models

Junzhe Zhang and **Elias Bareinboim**
Purdue University, USA
{zhang745, eb}@purdue.edu

Abstract

One of the fundamental tasks in causal inference is to decompose the observed association between a decision X and an outcome Y into its most basic structural mechanisms. In this paper, we introduce counterfactual measures for effects along with a specific mechanism, represented as a path from X to Y in an arbitrary structural causal model. We derive a novel non-parametric decomposition formula that expresses the covariance of X and Y as a sum over unblocked paths from X to Y contained in an arbitrary causal model. This formula allows a fine-grained path analysis without requiring a commitment to any particular parametric form, and can be seen as a generalization of Wright's decomposition method in linear systems (1923,1932) and Pearl's non-parametric mediation formula (2001).

1 INTRODUCTION

Analyzing the relative strength of different pathways between a decision X and an outcome Y is a topic that has interested both scientists and practitioners across disciplines for many decades. Specifically, path analysis allows scientists to explain how Nature's "black-box" works, and practically, it enables decision analysts to predict how an environment will change under a variety of policies and interventional conditions [Wright, 1923; Baron and Kenny, 1986; Bollen, 1989; Pearl, 2001].

More recently, understanding using causal inference tools how a black-box decision-making system operates has been a target of growing interest in the Artificial Intelligence community, most prominently in the context of Explainability, Transparency, and Fairness [Lu Zhang, 2017; Kusner *et al.*, 2017; Zafar *et al.*, 2017; Kilbertus *et al.*, 2017; Zhang and Bareinboim, 2018a]. For example, consider the *standard fairness model* described in Fig. 1(a) that is concerned with the relation between a hiring decision (Y) and an applicant's religious beliefs (X), which are *mediated by* the location (W), and *confounded by* the education background (Z) of the applicant. [1] Directed edges represent functional relations between variables. The relationship between X and Y is materialized through four different pathways in the system – the *direct* path $l_1 : X \to Y$, the *indirect* path $l_2 : X \to W \to Y$, and the *spurious* paths $l_3 : X \leftarrow Z \to Y$ and $l_4 : X \leftarrow Z \to W \to Y$.

Assuming, for simplicity's sake, that the functional relationships are linear and U_{V_i} is an independent "error" associated with each variable V_i (called the linear-standard model), Fig. 1(a) shows the structural coefficients corresponding to each edge – i.e., the value of the variable Y is decided by the structural function $Y \leftarrow \alpha_{YX}X+\alpha_{YZ}Z+\alpha_{YW}W+U_Y$. The celebrated result known as Wright's method of path coefficients [Wright, 1923, 1934], also known as Wright's rule, allows one to express the covariance of X and Y, denoted by $\mathrm{Cov}(X, Y)$, as the sum of the products of the structural coefficients along the paths from X to Y in the underlying causal model. [2] In particular, $\mathrm{Cov}(X, Y)$ is equal to:

$$\underbrace{\alpha_{YX}}_{X\to Y} + \underbrace{\alpha_{WX}\alpha_{YW}}_{X\to W\to Y} + \underbrace{\alpha_{XZ}\alpha_{YZ}}_{X\leftarrow Z\to Y} + \underbrace{\alpha_{XZ}\alpha_{WZ}\alpha_{YW}}_{X\leftarrow Z\to W\to Y}. \tag{1}$$

Using the observational covariance matrix, the decomposition above allows one to answer some compelling questions about the relationship between X and Y in the underlying model. For instance, the product $\alpha_{WX}\alpha_{YW}$ explains how much the indirect discrimination through the location (the path l_2) accounts for the observed disparities in the religion composition among hired employees.

The path analysis method gained momentum in the so-

[1] This specific setting has been called *standard fairness model* given its generality to representing a variety of decision-making scenarios [Zhang and Bareinboim, 2018a].

[2] For a survey on linear methods, see [Pearl, 2000, Ch. 5].

cial sciences during 1960's, becoming extremely popular in the form of the *mediation formula* in which the total effect of X on Y is decomposed into direct and indirect components [Baron and Kenny, 1986; Bollen, 1989; Duncan, 1975; Fox, 1980]. [3] The bulk of this literature, however, required a commitment to a particular parametric form, thus falling short of providing a general method for analyzing natural and social phenomena with nonlinearities and interactions [MacKinnon, 2008].

It took a few decades until this problem could be tackled in higher generality. In particular, the advent of non-parametric structural causal models (SCMs) allowed this leap, and a more fine-grained path-analysis with a much broader scope, including models with nonlinearities and arbitrarily complex interactions [Pearl, 2000, Ch. 7]. In particular, Pearl introduced the *causal mediation formula* for arbitrary non-parametric models, which decomposes the total effect $TE_{x_0,x_1}(Y) = \mathrm{E}[Y_{x_1}] - \mathrm{E}[Y_{x_0}]$, the difference between the causal effect of the intervention $do(x_1)$ and $do(x_0)$ [4], into what is now known as the natural direct (*NDE*) and indirect (*NIE*) effects [Pearl, 2001] (see also [Imai *et al.*, 2010, 2011; VanderWeele, 2015]). In the case of the specific linear-standard causal model,

$$TE_{0,1}(Y) = \underbrace{\alpha_{YX}}_{NDE} + \underbrace{\alpha_{WX}\alpha_{YW}}_{NIE}$$

for $x_0 = 0$ and $x_1 = 1$ levels. Remarkably, when compared with Eq. 1, *NDE* and *NIE* capture the effects along with the direct and indirect paths, but omits the spurious (non-causal) paths between X and Y (in this case, l_3, l_4). The mediation formula was recently generalized to account for these spurious paths (more akin to Wright's rules), which appears under the rubric of the *causal explanation formula* [Zhang and Bareinboim, 2018a]. This formula decomposes the total variation $TV_{x_0,x_1}(Y) = \mathrm{E}[Y|x_1] - \mathrm{E}[Y|x_0]$ (difference in conditional distributions) into counterfactual measures of the direct (*Ctf-DE*), indirect (*Ctf-IE*), and spurious (*Ctf-SE*) effects. In the linear-standard model, for $x_0 = 0, x_1 = 1$,

$$TV_{0,1}(Y) = \underbrace{\alpha_{YX}}_{Ctf\text{-}DE} + \underbrace{\alpha_{WX}\alpha_{YW}}_{Ctf\text{-}IE} + \underbrace{\alpha_{XZ}\alpha_{YZ} + \alpha_{XZ}\alpha_{WZ}\alpha_{YW}}_{Ctf\text{-}SE}$$

Despite the generality of such results, there are still outstanding challenges when performing path analysis in non-parametric models, i.e.: (1) Estimands are defined relative to specific values assigned to the treatment x_1 and its baseline x_0, which may be difficult to select in some non-linear settings; (2) Mediators and confounders are collapsed and considered *en bloc*, leading to a coarse decomposition of the relationship between X and Y [Pearl, 2001; Vansteelandt and VanderWeele, 2012; Tchetgen and Shpitser, 2012; VanderWeele *et al.*, 2014; Daniel *et al.*, 2015; Zhang and Bareinboim, 2018a]; (3) Path-specific estimands are well-defined [Pearl, 2001; Avin *et al.*, 2005], but not in a way that they sum up to either the total effect (TE) or variation (TV), precluding the comparison of their relative strengths.

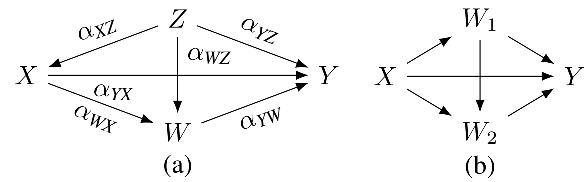


Figure 1: Causal diagrams for (a) the standard fairness model where X stands for the protected attribute, Y for the outcome, W the mediators, and Z the confounders; (b) the two-mediators setting where causal paths from X to Y are mediated by W_1, W_2.

This paper aims to circumvent these problems. In particular, we decompose the covariance of a treatment X and an outcome Y over effects along different mechanisms between X and Y. We define a set of novel counterfactual estimands for measuring the relative strength of a specific mechanism represented as a path from X to Y in an arbitrary causal model. These estimands lead to a non-parametric decomposition formula, which expresses the covariance $\mathrm{Cov}(X, Y)$ as a sum of the unblocked paths from X to Y in the causal graph. This formula allows a more fine-grained analysis of the total observed variations of Y due to X (both through causal and spurious mechanisms) when compared to the state-of-art methods. More specifically, our contributions are: (1) counterfactual covariance measures for a specific pathway from X to Y (causal and spurious) in an arbitrary causal model (Defs. 8, 11-12); (2) non-parametric decomposition formulae of the covariance $\mathrm{Cov}(X, Y)$ over paths from X to Y in the causal model (Thm. 5); (3) identification conditions for estimating the proposed path-specific decomposition from the passively-collected data in the standard model (Thms. 6-7).

2 PRELIMINARIES

In this section, we introduce notations used throughout the paper. We will use capital letters to denote variables (e.g., X), and small letters for their values (x). The abbreviation $P(x)$ represents the probabilities $P(X = x)$. For arbitrary sets A and B, let $A - B$ denote their differ-

[3] Just to give an idea of this popularity, Baron and Kenny's original paper counts more than 70,000 citations.

[4] By convention [Pearl, 2000], the post-interventional distribution is represented interchangeably by $P(y_x)$ and $P(y|do(x))$. General notation is discussed in the next section.

ence, and let $|A|$ be the dimension of set A. $V_{[i,j]}$ stands for a set $\{V_i, \ldots, V_j\}$ ($\emptyset$ if $i > j$). We use graphical family abbreviations: $An(X)_G$, $De(X)_G$, $Non\text{-}De(X)_G$, $Pa(X)_G$, $Ch(X)_G$, which stand for the set of ancestors, descendants, non-descendants, parents and children of X in G. We omit the subscript G when obvious.

The basic semantical framework of our analysis rests on *structural causal models* (SCM) [Pearl, 2000, Ch. 7; Bareinboim and Pearl, 2016]. A SCM M consists of a set of endogenous variables V (often observed) and exogenous variables U (often unobserved). The values of each $V_i \in V$ are determined by a structural function f_i taking as argument a combination of the other endogenous and exogenous variables (i.e., $V_i \leftarrow f_i(PA_i, U_i), PA_i \subseteq V, U_i \subseteq U$)). Values of U are drawn from a distribution $P(u)$. A SCM M is called *Markovian* when the exogenous are mutually independent and each $U_i \in U$ is associated with only one endogenous $V_i \in V$. If U_i is associated with two or more endogenous variables, M is called *semi-Markovian*.

Each recursive SCM M has an associated causal diagram in the form of a directed acyclic graph (DAG) G, where nodes represent endogenous variables and directed edges represent functional relations (e.g., Figs. 1-2). By convention, the exogenous U are not explicitly shown in the graph; a dashed-bidirected arrow between V_i and V_j indicates the presence of an unobserved confounder (UC) U_k affecting both V_i and V_j (e.g., the path $V_i \leftarrow U_k \rightarrow V_j$).

A path from X to Y is a sequence of edges which does not include a particular node more than once. It may go either along or against the direction of the edges. Paths of the form $X \rightarrow \cdots \rightarrow Y$ are *causal* (from X to Y). We use d-separation and blocking interchangeably, following the convention in [Pearl, 2000]. Any unblocked path that is not causal is called *spurious*. The direct link $X \rightarrow Y$ is the *direct* path and all the other causal paths from X to Y are called *indirect*. The set of unblocked paths from X to Y given a set Z in a causal diagram G is denoted by $\Pi(X, Y|Z)_G$; causal, indirect, and spurious paths are denoted by $\Pi^c(X, Y|Z)_G$, $\Pi^i(X, Y|Z)_G$, and $\Pi^s(X, Y|Z)_G$ (G will be omitted when obvious). For a causal path g including nodes V_1, V_2, we denote $g(V_1, V_2)$ a subpath of g from V_1 to V_2.[5]

An intervention on a set of endogenous variables X and exogenous variables U_i, denoted by $do(x^*, u_i^*)$, is an operation where values of X, U_i are set to x^*, u_i^*, respectively, without regard to how they were ordinarily determined (X through f_X and U_i through $P(U_i)$). Formally, we can rewrite the definition of potential response [Pearl, 2000, Ch. 7.1] to account for operation on U_i, namely:

Definition 1 (Potential Response)**.** Let M be a SCM, X, Y sets of arbitrary variables in V, and U_i a set of arbitrary variables in U. Let $U_{-i} = U - U_i$. The potential response of Y to the intervention $do(x^*, u_i^*)$ in the situation $U = u$, denoted by $Y_{x^*, u_i^*}(u)$, is the solution for Y with $U_{-i} = u_{-i}, U_i = u_i^*$ in the modified submodel M_{x^*} where functions f_X are replaced by constant functions $X = x^*$, i.e., $Y_{x^*, u_i^*}(u) \triangleq Y_{M_{x^*}}(u_i^*, u_{-i})$.[6]

$Y_{x^*, u_i^*}(u)$ can be read as the counterfactual sentence "the value that Y would have obtained in situation $U_{-i} = u_{-i}$, had the treatment X been x^* and the situation U_i been u_i^*." Averaging u over the distribution $P(u)$, we obtain a counterfactual random variable Y_{x^*, u_i^*}. If the values of x^*, u_i^* follow random variables X^*, U_i^*, we denote the resulting counterfactual Y_{X^*, U_i^*}.

3 A COARSE COVARIANCE DECOMPOSITION

In this section, we introduce counterfactual measures that will allow us to non-parametrically decompose the covariance $\mathrm{Cov}(X, Y)$ in terms of direct, indirect and spurious pathways from X to Y. Given space constraints, all proofs are included in Appendix 1.

If there exists no spurious path from X to Y, then treatment X is independent of the counterfactual Y_{x^*}, i.e., $(X \perp\!\!\!\perp Y_{x^*})$ [Pearl, 2000, Ch. 11.3.2]. The *spurious covariance* can then be defined as the correlation between the factual variable X and counterfactual Y_{x^*}.

Definition 2 (Spurious Covariance)**.** The spurious covariance between treatment $X = x^*$ and outcome Y is:

$$\mathrm{Cov}^s_{x^*}(X, Y) = \mathrm{Cov}(X, Y_{x^*}). \tag{2}$$

Property 1. $|\Pi^s(X, Y)| = 0 \Rightarrow \mathrm{Cov}^s_{x^*}(X, Y) = 0.$

The *causal covariance* can naturally be defined as the difference between the total and spurious covariance.

Definition 3 (Causal Covariance)**.** The causal covariance of the treatment $X = x^*$ and the outcome Y is:

$$\mathrm{Cov}^c_{x^*}(X, Y) = \mathrm{Cov}(X, Y - Y_{x^*}). \tag{3}$$

Prop. 2 establishes the correspondence between the causal paths and the causal covariance – if there is no causal path from X to Y in the underlying model, the causal covariance equates to zero.

Property 2. $|\Pi^c(X, Y)| = 0 \Rightarrow \mathrm{Cov}^c_{x^*}(X, Y) = 0.$

We consider more detailed measures corresponding to the different causal pathways, and first, the direct path:

[5] Mediators (relative to X and Y) are a set of variables $W \subseteq De(X) \cap Non\text{-}De(Y)$ such that $|\Pi^i(X, Y|W)| = 0$.

[6] An alternative way to see the replacement operation relative to U_i is to envision a system where U_i is observed.

Definition 4 (Direct Covariance)**.** Given a semi-Markovian model M, let the set W be the mediators between X and Y. The pure ($\mathrm{Cov}^{dp}_{x^*}(X,Y)$) and total ($\mathrm{Cov}^{dt}_{x^*}(X,Y)$) direct covariance of the treatment $X = x^*$ on the outcome Y are defined respectively as

$$\mathrm{Cov}^{dp}_{x^*}(X,Y) = \mathrm{Cov}(X, Y - Y_{x^*,W}), \quad (4)$$

$$\mathrm{Cov}^{dt}_{x^*}(X,Y) = \mathrm{Cov}(X, Y_{W_{x^*}} - Y_{x^*}). \quad (5)$$

By the composition axiom [Pearl, 2000, Ch. 7.3], Eqs. 4 and 5 can be explicitly written as follows [7]:

$$\mathrm{Cov}(X, Y - Y_{x^*,W}) = \mathrm{Cov}(X, Y_{X,W} - Y_{x^*,W}),$$
$$\mathrm{Cov}(X, Y_{W_{x^*}} - Y_{x^*}) = \mathrm{Cov}(X, Y_{X,W_{x^*}} - Y_{x^*,W_{x^*}}).$$

The counterfactual pure direct covariance (Eq. 4) is shown graphically in Fig. 2, where (a) corresponds to the Y-side, and (b) to the $Y_{x^*,W}$-side. Note that from the mediator W perspective, X remains at the level that it would naturally have attained, while the "direct" input from X to Y varies from its natural level (Fig. 2a) to $do(x^*)$ (b). The change of the outcome Y thus measures the effect of the direct path. A similar analysis also applies to the total direct covariance (Eq. 5).

Property 3. $\mathrm{Cov}^{dp}_{x^*}(X,Y) = \mathrm{Cov}^{dt}_{x^*}(X,Y) = 0$ *if* X *is not a parent of* Y *(i.e.,* $X \notin Pa(Y)$*).*

We can turn around the definitions of direct covariance and provide operational estimands for indirect paths.

Definition 5 (Indirect Covariance)**.** Given a semi-Markovian model M, let the set W be the mediators between X and Y. The pure ($\mathrm{Cov}^{ip}_{x^*}(X,Y)$) and total ($\mathrm{Cov}^{it}_{x^*}(X,Y)$) indirect covariance of the treatment $X = x^*$ on the outcome Y are defined respectively as:

$$\mathrm{Cov}^{ip}_{x^*}(X,Y) = \mathrm{Cov}(X, Y - Y_{W_{x^*}}), \quad (6)$$

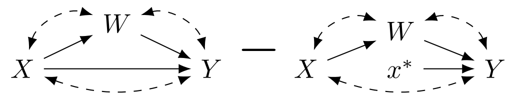
$$\mathrm{Cov}^{it}_{x^*}(X,Y) = \mathrm{Cov}(X, Y_{x^*,W} - Y_{x^*}). \quad (7)$$

Eqs. 6 and 7 correspond to the indirect paths, since they capture the covariance of X and Y, but only via paths mediated by W. The first argument of Y is the same in both halves of the contrast, but this value can either be x^* (Eq. 7) or at the level that X would naturally attain without intervention (Eq. 6).

Property 4. $|\Pi^i(X,Y)| = 0 \Rightarrow \mathrm{Cov}^{ip}_{x^*}(X,Y) = \mathrm{Cov}^{it}_{x^*}(X,Y) = 0.$

Putting these definitions together, we can prove a general non-parametric decomposition of $\mathrm{Cov}(X,Y)$:

[7]Consider Eq. 4 as an example. For any $U = u$, $Y_{X(u),W(u)}(u) = Y_{x^*,w}(u)$ if $X(u) = x^*, W(u) = w$. By the composition axiom, $X(u) = x^*, W(u) = w$ implies $Y(u) = Y_{x^*,w}(u)$, which in turn gives $Y_{X(u),W(u)}(u) = Y(u)$. Averaging u over $P(u)$, we obtain $Y_{X,W} = Y$.



(a) Y (b) $Y_{x^*,W}$

Figure 2: The graphical representation of measuring the pure direct covariance $\mathrm{Cov}^{dp}_{x^*}(X,Y)$.

Theorem 1. $\mathrm{Cov}(X,Y)$*,* $\mathrm{Cov}^{s}_{x^*}(X,Y)$ *and* $\mathrm{Cov}^{c}_{x^*}(X,Y)$ *obey the following non-parametric relationship:*

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}^{c}_{x^*}(X,Y) + \mathrm{Cov}^{s}_{x^*}(X,Y), \quad (8)$$

where $\mathrm{Cov}^{c}_{x^*}(X,Y) = \mathrm{Cov}^{dp}_{x^*}(X,Y) + \mathrm{Cov}^{it}_{x^*}(X,Y) = \mathrm{Cov}^{dt}_{x^*}(X,Y) + \mathrm{Cov}^{ip}_{x^*}(X,Y)$*.*

In other words, the covariance between X and Y can be partitioned into its corresponding direct, indirect, and spurious components. In particular, Thm. 1 coincides with Eq. 1 in the linear-standard model.

Corollary 1. *In the linear-standard model, for any* x^**,* $\mathrm{Cov}^{s}_{x^*}(X,Y)$*,* $\mathrm{Cov}^{dp}_{x^*}(X,Y)$*,* $\mathrm{Cov}^{dt}_{x^*}(X,Y)$*,* $\mathrm{Cov}^{ip}_{x^*}(X,Y)$ *and* $\mathrm{Cov}^{it}_{x^*}(X,Y)$ *are equal to:*

$$\mathrm{Cov}^{s}_{x^*}(X,Y) = \alpha_{XZ}\alpha_{YZ} + \alpha_{XZ}\alpha_{WZ}\alpha_{YW},$$
$$\mathrm{Cov}^{dp}_{x^*}(X,Y) = \mathrm{Cov}^{dt}_{x^*}(X,Y) = \alpha_{YX},$$
$$\mathrm{Cov}^{ip}_{x^*}(X,Y) = \mathrm{Cov}^{it}_{x^*}(X,Y) = \alpha_{WX}\alpha_{YW}.$$

Corol. 1 says that the proposed decomposition (Thm. 1) does not depend on the value of $do(x^*)$ in the linear model of Fig. 1(a), which is not achievable in previous value-specific decompositions [Pearl, 2001; Zhang and Bareinboim, 2018a].[8]

4 DECOMPOSING CAUSAL RELATIONS

We now focus on the challenge of decomposing the causal covariance into more elementary components. We use the two-mediators setting (Fig. 1(b)) as example, where X and Y are connected through four causal paths: through both W_1, W_2 ($g_1 : X \to W_1 \to W_2 \to Y$), only through W_1 ($g_2 : X \to W_1 \to Y$), only through W_2 ($g_3 : X \to W_2 \to Y$), and directly ($g_4 : X \to Y$). Our goal is to decompose the $\mathrm{Cov}^{c}_{x^*}(X,Y)$ over the paths $g_{[1,4]}$. Our analysis applies to semi-Markovian models, without loss of generality, and the Markovian example (Fig. 1(b)) is used for simplicity of the exposition.

[8]For the nonlinear models, the decomposing terms (e.g., $\mathrm{Cov}^{s}_{x^*}(X,Y)$) are still sensitive to the target level $do(x^*)$. To circumvent the challenges of picking a specific decision value, one could assign a randomized treatment $do(x^* \sim P(X))$, where $P(X)$ is the distribution over the treatment X induced by the underlying causal model.

For a node $S_i \in Pa(Y)$ and a set of causal paths π, the edge $S_i \to Y$ defines a funnel operator $\lhd_{S_i \to Y}$, which maps from π to the set of paths $\lhd_{S_i \to Y}(\pi)$ obtained from π by replacing all paths of the form $X \to \cdots \to S_i \to Y$ with $X \to \cdots \to S_i$, and removing all the other paths. As an example, for $\pi = \{g_1, g_2, g_3\}$, $\lhd_{W_2 \to Y}(\pi) = \{g_1(X, W_2), g_3(X, W_2)\}$, where $g_1(X, W_2)$ is the subpath $X \to W_1 \to W_2$ and $g_3(X, W_2)$ is the subpath $X \to W_2$. We next formalize the notion of path-specific interventions, which isolates the influence of the intervention $do(x^*)$ passing through a subset π of causal paths from X, denoted by $do(\pi[x^*])$ (a similar notion has been introduced by [Pearl, 2001], and then [Avin *et al.*, 2005; Shpitser and Tchetgen, 2016]).

Definition 6 (Path-Specific Potential Response)**.** For a SCM M and an arbitrary variable $Y \in V$, let π be a set of causal paths. Let X be the source variables of paths in π. Further, let $X_{\pi \to Y} = \{X_i : \forall X_i \in X, X_i \to Y \in \pi\}$ and $S = (Pa(Y)_G \cap V) - X_{\pi \to Y}$. The π-specific potential response of Y to the intervention $do(\pi[x^*])$ in the situation $U = u$, denoted by $Y_{\pi[x^*]}(u)$, is defined as:

$$Y_{\pi[x^*]}(u) = \begin{cases} Y_{x^*_{\pi \to Y}, S_{\lhd_{S \to Y}(\pi)[x^*]}(u)} & \text{if } \pi \neq \emptyset \\ Y(u) & \text{otherwise} \end{cases}$$

where $S_{\lhd_{S \to Y}(\pi)[x^*]}(u)$ is a set of π-specific potential response $\{S_{i_{\lhd_{S_i \to Y}(\pi)[x^*]}}(u) : S_i \in S\}$.[9]

Despite the non-trivial notation, the π-specific counterfactual $Y_{\pi[x^*]}$ is simply assigning the treatment $do(x^*)$ exclusively to the causal paths in π, while allowing all the other causal paths to behave naturally. This contrasts with the counterfactual Y_{x^*}, which can be seen as assigning the treatment $do(x^*)$ to all causal paths from X to Y. For instance, repeatedly applying Def. 6 to $g_1 : X \to W_1 \to W_2 \to Y$ (see Appendix 2.1), we obtain the g_1-specific potential response $Y_{g_1[x^*]}$ as

$$Y_{g_1[x^*]} = Y_{X, W_1, W_{2_{X, W_{1_{x^*}}}}} = Y_{W_{2_{W_{1_{x^*}}}}}.$$

The intervention $do(g_1[x^*])$ can be visualized more immediately through its graphical representation (Fig. 3(b)) – the treatment $do(x^*)$ is assigned throughout g_1 while all the other paths are kept at the level that it would have attained "naturally" following X. The difference of the outcome Y (induced by $do(g_1[x^*])$) and the unintervened Y (Fig. 3(a)) measures the relative strength of g_1 itself, which leads to the following definition.

Definition 7 (Pure Path-Specific Causal Covariance)**.** For a semi-Markovian model M and an arbitrary causal path g from X, the pure g-specific causal covariance of the treatment $X = x^*$ on the outcome Y is defined as:

$$\mathrm{Cov}^c_{g[x^*]}(X, Y) = \mathrm{Cov}(X, Y - Y_{g[x^*]}). \quad (9)$$

[9]For a single causal path g, let $Y_{g[x^*]}(u) = Y_{\{g\}[x^*]}(u)$. Averaging u over $P(u)$, we obtain a random variable $Y_{\pi[x^*]}$.

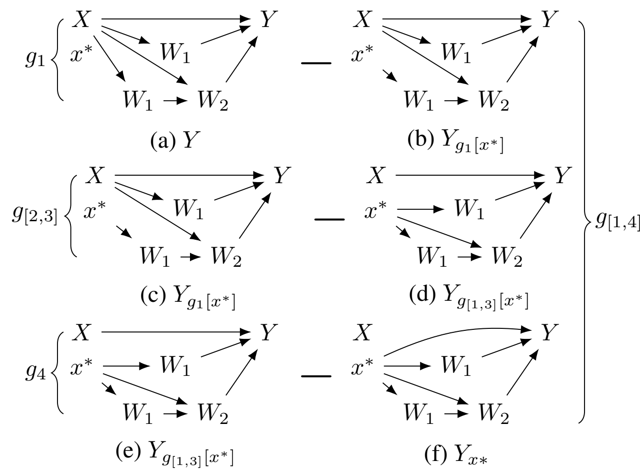


Figure 3: Graphical representations of the causal covariance specific to g_1 (a-b), $g_{[2,3]}$ (c-d) and g_4 (e-f).

In the previous example, more explicitly, the pure g_1-specific causal covariance is equal to (Fig. 3(a-b)):

$$\mathrm{Cov}^c_{g_1[x^*]}(X, Y) = \mathrm{Cov}\Big(X, Y - Y_{W_{2_{W_{1_{x^*}}}}}\Big). \quad (10)$$

For $U = u$, the counterfactual $Y_{\emptyset[x^*]}(u)$ stands for the values of Y when all causal paths are under the natural regime. Eq. 9 can then be rewritten as:

$$\mathrm{Cov}^c_{g[x^*]}(X, Y) = \mathrm{Cov}(X, Y_{\emptyset[x^*]} - Y_{g[x^*]}).$$

The pure path-specific causal covariance for g can be seen as a function of the difference between two path-specific potential response $Y_{\pi_0[x^*]}$ and $Y_{\pi_1[x^*]}$ such that $g \notin \pi_0$ and $\pi_1 = \pi_0 \cup \{g\}$ (i.e., the difference between π_1 and π_0 is g). The difference $Y_{\pi_1[x^*]} - Y_{\pi_0[x^*]}$, therefore, measures precisely the effects of $do(x^*)$ along the target causal path g. Def. 7 can be generalized to account for the path-specific covariance in terms of path-differences.

Definition 8 (Path-Specific Causal Covariance)**.** For a semi-Markovian model M and an arbitrary causal path g from X, let π be a function mapping g to a set of causal paths $\pi(g)$ from X such that $g \notin \pi(g)$. The g-specific causal covariance of the treatment $X = x^*$ on the outcome Y is defined as:

$$\mathrm{Cov}^c_{g[x^*]}(X, Y)_\pi = \mathrm{Cov}(X, Y_{\pi(g)[x^*]} - Y_{\pi(g) \cup \{g\}[x^*]}).$$

The following property establishes the correspondence between a causal path and its path-specific estimand.

Property 5. $g \notin \Pi^c(X, Y) \Rightarrow \mathrm{Cov}^c_{g[x^*]}(X, Y)_\pi = 0.$

Prop. 5 follows immediately as a corollary of Lem. 1, which implies that the counterfactuals $Y_{\pi(g)[x^*]}$ and $Y_{\pi(g) \cup \{g\}[x^*]}$ define the same variable over U if g is not a causal path from X to Y.

Lemma 1. $g \notin \Pi^c(X,Y) \Rightarrow Y_{\pi(g)[x^*]}(u) = Y_{\pi(g)\cup\{g\}[x^*]}(u)$.

Considering again the model in Fig. 1(b), let $g_{[i,j]} = \{g_k\}_{i\leq k\leq j}$ ($\emptyset$ if $i > j$). Recall that $g_4 = \{X \rightarrow Y\}$, and note that the g_4-specific causal covariance can be computed using $\pi(g_4) = g_{[1,3]}$, yielding:

$$
\begin{aligned}
&\mathrm{Cov}^c_{g_4[x^*]}(X,Y)_\pi = \mathrm{Cov}(X, Y_{g_{[1,3]}[x^*]} - Y_{g_{[1,4]}[x^*]}) \\
&= \mathrm{Cov}(X, Y_{W_{1_{x^*}}, W_{2_{x^*}}} - Y_{x^*}), \qquad (11)
\end{aligned}
$$

which coincides with the direct effect (Eq. 5 with $W = \{W_1, W_2\}$). Fig. 3(e-f) shows a graphical representation of this procedure.

The path-specific quantity given in Def. 8 has another desirable property, namely, the causal covariance $\mathrm{Cov}^c_x(X,Y)$ can be decomposed as a summation over causal paths from X to Y. To witness, first let an order over $\Pi^c(X,Y)$ be $\mathcal{L}^c : g_1 < \cdots < g_n$. For a path $g_i \in \Pi^c(X,Y)$, the order $\mathcal{L}^c$ defines a function $\mathcal{L}^c_\pi$ which maps from g_i to a set of paths $\mathcal{L}^c_\pi(g_i)$ that precede g_i in $\mathcal{L}^c$, i.e., $\mathcal{L}^c_\pi(g_i) = g_{[1,i-1]}$. We derive in the sequel a path-specific decomposition formula for the causal covariance relative to an order $\mathcal{L}^c$.

Theorem 2. *For a semi-Markovian model M, let $\mathcal{L}^c$ be an order over $\Pi^c(X,Y)$. For any x^*, the following non-parametric relationship hold:*

$$
\mathrm{Cov}^c_x(X,Y) = \sum_{g\in\Pi^c(X,Y)} \mathrm{Cov}^c_{g[x^*]}(X,Y)_{\mathcal{L}^c_\pi}.
$$

Thm. 2 can be demonstrated in the model of Fig. 1(a). Let an order $\mathcal{L}^c$ over $g_{[1,4]}$ be $g_i < g_j$ if $i < j$. First note that the path-specific causal covariance of g_2 ($\mathrm{Cov}^c_{g_2[x^*]}(X,Y)_{\mathcal{L}^c_\pi}$) and g_3 ($\mathrm{Cov}^c_{g_3[x^*]}(X,Y)_{\mathcal{L}^c_\pi}$) are equal to, respectively,

$$
\mathrm{Cov}\Big(X, Y_{W_{2_{W_{1_{x^*}}}}} - Y_{W_{2_{W_{1_{x^*}}}}, W_{1_{x^*}}}\Big) \qquad (12)
$$

$$
\mathrm{Cov}\Big(X, Y_{W_{2_{W_{1_{x^*}}}}, W_{1_{x^*}}} - Y_{W_{1_{x^*}}, W_{2_{x^*}}}\Big) \qquad (13)
$$

The causal covariance $\mathrm{Cov}^c_x(X,Y)$ can then be decomposed as the sum of Eqs. 10-13, respectively, g_1, g_4, g_2, g_3. Fig. 3 describes this decomposition procedure: we measures the difference of the outcome Y as the intervention $do(x^*)$ propagates through paths g_1, g_2, g_3, g_4. The sum of these differences thus equate to the total influence of the intervention $do(x^*)$ to the outcome Y, i.e., the causal covariance $\mathrm{Cov}^c_{x^*}(X,Y)$.

5 DECOMPOSING SPURIOUS RELATIONS

We introduce in the sequel a new strategy to decompose the spurious covariance (Def. 2), which will play a central role in the analysis of the spurious relations relative to the pair X, Y. The spurious covariance measures the correlation between the observational X and the counterfactual Y_{x^*} (Def. 2). We will employ in our analysis the twin network [Balke and Pearl, 1994; Pearl, 2000, Sec. 7.1.4], which is a graphical method to analyzing the relation between observational and counterfactual variables.

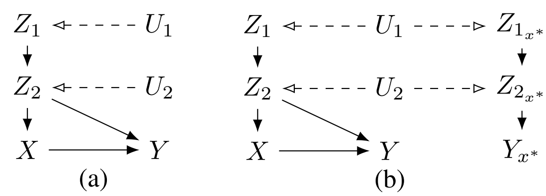


Figure 4: Causal diagrams for (a) the one-confounder setting where X and Y are confounded by the variable Z_2, of which Z_1 is a parent node; (b) the twin network for the model of (a) under $do(x^*)$.

Consider the causal model M in Fig. 4(a), for example, where the exogenous variables $\{U_1, U_2\}$ are shown explicitly. Its twin network is the union of the model M (factual) and the submodel M_{x^*} (counterfactual) under intervention $do(x^*)$, which is shown in Fig. 4(b). The factual (M) and counterfactual (M_{x^*}) worlds share only the exogenous variables (in this case, U_1, U_2), which constitute the invariances shared across worlds. In this twin network, the observational X and the counterfactual Y_{x^*} are connected through two paths: one through U_1 and the other through U_2. These paths correspond to two pathways from X to Y in the original causal diagram: $\tau_1 : X \leftarrow Z_2 \leftarrow Z_1 \leftarrow U_1 \rightarrow Z_1 \rightarrow Z_2 \rightarrow Y$, and $\tau_2 : X \leftarrow Z_2 \leftarrow U_2 \rightarrow Z_2 \rightarrow Y$.

Note that when considering the corresponding paths in the original graph (Fig. 4(a)), these paths (τ_1, τ_2) are not necessarily simple, i.e., they may contain a particular node more than once. Furthermore, each path can be partitioned into a pair of causal paths (say, g_l, g_r) from a common source $U_i \in U$ (e.g., τ_1 consists of a pair (g_{l_1}, g_{r_1}), where $g_{l_1} : U_1 \rightarrow Z_1 \rightarrow Z_2 \rightarrow X$, and $g_{r_1} : U_1 \rightarrow Z_1 \rightarrow Z_2 \rightarrow Y$). Indeed, these non-simple paths are referred to as *treks* in the causal inference literature, which usually has been studied in the context of linear models [Spirtes *et al.*, 2001; Sullivant *et al.*, 2010].

Definition 9 (Trek). A trek τ in G (from X to Y) is an ordered pair of causal paths (g_l, g_r) with a common exogenous source $U_i \in U$ such that $g_l \in \Pi^c(U_i, X)$ and $g_r \in \Pi^c(U_i, Y)$. The common source U_i is called the top of the trek, denoted *top*(g_l, g_r). A trek is spurious if $g_r \in \Pi^c(U_i, Y|X)$, i.e., g_r is a causal path from U_i to Y that is not intercepted by X.

We denote the set of treks from X to Y in G by $\mathcal{T}(X,Y)_G$ and spurious treks by $\mathcal{T}^s(X,Y)_G$ (G will be omitted when obvious). We introduce next an estimand for a specific spurious trek. For a spurious trek $\tau = (g_l, g_r)$ with $U_i = top(\tau)$, first let X_{g_l} denote the path-specific potential response $X_{g_l[U_i^l]}$, where U_i^l is an i.i.d. draw from the distribution $P(U_i)$. Similarly, let $Y_{x^*,g_r} = Y_{x^*,g_r[U_i^r]}$[10], where $U_i^r \sim P(U_i)$. Pure trek-specific covariance can then finally be defined.

Definition 10 (Pure Trek-Specific Spurious Covariance)**.** For a semi-Markovian model M and a spurious trek $\tau = (g_l, g_r)$ with $U_i = top(g_l, g_r)$, the pure τ-specific spurious covariance of the treatment $X = x^*$ on the outcome Y is defined as:

$$\mathrm{Cov}^{ts}_{\tau[x^*]}(X,Y) = \mathrm{Cov}(X - X_{g_l}, Y_{x^*} - Y_{x^*,g_r}).$$

In words, the differences $X - X_{g_l}$ and $Y_{x^*} - Y_{x^*,g_r}$ are simply measuring the effects of the causal paths g_l and g_r (Lem. 1), while the $\mathrm{Cov}(\cdot)$ operator is in charge of compounding them. (In the extreme case when g_l or g_r are disconnected, the pure τ-specific spurious covariance will equate to zero.) For example, the pure τ_1-specific spurious covariance $\mathrm{Cov}^{ts}_{\tau_1[x^*]}(X,Y)$ in Fig. 4(a) is

$$\mathrm{Cov}(X - X_{g_{l_1}}, Y_{x^*} - Y_{x^*,g_{r_1}}). \tag{14}$$

Note that the counterfactuals $X_{g_{l_1}}$ and $Y_{x^*,g_{r_1}}$ assign the randomized interventions $do(U_1^l), do(U_1^r)$ to the paths g_{l_1}, g_{r_1}, respectively. By Def. 6, Eq. 14 is equal to:

$$\mathrm{Cov}(X - X_{U_1^l}, Y_{x^*} - Y_{x^*,U_1^r}).$$

This quantity can be more easily seen through its graphical representation in Fig. 5 (top). The main idea is to decompose U_1 into two independent components U_1^l, U_1^r (Fig. 5b), which is then contrasted with the world in which U_1 is kept intact (a).[11] [12] We note that by Def. 6, $X = X_\emptyset$ and $Y_{x^*} = Y_{x^*,\emptyset}$. The pure τ_1-specific spurious covariance can be written as:

$$\mathrm{Cov}^{ts}_{\tau_1[x^*]}(X,Y) = \mathrm{Cov}(X_\emptyset - X_{g_{l_1}}, Y_{x^*,\emptyset} - Y_{x^*,g_{r_1}}).$$

More generally, the pure trek-specific spurious covariance for $\tau = (g_l, g_r)$ measures the covariance of variables $X_{\pi_l} - X_{\pi_l \cup \{g_l\}}$ and $Y_{x^*,\pi_r} - Y_{x^*,\pi_r \cup \{g_r\}}$, where π_l (π_r) is an arbitrary set of causal paths from U that does not contain g_l (g_r). This observation will be useful later on, which leads to the trek-specific spurious covariance.

[10] $Y_{x^*,g_r[U_i^r]}$ is the g_r-specific potential response of Y to $do(g_r[U_i^r])$ in the submodel M_{x^*}.

[11] This operation can be seen as the parallel to the pure path-specific covariance (Def. 7), with the distinct requirement that the replacement operator, used to generate the differences, is not relative to the observed X, but the corresponding U_i.

[12] To avoid clutter, Fig. 5 is a projected version of the original twin network focused on the relevant quantities (w.l.g.).

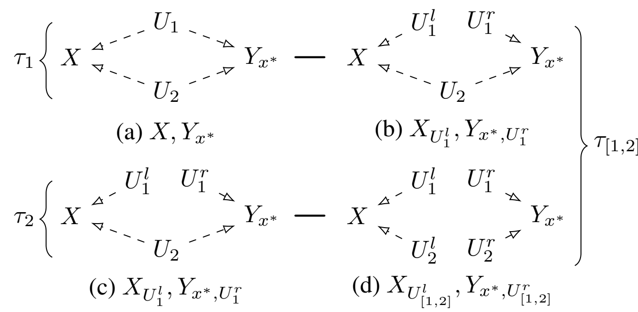


Figure 5: The decomposition procedure of the spurious covariance over the spurious treks τ_1, τ_2 (Thm. 3).

Definition 11 (Trek-Specific Spurious Covariance)**.** For a semi-Markovian model M, let τ be a spurious trek (g_l, g_r) and π is a function mapping τ to a pair $\pi(\tau) = (\pi_l, \pi_r)$ where π_l and π_r are sets of causal paths from U such that $g_l \notin \pi_l$ and $g_r \notin \pi_r$. The τ-specific spurious covariance of the treatment $X = x^*$ on the outcome Y, denoted by $\mathrm{Cov}^{ts}_{\tau[x^*]}(X,Y)_\pi$, is defined as

$$\mathrm{Cov}(X_{\pi_l} - X_{\pi_l \cup \{g_l\}}, Y_{x^*,\pi_r} - Y_{x^*,\pi_r \cup \{g_r\}}).$$

The next proposition establishes the relationship between Def. 11 and the corresponding spurious treks. This property can be seen as a necessary condition for any measure of strength for spurious relations.

Property 6. $\tau \notin \mathcal{T}^s(X,Y) \Rightarrow \mathrm{Cov}^{ts}_{\tau[x^*]}(X,Y)_\pi = 0.$

As an example of Def. 11, the trek τ_2 in Fig. 4(a) consists of paths $g_{l_2} : U_2 \to Z_2 \to X$ and $g_{r_2} : U_2 \to Z_2 \to Y$. If we set $\pi(\tau_2) = (\{g_{l_1}\}, \{g_{r_1}\})$, the τ_2-specific spurious covariance can be measured by $\mathrm{Cov}^{ts}_{\tau_2[x^*]}(X,Y)_\pi$, i.e.,

$$\mathrm{Cov}(X_{g_{l_1}} - X_{g_{l_{[1,2]}}}, Y_{x^*,g_{r_1}} - Y_{x^*,g_{r_{[1,2]}}}) \tag{15}$$

$$= \mathrm{Cov}(X_{U_1^l} - X_{U^l_{[1,2]}}, Y_{x^*,U_1^r} - Y_{x^*,U^r_{[1,2]}}). \tag{16}$$

Eq. 16 is graphically represented in Fig. 5(c-d), where the effect of the trek τ_2 is measured. In words, the difference between Fig. 5(c) and (d) is the effect of the causal paths g_{l_2} and g_{r_2} when U_2 is kept intact versus when divided into two independent components (U_2^l, U_2^r).

Armed with the definition of trek-specific spurious covariance, we can finally study the decomposability of the spurious covariance $\mathrm{Cov}^s_{x^*}(X,Y)$ (Def. 2). First, let $U^s \subseteq U$ denote the maximal set of exogenous variables that simultaneously affect variables X and Y_{x^*} (common exogenous ancestors), and let an order over U^s be $\mathcal{L}^s_u : U_1 < \cdots < U_n$. For each $U_i \in U^s$, let $\mathcal{L}^s_{l_i}$ be an order $g^i_{l_1} < \cdots < g^i_{l_n}$ over the set $\Pi^c(U_i, X)$. Similarly, we define $\mathcal{L}^s_{r_i}$ for $\Pi^c(U_i, Y|X)$. The tuple $\mathcal{L}^s = \langle \mathcal{L}^s_u, \{(\mathcal{L}^s_{l_i}, \mathcal{L}^s_{r_i})\}_{1 \le i \le |U^s|} \rangle$ thus defines an order for the spurious treks $\mathcal{T}^s(X,Y)$. We denote $\mathcal{L}^s_\pi$ a function which maps from a trek τ to sets of paths $\mathcal{L}^s_\pi(\tau)$ covered by the spurious treks preceding τ in $\mathcal{L}^s$. Formally, given a spurious trek $\tau = (g^i_{l_j}, g^i_{r_k})$, $\mathcal{L}^s_\pi(\tau)$ is equal to

$$(\Pi^c(U_{[1,i-1]}, X) \cup g^i_{l_{[1,j-1]}}, \Pi^c(U_{[1,i-1]}, Y|X) \cup g^i_{r_{[1,k-1]}}).$$

We are now ready to derive the decomposition formula for the spurious covariance $\mathrm{Cov}^s_{x^*}(X,Y)$.

Theorem 3. *For a semi-Markovian model M, let $\mathcal{L}^s = \langle \mathcal{L}^s_u, \{(\mathcal{L}^s_{l_i}, \mathcal{L}^s_{r_i})\}_{1 \le i \le |U^s|}\rangle$ be an order over spurious treks $\mathcal{T}^s(X,Y)$. For any x^*, the following non-parametric relationship hold:*

$$\mathrm{Cov}^s_{x^*}(X,Y) = \sum_{\tau \in \mathcal{T}^s(X,Y)} \mathrm{Cov}^{ts}_{\tau[x^*]}(X,Y)_{\mathcal{L}^s_\pi}$$

For example, in the model of Fig. 4(a), $U^s = \{U_1, U_2\}$. τ_1 (τ_2) is the only spurious trek associated with U_1 (U_2). If we consider the order $\mathcal{L}^s$ such that $\mathcal{L}^s_u : U_1 < U_2$, Thm. 3 dictates that $\mathrm{Cov}^s_{x^*}(X,Y)$ should be decomposed as the sum of Eqs. 14 and 15. Fig. 5 shows the graphical representation of this decomposition procedure: we measure the change of the covariance between X and Y_{x^*} as we disconnect the relations going through τ_1 (assocaited with U_1) and τ_2 (U_2), sequentially. The sum of these changes thus equates to the correlations of X and Y along the spurious pathways, i.e., the spurious covariance $\mathrm{Cov}^s_{[x^*]}(X,Y)$. (See Appendix 2 for more examples.)

6 NON-PARAMETRIC PATH ANALYSIS

In this section, we put the results of the previous sections together and derive a general path-specific decomposition for the covariance of the treatment X and the outcome Y without assuming any specific parametric form.

We start by noting that each spurious path from X to Y corresponds to a unique set of spurious treks that start on X and end in Y. Recall that a spurious path l can be seen as a pair of causal paths (g_l, g_r), where the only node shared among g_l and g_r is the common source. For example, the spurious path $l : X \leftarrow Z_2 \rightarrow Y$ is a pair (g_l, g_r) such that $g_l : Z_2 \rightarrow X$ and $g_r : Z_2 \rightarrow Y$. We can thus define a rule f which maps a trek $\tau \in \mathcal{T}^s(X,Y)$ to a spurious path $l \in \Pi^s(X,Y)$. For $\tau = (g_l, g_r)$, let V_t be the most distant recurring node from *top*(g_l, g_r) such that V_t is the only node shared among subpaths $g_l(V_t, X)$ and $g_r(V_t, Y)$; the pair $(g_l(V_t, X), g_r(V_t, Y))$ corresponds to a path l in $\Pi^s(X,Y)$. As an example, the trek τ_1 in Fig. 4(a) has $V_t = Z_2$, which corresponds to the spurious path $l : X \leftarrow Z_2 \rightarrow Y$, and similarly, $f(\tau_1) = l$ as well as $f(\tau_2) = l$. Lem. 2 shows that the rule f forms a valid surjective function.

Lemma 2. *For a semi-Markovian model M, for each spurious trek $\tau \in \mathcal{T}^s(X,Y)$, there always exists a unique most distant recurring node V_t.*

For a spurious path l, let $\mathcal{T}^s(l) = f^{-1}(l)$ denote its corresponding treks. Specifically, if $l \notin \Pi^s(X,Y)$, then for each $\tau \in \mathcal{T}^s(l)$, we must have $\tau \notin \mathcal{T}^s(X,Y)$. For instance, if the spurious l in Fig. 4(a) is disconnected, e.g., $Z_2 \not\rightarrow X$, treks τ_1, τ_2 are both disconnected as well. From this observation, we could naturally define the spurious covariance of a path l as a sum over treks in $\mathcal{T}^s(l)$.

Definition 12 (Path-Specific Spurious Covariance)**.** For a semi-Markovian model M with an associated causal diagram G, let l be an arbitrary spurious path in G. Let π be a function that maps a trek $\tau = (g_l, g_r) \in \mathcal{T}^s(l)$ to a pair $\pi(\tau) = (\pi_l, \pi_r)$, where π_l and π_r are arbitrary sets of causal paths from U such that $g_l \notin \pi_l$ and $g_r \notin \pi_r$. The l-specific spurious covariance of the treatment $X = x^*$ on the outcome Y is defined as

$$\mathrm{Cov}^s_{l[x^*]}(X,Y)_\pi = \sum_{\tau \in \mathcal{T}^s(l)} \mathrm{Cov}^{ts}_{\tau[x^*]}(X,Y)_\pi$$

Property 7. $l \notin \Pi^s(X,Y) \Rightarrow \mathrm{Cov}^s_{l[x^*]}(X,Y)_\pi = 0.$

The surjectivity of the function f assures that the set $\{\mathcal{T}^s(l)\}_{l \in \Pi^s(X,Y)}$ forms a partition over the spurious treks $\mathcal{T}^s(X,Y)$. From Thm. 3, it follows immediately that the path-specific spurious covariance (Def. 12) has the property that expresses the spurious covariance $\mathrm{Cov}^s_{x^*}(X,Y)$ as a sum over $\Pi^s(X,Y)$.

Theorem 4. *For a semi-Markovian model M, let $\mathcal{L}^s = \langle \mathcal{L}^s_u, \{(\mathcal{L}^s_{l_i}, \mathcal{L}^s_{r_i})\}_{1 \le i \le |U^s|}\rangle$ be an order over spurious treks $\mathcal{T}^s(X,Y)$. For any x^*, the following non-parametric relationship hold:*

$$\mathrm{Cov}^s_{x^*}(X,Y) = \sum_{l \in \Pi^s(X,Y)} \mathrm{Cov}^s_{l[x^*]}(X,Y)_{\mathcal{L}^s_\pi}$$

As an example, the path $l : X \leftarrow Z_2 \rightarrow Y$ in Fig. 4(a) corresponds to $\mathcal{T}^s(l) = \{\tau_1, \tau_2\}$. For an arbitrary order $\mathcal{L}^s$, Thm. 4 is applicable and immediately yields $\mathrm{Cov}^s_{x^*}(X,Y) = \mathrm{Cov}^s_{l[x^*]}(X,Y)_{\mathcal{L}^s_\pi}$, which means that the path l accounts for all the spurious relations between X and Y. In other words, the spurious joint variability of X and Y is fully explained by the variance of Z_2, which is a function of the exogenous variables U_1 (through τ_1) and U_2 (through τ_2).

Thms. 1-2 and 4 together lead to a general path-specific decomposition formula, which allows one to non-parametrically decompose the covariance $\mathrm{Cov}(X,Y)$ over all open paths from X to Y in the underlying model.

Theorem 5 (Path-Specific Decomposition)**.** *For a semi-Markovian model M, let $\mathcal{L}^c$ be an order over $\Pi^c(X,Y)$*

and $\mathcal{L}^s = \langle \mathcal{L}^s_u, \{(\mathcal{L}^s_{l_i}, \mathcal{L}^s_{r_i})\}_{1 \leq i \leq |U^s|}\rangle$ be an order over $\mathcal{T}^s(X, Y)$. For any x^*, the following non-parametric relationship hold:

$$\begin{aligned}\operatorname{Cov}(X,Y) = &\sum_{l \in \Pi^c(X,Y)} \operatorname{Cov}^c_{l[x^*]}(X,Y)_{\mathcal{L}^c_\pi} \\ &+ \sum_{l \in \Pi^s(X,Y)} \operatorname{Cov}^s_{l[x^*]}(X,Y)_{\mathcal{L}^s_\pi}.\end{aligned} \tag{17}$$

We illustrate the use of Thm. 5 using the model discussed in Sec. 1 (Fig. 1(a)). Recall that X and Y are connected through the causal paths l_1, l_2 and spurious paths l_3, l_4. Note that $U^s = \{U_Z\}$ spuriously affects the treatment X through the path $g_l = U_Z \to Z \to X$, and the outcome Y through the paths $g_{r_1} = U_Z \to Z \to Y$ and $g_{r_2} = U_Z \to Z \to W \to Y$. Let order $\mathcal{L}^c$ be $l_1 < l_2$ and $\mathcal{L}^s_r$ be $g_{r_1} < g_{r_2}$. For any level x^*, Thm. 5 equates the covariance $\operatorname{Cov}(X, Y)$ to the sum of $\{\operatorname{Cov}^c_{l_i[x^*]}(X,Y)_{\mathcal{L}^c_\pi}\}_{i=1,2}$ and $\{\operatorname{Cov}^s_{l_i[x^*]}(X,Y)_{\mathcal{L}^s_\pi}\}_{i=3,4}$, which can be written as

$$\begin{aligned}&\underbrace{\operatorname{Cov}(X, Y - Y_{x^*,W})}_{l_1: X \to Y} + \underbrace{\operatorname{Cov}(X, Y_{x^*,W} - Y_{x^*})}_{l_2: X \to W \to Y} \\ &+ \underbrace{\operatorname{Cov}(X - X_{U^l_Z}, Y_{x^*} - Y_{x^*, W_{x^*} Z_{U^r_Z}})}_{l_3: X \leftarrow Z \to Y} \\ &+ \underbrace{\operatorname{Cov}(X - X_{U^l_Z}, Y_{x^*, W_{x^*}, Z_{U^r_Z}} - Y_{x^*, U^r_Z})}_{l_4: X \leftarrow Z \to W \to Y},\end{aligned} \tag{18}$$

which are all well-defined, computable from the structural causal model [Def. 1; Pearl, 2000, Sec. 7.1].

7 IDENTIFYING PATH-SPECIFIC DECOMPOSITION

By and large, identifiability is one of the most studied topics in causal inference. It is acknowledged in the literature that obtaining identifiability may be non-trivial even in the context of less granular measures of causal effects, including quantities without nested counterfactual and following the do-calculus analysis.

In this section, we start the study of identifiability conditions for when the path-specific decomposition formula (Thm. 5) can be estimated from data, when the SCM is not fully known. We'll analyze the causal model discussed in Fig. 1(a) given its generality and potential to encode more complex models. The main assumption encoded in this model is Markovianity, i.e., that all exogenous variables are independent. We show next that identifiability can be obtained under these assumptions.

Theorem 6. *The path-specific decomposition of Eq. 18 is identifiable if the distributions $P(x, y_{x^*}), P(x, y_{x^*,W})$ and $P(x, y_{x^*, W_{x^*}, Z_{U^r_Z}})$ are identifiable. Specifically, in the model of Fig. 1(a), $P(x, y_{x^*}), P(x, y_{x^*,W})$, and $P(x, y_{x^*, W_{x^*}, Z_{U^r_Z}})$ can be estimated, respectively, from the observational distribution $P(x, y, z, w)$ as follows:*

$$P(x, y_{x^*}) = \sum_{z,w} P(y|x^*, w, z) P(w|x^*, z) P(x, z)$$

$$P(x, y_{x^*,W}) = \sum_{z,w} P(y|x^*, z, w) P(x, z, w)$$

$$P(x, y_{x^*, W_{x^*}, Z_{U^r_Z}}) = \sum_{z,z',w} P(y|x^*, z, w) P(w|x^*, z') P(x, z') P(z)$$

Note that all the quantities listed in Thm. 6 are expressible in terms of conditional distributions and do not involve any counterfactual (simple nor nested), which are readily estimable from the observational distribution. As an example, the l_2-specific causal covariance $\operatorname{Cov}^c_{l_2[x^*]}(X,Y)_{\mathcal{L}^c_\pi}$ in Eq. 18 can be written as $\operatorname{Cov}(X, Y_{x^*,W}) - \operatorname{Cov}(X, Y_{x^*})$, which is computed from the counterfactual distributions $P(x, y_{x^*})$ and $P(x, y_{x^*,W})$, respectively. These distributions can be estimated from the observational distribution $P(x, y, z, w)$ following Thm. 6. Indeed, the path-specific decomposition formula (Thm. 5) is identifiable in the model of Fig. 1(a) regardless of the order $\mathcal{L}^c$ and $\mathcal{L}^s$. (For identifications of other decompositions, see Appendix 1.)

We further considered the identifiability conditions for the path-specific decomposition formula when the more stringent assumption that the underlying structural functions are linear is imposed.

Theorem 7. *Under the assumption of linearity and the assumption of Fig. 1(a), for any arbitrary orders $\mathcal{L}^c$ and $\mathcal{L}^s$, for any x^*, the path-specific covariance of l_1, l_2, l_3 and l_4 are equal to:*

$\operatorname{Cov}^c_{l_1[x^*]}(X,Y)_{\mathcal{L}^c_\pi} = \alpha_{YX}$, $\operatorname{Cov}^c_{l_2[x^*]}(X,Y)_{\mathcal{L}^c_\pi} = \alpha_{WX}\alpha_{YW}$

$\operatorname{Cov}^s_{l_3[x^*]}(X,Y)_{\mathcal{L}^s_\pi} = \alpha_{XZ}\alpha_{YZ}$, $\operatorname{Cov}^s_{l_4[x^*]}(X,Y)_{\mathcal{L}^s_\pi} = \alpha_{XZ}\alpha_{WZ}\alpha_{YW}$

The parameters α can be estimated from the corresponding (partial) regression coefficients [Pearl, 2000, Ch. 5].

Clearly, after applying Thm. 7 to Eq. 18, the resulting decomposition coincides with Wright's method of path coefficients in the linear-standard model (Eq. 1).

8 CONCLUSIONS

We introduced novel covariance-based counterfactual measures to account for effects along with a specific path from a treatment X to an outcome Y (Defs. 8, 11-12). We developed machinery to allow, for the first time, the non-parametric decomposition of the covariance of X and Y as a summation over the different pathways in the underlying causal model (Thm. 5). We further provided identification conditions under which the decomposition formula can be estimated from data (Thm. 6-7).

Acknowledgments

Bareinboim and Zhang are supported in parts by grants from NSF IIS-1704352 and IIS-1750807 (CAREER).

References

C. Avin, I. Shpitser, and J. Pearl. Identifiability of path-specific effects. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence IJCAI-05*, pages 357–363, Edinburgh, UK, 2005. Morgan-Kaufmann Publishers.

A. Balke and J. Pearl. Counterfactual probabilities: Computational methods, bounds, and applications. In R. Lopez de Mantaras and D. Poole, editors, *Uncertainty in Artificial Intelligence 10*, pages 46–54. Morgan Kaufmann, San Mateo, CA, 1994.

E. Bareinboim and J. Pearl. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113:7345–7352, 2016.

Reuben M Baron and David A Kenny. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6):1173, 1986.

K.A. Bollen. *Structural Equations with Latent Variables*. John Wiley, NY, 1989.

RM Daniel, BL De Stavola, SN Cousens, and Stijn Vansteelandt. Causal mediation analysis with multiple mediators. *Biometrics*, 71(1):1–14, 2015.

O.D. Duncan. *Introduction to Structural Equation Models*. Academic Press, New York, 1975.

J. Fox. Effect analysis in structural equation models. *Sociological Methods and Research*, 9(1):3–28, 1980.

Kosuke Imai, Luke Keele, and Teppei Yamamoto. Identification, inference and sensitivity analysis for causal mediation effects. *Statist. Sci.*, 25(1):51–71, 02 2010.

Kosuke Imai, Luke Keele, Dustin Tingley, and Teppei Yamamoto. Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, 105(4):765–789, 2011.

Niki Kilbertus, Mateo Rojas Carulla, Giambattista Parascandolo, Moritz Hardt, Dominik Janzing, and Bernhard Schölkopf. Avoiding discrimination through causal reasoning. In *Advances in Neural Information Processing Systems*, pages 656–666, 2017.

Matt J Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. Counterfactual fairness. In *Advances in Neural Information Processing Systems*, pages 4069–4079, 2017.

Xintao Wu Lu Zhang, Yongkai Wu. A causal framework for discovering and removing direct and indirect discrimination. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 3929–3935, 2017.

D.P. MacKinnon. *An Introduction to Statistical Mediation Analysis*. Lawrence Erlbaum Associates, New York, 2008.

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, 2000.

J. Pearl. Direct and indirect effects. In *Proc. of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pages 411–420. Morgan Kaufmann, CA, 2001.

Ilya Shpitser and Eric Tchetgen Tchetgen. Causal inference with a graphical hierarchy of interventions. *Annals of statistics*, 44(6):2433, 2016.

Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, prediction, and search*. MIT press, 2001.

Seth Sullivant, Kelli Talaska, and Jan Draisma. Trek separation for gaussian graphical models. *The Annals of Statistics*, pages 1665–1685, 2010.

Eric J Tchetgen Tchetgen and Ilya Shpitser. Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. *Annals of statistics*, 40(3):1816, 2012.

Tyler J VanderWeele, Stijn Vansteelandt, and James M Robins. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology*, 25(2):300, 2014.

Tyler VanderWeele. *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press, New York, 2015.

Stijn Vansteelandt and Tyler J VanderWeele. Natural direct and indirect effects on the exposed: effect decomposition under weaker assumptions. *Biometrics*, 68(4):1019–1027, 2012.

S. Wright. The theory of path coefficients: A reply to Niles' criticism. *Genetics*, 8:239–255, 1923.

Sewall Wright. The method of path coefficients. *The annals of mathematical statistics*, 5(3):161–215, 1934.

Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rodriguez, and Krishna P Gummadi. Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *Proceedings of the 26th International Conference on WWW*, pages 1171–1180, 2017.

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

Junzhe Zhang and Elias Bareinboim. Non-parametric path analysis in structural causal models. Technical Report R-34, AI Lab, Purdue University., 2018b.

Non-Parametric Path Analysis in Structural Causal Models Supplemental Material

1 PROOFS

Proofs build on the exclusion and independence restrictions rules of SCMs [Pearl, 2000, pp. 232], and three axioms of structural counterfactuals: composition, effectiveness, and reversibility [Pearl, 2000, Ch.7.3.1].

Proof of Property 1. If X has no spurious path connecting Y in G, the independence relation $Y_{x^*} \perp\!\!\!\perp X$ must hold for any x^* [Pearl, 2000, Ch. 11.3.2], which gives:

$$\mathrm{Cov}^s_{x^*}(X,Y) = \mathrm{Cov}(X, Y_{x^*}) = 0. \qquad \square$$

Proof of Property 2. If X has no causal path connecting Y in G, then for any x^*, $Y_{x^*} = Y$. This implies:

$$\mathrm{Cov}^c_{x^*}(X,Y) = \mathrm{Cov}(X, Y - Y) = 0 \qquad \square$$

Proof of Property 3. We first consider the total direct covariance. To prove $\mathrm{Cov}^{dt}_{x^*}(X,Y) = 0$, it suffices to show that for any x, x^*, y,

$$P(x, y_{x^*,W}) = P(x,y). \tag{1}$$

Let $PA = Pa(Y)$. Conditioned on PA, W, $P(x, y_{x^*,W})$ can be written as:

$$\begin{aligned} P(x, y_{x^*,W}) &= \sum_{w,pa} P(x, y_{x^*,w} | pa, w) P(pa, w) \\ &= \sum_{w,pa} P(x, y_{x^*,w} | pa_w, w) P(pa, w). \end{aligned}$$

The last step holds by the composition axiom: for any u, if $W(u) = w$, then $PA(u) = PA_w(u)$. We will next show that for any u, w, x^*,

$$PA_w(u) = PA_{x^*,w}(u). \tag{2}$$

We will prove this statement by contradictions. If Eq. 2 does not hold, there must exist a unblocked causal path from X to a node in PA given W [Galles and Pearl, 1997]. Since PA are the parents of Y and $X \notin Pa(Y)$, we can find a indirect path from X to Y given W, which contradicts the definition of mediators. Eq. 2 implies that:

$$\begin{aligned} &\sum_{w,pa} P(x, y_{x^*,w} | pa_w, w) P(pa, w) \\ &= \sum_{w,pa} P(x, y_{x^*,w} | pa_{x^*,w}, w) P(pa, w) \\ &= \sum_{w,pa} P(x, y_{x^*,pa,w} | pa_{x^*,w}, w) P(pa, w) \\ &= \sum_{w,pa} P(x, y_{pa,w} | pa_{x^*,w}, w) P(pa, w) \end{aligned}$$

The last steps hold by the assumption that $X \not\to Y$: since all parents of Y are fixed, the exclusion restrictions rule gives $Y_{x^*,pa,w}(u) = Y_{pa,w}(u)$ for any u, x^*, pa, w. Applying Eq. 2 and the composition axiom again gives:

$$\begin{aligned} &P(x, y_{x^*,W}) \\ &= \sum_{w,pa} P(x, y_{pa,w} | pa_{x^*,w}, w) P(pa, w) \\ &= \sum_{w,pa} P(x, y_{pa,w} | pa, w) P(pa, w) \\ &= \sum_{w,pa} P(x, y | pa, w) P(pa, w) = P(x,y), \end{aligned}$$

which gives Eq. 1. To prove the pure direct covariance $\mathrm{Cov}^{dp}_{x^*}(X,Y) = 0$, it suffices to show that for any x, x^*, y,

$$P(x, y_{x^*}) = P(x, y_{W_{x^*}}). \tag{3}$$

By expanding on $W_{x^*}, PA_{W_{x^*}}$, $P(y_{W_{x^*}})$ is equal to:

$$\begin{aligned} P(x, y_{W_{x^*}}) &= \sum_{w,pa} P(x, y_w | pa_w, w_{x^*}) P(pa_w, w_{x^*}) \\ &= \sum_{w,pa} P(x, y_{pa,w} | pa_w, w_{x^*}) P(pa_w, w_{x^*}). \end{aligned}$$

The last step holds by the composition axiom: If $PA_w = pa$, then $Y_w = Y_{pa,w}$. Since $X \not\to Y$ in G, we have $Y_{pa,w} = Y_{x^*,pa,w}$, which gives:

$$\sum_{w,pa} P(x, y_{pa,w}|pa_w, w_{x^*})P(pa_w, w_{x^*}) \\ = \sum_{w,pa} P(x, y_{x^*,pa,w}|pa_w, w_{x^*})P(pa_w, w_{x^*}).$$

Applying Eq. 2 gives:

$$\sum_{w,pa} P(x, y_{x^*,pa,w}|pa_w, w_{x^*})P(pa_w, w_{x^*}) \\ = \sum_{w,pa} P(x, y_{x^*,pa,w}|pa_{x^*,w}, w_{x^*})P(pa_{x^*,w}, w_{x^*}) \\ = \sum_{w,pa} P(x, y_{x^*,pa,w}|pa_x, w_{x^*})P(pa_x, w_{x^*}).$$

The last step holds by the composition axiom: if $W_{x^*}(u) = w$, $PA_{x^*,w}(u) = PA_x(u)$. Apply the composition axiom on $Y_{x^*,pa,w}(u)$ again gives:

$$P(x, y_{W_{x^*}}) \\ = \sum_{w,pa} P(x, y_{x^*,pa,w}|pa_x, w_{x^*})P(pa_x, w_{x^*}) \\ = \sum_{w,pa} P(x, y_x|pa_x, w_{x^*})P(pa_x, w_{x^*}) = P(x, y_{x^*}),$$

which proves Eq. 3. □

Proof of Property 4. Without loss of generality, we suppose the mediators $|W| > 0$. To prove the pure indirect covariance $\mathrm{Cov}^{ip}_{x^*}(X, Y) = 0$, it suffices to show that for any x^*, u,

$$Y_{W_{x^*(u)}}(u) = Y(u). \quad (4)$$

We will first show that if $|\Pi^i(X, Y)| = 0$, then for any x^*, u, w, one of the following equation must hold

$$Y_w(u) = Y(u), \quad (5)$$
$$W_{x^*}(u) = W(u). \quad (6)$$

Suppose that Eq. 5 and 6 both fail, there must exist a unblocked causal path from X to W and a unblocked causal path from W to Y. We then find an indirect path from X to Y, which is a contradiction. Either Eq. 5 or 6 imply Eq. 4.

To prove the total indirect covariance $\mathrm{Cov}^{it}_{x^*}(X, Y) = 0$, it suffices to show that for any x^*, u,

$$Y_{x^*}(u) = Y_{x^*,W(u)}(u). \quad (7)$$

Similarly, We will show that if $|\Pi^i(X, Y)| = 0$, then for any x^*, u, w, Eq. 6 and the following equation cannot both be false:

$$Y_{x^*,w}(u) = Y_{x^*}(u), \quad (8)$$

Suppose Eq. 6 and 8 both fail, there must exist a unblocked causal path from X to W and a unblocked causal path from W to Y given X. Since removing conditioning nodes only opens up more causal path, we then find a indirect path from X to Y, which is a contradiction. Either Eq. 6 or 8 imply Eq. 7. □

Proof of Theorem 1. By basic mathematical operations, $\mathrm{Cov}(X, Y)$ can be written as:

$$\mathrm{Cov}(X, Y) = \mathrm{Cov}(X, Y - Y_{x^*}) + \mathrm{Cov}(X, Y_{x^*}) \\ = \mathrm{Cov}^c_{x^*}(X, Y) + \mathrm{Cov}^s_{x^*}(X, Y).$$

$\mathrm{Cov}^c_{x^*}(X, Y)$ can be further decomposed as:

$$\mathrm{Cov}^c_{x^*}(X, Y) = \mathrm{Cov}(X, Y) - \mathrm{Cov}(X, Y_{x^*}) \\ = \mathrm{Cov}(X, Y) - \mathrm{Cov}(X, Y_{x^*,W}) \\ + \mathrm{Cov}(X, Y_{x^*,W}) - \mathrm{Cov}(X, Y_{x^*}) \\ = \mathrm{Cov}^{dp}_{x^*}(X, Y) + \mathrm{Cov}^{it}_{x^*}(X, Y).$$

By replacing the term $\mathrm{Cov}(X, Y_{x^*,W})$ in the above equation with $\mathrm{Cov}(X, Y_{W_{x^*}})$, we have:

$$\mathrm{Cov}^c_{x^*}(X, Y) = \mathrm{Cov}(X, Y) - \mathrm{Cov}(X, Y_{x^*}) \\ = \mathrm{Cov}(X, Y) - \mathrm{Cov}(X, Y_{W_{x^*}}) \\ + \mathrm{Cov}(X, Y_{W_{x^*}}) - \mathrm{Cov}(X, Y_{x^*}) \\ = \mathrm{Cov}^{ip}_{x^*}(X, Y) + \mathrm{Cov}^{dt}_{x^*}(X, Y). \quad \square$$

Proof of Corollary 1. In the linear-standard model, values of X, Y, Z, W are decided by the following functions:

$$z = u_Z, \quad x = \alpha_{XZ} z + u_X, \quad w = \alpha_{WX} x + \alpha_{WZ} z + u_W, \\ y = \alpha_{YX} x + \alpha_{YZ} z + \alpha_{YW} w + u_Y.$$

Computing $\mathrm{Cov}^s_{x^*}(X, Y)$ gives:

$$\mathrm{Cov}^s_{x^*}(X, Y) = \mathrm{Cov}(X, Y_{x^*}) \\ = \mathrm{Cov}(X, \alpha_{YX} x^* + \alpha_{YZ} Z + \alpha_{YW} W_{x^*} + U_Y) \\ = \alpha_{YZ}\mathrm{Cov}(X, Z) + \alpha_{YW}\mathrm{Cov}(X, W_{x^*}) \\ = \alpha_{YZ}\mathrm{Cov}(X, Z) + \alpha_{YW}\mathrm{Cov}(X, \alpha_{WX} x^* + \alpha_{WZ} Z + U_W) \\ = (\alpha_{YZ} + \alpha_{YW}\alpha_{WZ})\mathrm{Cov}(X, Z) \\ = (\alpha_{YZ} + \alpha_{YW}\alpha_{WZ})\mathrm{Cov}(\alpha_{XZ} Z + U_X, Z) \\ = (\alpha_{YZ} + \alpha_{YW}\alpha_{WZ})\alpha_{XZ}\mathrm{Cov}(Z, Z) \\ = (\alpha_{YZ} + \alpha_{YW}\alpha_{WZ})\alpha_{XZ}$$

The last step holds since $\mathrm{Cov}(Z, Z) = \mathrm{Var}(Z) = 1$. We can compute $\mathrm{Cov}^{dt}_{x^*}(X, Y)$ as:

$$\begin{aligned}&\mathrm{Cov}^{dt}_{x^*}(X, Y) = \mathrm{Cov}(X, Y_{W_{x^*}} - Y_{x^*})\\&= \mathrm{Cov}(X, \alpha_{YX}X + \alpha_{YZ}Z + \alpha_{YW}W_{x^*} + U_Y)\\&- \mathrm{Cov}(X, \alpha_{YX}x^* + \alpha_{YZ}Z + \alpha_{YW}W_{x^*} + U_Y)\\&= \alpha_{YX}\mathrm{Cov}(X, X) = \alpha_{YX}.\end{aligned}$$

Similarly, $\mathrm{Cov}^{dp}_{x^*}(X, Y)$ is equal to:

$$\begin{aligned}&\mathrm{Cov}^{dp}_{x^*}(X, Y) = \mathrm{Cov}(X, Y - Y_{x^*,W})\\&= \mathrm{Cov}(X, \alpha_{YX}X + \alpha_{YZ}Z + \alpha_{YW}W + U_Y)\\&- \mathrm{Cov}(X, \alpha_{YX}x^* + \alpha_{YZ}Z + \alpha_{YW}W + U_Y)\\&= \alpha_{YX}\mathrm{Cov}(X, X) = \alpha_{YX}.\end{aligned}$$

Finally, $\mathrm{Cov}^{it}_{x^*}(X, Y)$ and $\mathrm{Cov}^{ip}_{x^*}(X, Y)$ can be written as:

$$\begin{aligned}&\mathrm{Cov}^{it}_{x^*}(X, Y) = \mathrm{Cov}(X, Y_{x^*,W} - Y_{x^*})\\&= \mathrm{Cov}(X, \alpha_{YX}x^* + \alpha_{YZ}Z + \alpha_{YW}W + U_Y)\\&- \mathrm{Cov}(X, \alpha_{YX}x^* + \alpha_{YZ}Z + \alpha_{YW}W_{x^*} + U_Y)\\&= \alpha_{YW}\mathrm{Cov}(X, W - W_{x^*})\\&= \alpha_{YW}\mathrm{Cov}(X, \alpha_{WX}X + \alpha_{WZ}Z + U_W)\\&- \alpha_{YW}\mathrm{Cov}(X, \alpha_{WX}x^* + \alpha_{WZ}Z + U_W)\\&= \alpha_{YW}\alpha_{WX}\mathrm{Cov}(X, X)\\&= \alpha_{YW}\alpha_{WX}.\\&\mathrm{Cov}^{ip}_{x^*}(X, Y) = \mathrm{Cov}(X, Y - Y_{W_{x^*}})\\&= \mathrm{Cov}(X, \alpha_{YX}X + \alpha_{YZ}Z + \alpha_{YW}W + U_Y)\\&- \mathrm{Cov}(X, \alpha_{YX}X + \alpha_{YZ}Z + \alpha_{YW}W_{x^*} + U_Y)\\&= \alpha_{YW}\mathrm{Cov}(X, W - W_{x^*})\\&= \alpha_{YW}\alpha_{WX}.\end{aligned}$$ □

Proof of Lemma 1. We will prove this lemma by showing a more general case. Let π_0, π_1 be arbitrary sets of causal paths from X such that $\pi_0 \subseteq \pi_1$. Let $\Pi^c(X, Y)_{\pi_0}$ denote the set of open causal paths in π_0 which connects Y from X, i.e., $\Pi^c(X, Y) \cap \pi_0$, so does $\Pi^c(X, Y)_{\pi_1} = \Pi^c(X, Y) \cap \pi_1$. Let $\Pi^c(X, Y)_{\pi_0,\pi_1}$ denote the difference of sets $\Pi^c(X, Y)_{\pi_1} - \Pi^c(X, Y)_{\pi_0}$. We would like to show that if $|\Pi^c(X, Y)_{\pi_0,\pi_1}| = 0$, then for any x^*, u,

$$Y_{\pi_0[x^*]}(u) = Y_{\pi_1[x^*]}(u). \tag{9}$$

We will prove this statement by induction on the length N of the longest causal path in π_1.

Base Case: If $N = 0$, the means that $\pi_0 = \pi_1 = \emptyset$. By definition, $Y_{\pi_0[x^*]}(u) = Y_{\pi_1[x^*]}(u) = Y(u)$, i.e., Eq. 9 holds.

Inductive Case: Assume that for an arbitrary variable $Y \in V$ and sets of causal paths π_0, π_1 where $\pi_0 \subseteq \pi_1$ and the length of all paths in π_1 is no greater than N, $|\Pi^c(X, Y)_{\pi_0,\pi_1}| = 0$ implies that $Y_{\pi_0[x^*]}(u) = Y_{\pi_1[x^*]}(u)$. We use this assumption to prove that for π_1 with the length of paths no greater than $N + 1$, if $|\Pi^c(X, Y)_{\pi_0,\pi_1}| = 0$, then for any x^*, u, Eq. 9 holds. We will prove its contra-positive statement: if Eq. 9 does not hold for some x^*, u, then we can find a causal path g from X to Y where g is contained in π_1 but not in π_0, i.e., $|\Pi^c(X, Y)_{\pi_0,\pi_1}| > 0$.

If Eq. 9 does not hold for some x^*, u, then one of the following cases must hold:

1. There exists a variable $U_i \in U$ such that $U_i \in X_{\pi_1\to Y}, U_i \notin X_{\pi_0\to Y}$ and its treatment assignment u^*_i is different from its natural value u_i.

2. There exists a variable $X_i \in X_{\pi_1\to Y}, X_i \notin X_{\pi_0\to Y}$ and $x^*_i \neq X_{i_{\lhd_{X_i\to Y}(\pi_0)[x^*]}}(u)$.

3. There exists a variable $S_i \in (Pa(Y)_G \cap V) - X_{\pi_i\to Y}$ such that $S_{i_{\lhd_{S_i\to Y}(\pi_0)[x^*]}}(u) \neq S_{i_{\lhd_{S_i\to Y}(\pi_1)[x^*]}}(u)$.

We will next show that for each of the above cases, one can find a causal path $g \in \Pi^c(X, Y)_{\pi_0,\pi_1}$. As for Case. 1 and 2, it immediately follows that the direct links $U_i \to Y$ and $X_i \to Y$ construct a causal paths $g \in \Pi^c(X, Y)_{\pi_0,\pi_1}$ respectively.

As for Case. 3, by the assumption, $S_{i_{\lhd_{S_i\to Y}(\pi_0)[x^*]}}(u) \neq S_{i_{\lhd_{S_i\to Y}(\pi_1)[x^*]}}(u)$ implies that there exists a path $g_s \in \Pi^c(X, S_i)_{\lhd_{S_i\to Y}(\pi_0),\lhd_{S_i\to Y}(\pi_1)}$. We can then construct a casual path g from X to Y by appending the edge $S_i \to Y$ to g_s. By the definition of the funnel operator $\lhd_{S_i\to Y}(\cdot)$, we must have $g \in \Pi^c(X, Y)_{\pi_0,\pi_1}$.

To prove Lem. 1, let $\pi_0 = \pi(g), \pi_1 = \pi(g) \cup \{g\}$. If $g \notin \Pi^c(X, Y)$, then $|\Pi^c(X, Y)_{\pi_0,\pi_1}| = 0$. This implies that Eq. 9 holds, i.e.,

$$Y_{\pi(g)[x^*]}(u) = Y_{\pi(g)\cup\{g\}[x^*]}(u).$$ □

Proof of Property 5. Lem. 1 implies that for any x^*, u,

$$g \notin \Pi^c(X, Y) \Rightarrow Y_{\pi(g)[x^*]}(u) = Y_{\pi(g)\cup\{g\}[x^*]}(u).$$

This gives

$$\begin{aligned}&\mathrm{Cov}^c_{g[x^*]}(X, Y)_\pi\\&= \mathrm{Cov}(X, Y_{\pi(g)[x^*]} - Y_{\pi(g)\cup\{g\}[x^*]}) = 0.\end{aligned}$$ □

Proof of Theorem 2. By definition,

$$Y_{\emptyset[x^*]}(u) = Y(u)$$

Following the order $\mathcal{L}^c$, let $g_{[1,n]}$ denote $\Pi^c(X,Y)$. In the model associated with the $g_{[1,n]}$-specific counterfactual $Y_{g_{[1,n]}[x^*]}(u)$, all variables are under the influence of the intervention $X = x^*$, i.e.,

$$Y_{g_{[1,n]}[x^*]}(u) = Y_{x^*}(u)$$

Thus, the causal covariance $\mathrm{Cov}^c_x(X,Y)$ is equal to:

$$\begin{aligned}
&\mathrm{Cov}^c_x(X,Y) = \mathrm{Cov}(X, Y - Y_{x^*}) \\
&= \mathrm{Cov}(X, Y_{\emptyset[x^*]} - Y_{g_{[1,n]}[x^*]}) \\
&= \mathrm{Cov}(X, Y_{\emptyset[x^*]} - Y_{g_1[x^*]}) \\
&+ \mathrm{Cov}(X, Y_{g_1[x^*]} - Y_{g_{[1,n]}[x^*]}) \\
&= \mathrm{Cov}(X, Y_{\mathcal{L}^c_\pi(g_1)[x^*]} - Y_{\mathcal{L}^c_\pi(g_1)\cup\{g_1\}[x^*]}) \\
&+ \mathrm{Cov}(X, Y_{g_1[x^*]} - Y_{g_{[1,n]}[x^*]}) \\
&= \mathrm{Cov}^c_{g_1[x^*]}(X,Y)_{\mathcal{L}^c_\pi} \\
&+ \mathrm{Cov}(X, Y_{\pi_1[x^*]} - Y_{\pi_n[x^*]}) \\
&\vdots \\
&= \sum_{i=1}^n \mathrm{Cov}^c_{g_i[x^*]}(X,Y)_{\mathcal{L}^c_\pi}.
\end{aligned}$$

Reorganizing the above equation gives

$$\mathrm{Cov}^c_x(X,Y) = \sum_{g\in\Pi^c(X,Y)} \mathrm{Cov}^c_{g[x^*]}(X,Y)_{\mathcal{L}^c_\pi} \qquad \square$$

Proof of Property 6. Let $U_i = top(\tau)$ and $(\pi_l, \pi_r) = \pi(\tau)$. $\tau = (g_l, g_r) \notin \mathcal{T}^s(X,Y)$ implies one the following conditions:

$$g_l \notin \Pi^c(U_i, X), \tag{10}$$
$$g_r \notin \Pi^c(U_i, Y|X). \tag{11}$$

By Lem. 1, Eq. 10 implies that for any x^*, u, u^l_i:

$$X_{\pi_l[u^l_i]}(u) = X_{\pi_l\cup\{g_l\}[u^l_i]}(u).$$

Eq. 11 implies that in the submodel M_{x^*} with an associated causal diagram G_{x^*} where all incoming edges of X are removed, $g_r \notin \Pi^c(U_i, Y)_{G_{x^*}}$. By the definition of the submodel M_{x^*} [Pearl, 2000, Ch. 7.1], the counterfactual Y_{x^*} is the outcome Y in the submodel M_{x^*}. By Lem. 1, we then have, for any x^*, u, u^r_i,

$$Y_{x^*,\pi_r[u^r_i]}(u) = Y_{x^*,\pi_r\cup\{g_r\}[u^r_i]}(u).$$

The τ-specific spurious covariance thus equates to:

$$\begin{aligned}
&\mathrm{Cov}^{ts}_{\tau[x^*]}(X,Y) \\
&= \mathrm{Cov}(X_\pi - X_{\pi\cup\{g_l\}}, Y_{x^*,\pi_r} - Y_{x^*,\pi_r\cup\{g_r\}}) = 0. \square
\end{aligned}$$

To prove Thm. 3, we first introduce two lemmas.

Lemma 3. *For a semi-Markovian model M, let an order over U^s be $\mathcal{L}^s_u : U_1 < \cdots < U_n$. For any x^*, $\mathrm{Cov}^s_{x^*}(X,Y)$ can be expressed as:*

$$\sum_{i=1}^n \underbrace{\mathrm{Cov}(X_{U^l_{[1,i-1]}} - X_{U^l_{[1,i]}}, Y_{x^*,U^r_{[1,i-1]}} - Y_{x^*,U^r_{[1,i]}})}_{\text{denoted by } \mathrm{Cov}^s_{x^*}(X,Y)_{U_i}}.$$

Proof. Following the order $\mathcal{L}^s_u$, let $U_{[1,n]}$ denote U^s. We will use $I\{\cdot\}$ to represent an indicator function. Since the exogenous variables $U, U^l_{[1,n]}, U^r_{[1,n]}$ explain all the uncertainties of variables $X_{U^l_{[1,n]}}$ and $Y_{x^*,U^r_{[1,n]}}$, we must have:

$$\begin{aligned}
&P(x_{U^l_{[1,n]}}, y_{x^*,U^r_{[1,n]}}) \\
&= \sum_u \sum_{u^l_{[1,n]}} \sum_{u^r_{[1,n]}} I\{X_{u^l_{[1,n]}}(u) = x\} I\{Y_{x^*,u^r_{[1,n]}}(u) = y\} \\
&\cdot P(u) P(u^l_{[1,n]}) P(u^r_{[1,n]}).
\end{aligned}$$

Let U_X denote the set of exogenous variables which affect X other than U^c. Similarly, we define U_Y for Y_{x^*}. Since U^s is the maximal set of exogenous variables that affects both X and Y_{x^*}, we must have $U_X \cap U_Y = \emptyset$. The above equation can thus be written as:

$$\begin{aligned}
&P(x_{U^l_{[1,n]}}, y_{x^*,U^r_{[1,n]}}) \\
&= \sum_u \sum_{u^l_{[1,n]}} \sum_{u^r_{[1,n]}} I\{X(u^l_{[1,n]}, u_X) = x\} \\
&\cdot I\{Y_{x^*}(u^r_{[1,n]}, u_Y) = y\} P(u) P(u^l_{[1,n]}) P(u^r_{[1,n]}) \\
&= \sum_{u_X} \sum_{u^l_{[1,n]}} I\{X(u^l_{[1,n]}, u_X) = x\} P(u_X) P(u^l_{[1,n]}) \\
&\cdot \sum_{u_Y} \sum_{u^r_{[1,n]}} I\{Y_{x^*}(u^r_{[1,n]}, u_Y) = y\} P(u_Y) P(u^r_{[1,n]}) \\
&= \sum_{u^r_{[1,n]}} P(X(u^l_{[1,n]}) = x) P(u^l_{[1,n]}) \\
&\cdot \sum_{u^r_{[1,n]}} P(Y_{x^*}(u^r_{[1,n]}) = y) P(u^r_{[1,n]}) \qquad (12) \\
&= P(x_{U^l_{[1,n]}}) P(y_{x^*,U^r_{[1,n]}}). \qquad (13)
\end{aligned}$$

Eq. 13 implies that $X_{U^l_{[1,n]}} \perp\!\!\!\perp Y_{x^*,U^r_{[1,n]}}$, i.e.,

$$\mathrm{Cov}(X_{U^l_{[1,n]}}, Y_{x^*,U^r_{[1,n]}}) = 0.$$

Since $U_{[1,0]} = \emptyset$, the spurious covariance $\mathrm{Cov}^s_{x^*}(X,Y)$ can be written as

$$\begin{aligned}
&\mathrm{Cov}^s_{x^*}(X,Y) = \mathrm{Cov}(X, Y_{x^*}) \\
&= \mathrm{Cov}(X_{U^l_{[1,0]}}, Y_{x^*,U^r_{[1,0]}}) - \mathrm{Cov}(X_{U^l_{[1,n]}}, Y_{x^*,U^r_{[1,n]}}) \\
&= \mathrm{Cov}(X_{U^l_{[1,0]}}, Y_{x^*,U^r_{[1,0]}}) - \mathrm{Cov}(X_{U^l_{[1,1]}}, Y_{x^*,U^r_{[1,1]}}) \\
&+ \mathrm{Cov}(X_{U^l_{[1,1]}}, Y_{x^*,U^r_{[1,1]}}) - \mathrm{Cov}(X_{U^l_{[1,n]}}, Y_{x^*,U^r_{[1,n]}})
\end{aligned}$$

$$\vdots$$

$$= \sum_{i=1}^{n} \mathrm{Cov}(X_{U^l_{[1,i-1]}}, Y_{x^*, U^r_{[1,i-1]}}) - \mathrm{Cov}(X_{U^l_{[1,i]}}, Y_{x^*, U^r_{[1,i]}})$$

We will next show that

$$\begin{aligned}&\mathrm{Cov}(X_{U^l_{[1,i-1]}}, Y_{x^*, U^r_{[1,i-1]}}) - \mathrm{Cov}(X_{U^l_{[1,i]}}, Y_{x^*, U^r_{[1,i]}})\\ &= \mathrm{Cov}(X_{U^l_{[1,i-1]}} - X_{U^l_{[1,i]}}, Y_{x^*, U^r_{[1,i-1]}} - Y_{x^*, U^r_{[1,i]}}).\end{aligned} \tag{14}$$

By the basic mathematical operations of covariance,

$$\begin{aligned}&\mathrm{Cov}(X_{U^l_{[1,i-1]}}, Y_{x^*, U^r_{[1,i-1]}}) - \mathrm{Cov}(X_{U^l_{[1,i]}}, Y_{x^*, U^r_{[1,i]}})\\ &= \mathrm{Cov}(X_{U^l_{[1,i-1]}} - X_{U^l_{[1,i]}}, Y_{x^*, U^r_{[1,i-1]}} - Y_{x^*, U^r_{[1,i]}})\\ &+ \mathrm{Cov}(X_{U^l_{[1,i]}}, Y_{x^*, U^r_{[1,i-1]}} - Y_{x^*, U^r_{[1,i]}})\\ &+ \mathrm{Cov}(X_{U^r_{[1,i-1]}} - X_{U^l_{[1,i]}}, Y_{x^*, U^r_{[1,i]}}).\end{aligned}$$

It suffices to prove that for any x, y,

$$P(x_{U^l_{[1,i]}}, y_{x^*, U^r_{[1,i-1]}}) = P(x_{U^l_{[1,i]}}, y_{x^*, U^r_{[1,i]}}) \tag{15}$$

$$P(x_{U^l_{[1,i-1]}}, y_{x^*, U^r_{[1,i]}}) = P(x_{U^l_{[1,i]}}, y_{x^*, U^r_{[1,i]}}). \tag{16}$$

Let us first consider Eq. 15. From Eq. 12, the distributions $P(x_{U^l_{[1,i]}}, y_{x^*, U^r_{[1,i-1]}})$ and $P(x_{U^l_{[1,i]}}, y_{x^*, U^r_{[1,i]}})$ can be written as:

$$\begin{aligned}&P(x_{U^l_{[1,i]}}, y_{x^*, U^r_{[1,i-1]}})\\ &= \sum_{u_{[i,n]}} \sum_{u^l_{[1,i]}} \sum_{u^r_{[1,i-1]}} P(X(u^l_{[1,i]}, u_{[i+1,n]}) = x)\\ &\cdot P(Y_{x^*}(u^r_{[1,i-1]}, u_{[i,n]}) = y) P(u^l_{[1,i]})\\ &\cdot P(u^r_{[1,i-1]}) P(u_i) P(u_{[i+1,n]}),\\ &P(x_{U^l_{[1,i]}}, y_{x^*, U^r_{[1,i]}})\\ &= \sum_{u_{[i+1,n]}} \sum_{u^l_{[1,i]}} \sum_{u^r_{[1,i]}} P(X(u^l_{[1,i]}, u_{[i+1,n]}) = x)\\ &\cdot P(Y_{x^*}(u^r_{[1,i]}, u_{[i+1,n]}) = y) P(u^l_{[1,i]})\\ &\cdot P(u^r_{[1,i-1]}) P(u^r_i) P(u_{[i+1,n]}).\end{aligned}$$

Since U and U^r are i.i.d. draws from the exogenous distribution $P(u)$, we have for $u_i = u^r_i$, $P(u_i) = P(u^r_i)$. Replacing u^r_i with u_i in the above equations gives Eq. 15. Similarly, we can prove Eq. 16. Eqs. 15-16 together prove Eq. 14. $\square$

Lemma 4. *For a semi-Markovian model M, let $\mathcal{T}^s(X, Y; U_i)$ denote the set of spurious treks from X to Y with a common source U_i. Let $\mathcal{L}^s = \langle \mathcal{L}^s_u, \{(\mathcal{L}^s_{l_i}, \mathcal{L}^s_{r_i})\}_{1 \le i \le |U^s|} \rangle$ be an order over spurious treks $\mathcal{T}^s(X, Y)$. For any x^*, the following non-parametric relationships hold:*

$$\mathrm{Cov}^s_{x^*}(X, Y)_{U_i} = \sum_{\tau \in \mathcal{T}^s(X,Y;U_i)} \mathrm{Cov}^{ts}_{\tau[x^*]}(X, Y)_{\mathcal{L}^s_\pi}$$

Proof. Let (π_l, π_r) denote the pair

$$(\Pi^c(U_{[1,i-1]}, X), \Pi^c(U_{[1,i-1]}, Y|X)).$$

Following the order $\mathcal{L}^s$, let $g^i_{l_{[1,n]}} = \Pi^c(U_i, X)$ and $g^i_{r_{[1,m]}} = \Pi^c(U_i, Y|X)$. Since the intervention $do(U^l_i)$ ($do(U^r_i)$) assigns a randomized treatment U^l_i (U^r_i) to all causal paths in $g^i_{l_{[1,n]}}$ ($g^i_{r_{[1,m]}}$). The term $\mathrm{Cov}^s_{x^*}(X, Y)_{U_i}$ can thus be written as:

$$\begin{aligned}&\mathrm{Cov}(X_{U^l_{[1,i-1]}} - X_{U^l_{[1,i]}}, Y_{x^*, U^r_{[1,i-1]}} - Y_{x^*, U^r_{[1,i]}})\\ &= \mathrm{Cov}(X_{\pi_l} - X_{\pi_l \cup g^i_{l_{[1,n]}}}, Y_{x^*, \pi_r} - Y_{x^*, \pi_r \cup g^i_{r_{[1,m]}}}).\end{aligned}$$

The above equation can be decomposed over causal paths in $g^i_{l_{[1,n]}}$:

$$\begin{aligned}&\mathrm{Cov}(X_{\pi_l} - X_{\pi_l \cup g^i_{l_{[1,n]}}}, Y_{x^*, \pi_r} - Y_{x^*, \pi_r \cup g^i_{r_{[1,m]}}})\\ &= \mathrm{Cov}(X_{\pi_l \cup g^i_{l_{[1,0]}}} - X_{\pi_l \cup g^i_{l_{[1,1]}}}, Y_{x^*, \pi_r} - Y_{x^*, \pi_r \cup g^i_{r_{[1,m]}}})\\ &+ \mathrm{Cov}(X_{\pi_l \cup g^i_{l_{[1,1]}}} - X_{\pi_l \cup g^i_{l_{[1,n]}}}, Y_{x^*, \pi_r} - Y_{x^*, \pi_r \cup g^i_{r_{[1,m]}}})\\ &= \mathrm{Cov}(X_{\pi_l \cup g^i_{l_{[1,0]}}} - X_{\pi_l \cup g^i_{l_{[1,1]}}}, Y_{x^*, \pi_r} - Y_{x^*, \pi_r \cup g^i_{r_{[1,m]}}})\\ &+ \mathrm{Cov}(X_{\pi_l \cup g^i_{l_{[1,1]}}} - X_{\pi_l \cup g^i_{l_{[1,2]}}}, Y_{x^*, \pi_r} - Y_{x^*, \pi_r \cup g^i_{r_{[1,m]}}})\\ &+ \mathrm{Cov}(X_{\pi_l \cup g^i_{l_{[1,2]}}} - X_{\pi_l \cup g^i_{l_{[1,n]}}}, Y_{x^*, \pi_r} - Y_{x^*, \pi_r \cup g^i_{r_{[1,m]}}})\\ &\vdots\\ &= \sum_{j=1}^{n} \mathrm{Cov}(X_{\pi_l \cup g^i_{l_{[1,j-1]}}} - X_{\pi_l \cup g^i_{l_{[1,j]}}},\\ &\qquad Y_{x^*, \pi_r} - Y_{x^*, \pi_r \cup g^i_{r_{[1,m]}}}).\end{aligned}$$

The summation term can be further decomposed over paths in $g^i_{r_{[1,m]}}$:

$$\begin{aligned}&\mathrm{Cov}(X_{\pi_l \cup g^i_{l_{[1,j-1]}}} - X_{\pi_l \cup g^i_{l_{[1,j]}}}, Y_{x^*, \pi_r} - Y_{x^*, \pi_r \cup g^i_{r_{[1,m]}}})\\ &= \mathrm{Cov}(X_{\pi_l \cup g^i_{l_{[1,j-1]}}} - X_{\pi_l \cup g^i_{l_{[1,j]}}},\\ &\qquad Y_{x^*, \pi_r \cup g^i_{r_{[1,0]}}} - Y_{x^*, \pi_r \cup g^i_{r_{[1,1]}}})\\ &+ \mathrm{Cov}(X_{\pi_l \cup g^i_{l_{[1,j-1]}}} - X_{\pi_l \cup g^i_{l_{[1,j]}}},\\ &\qquad Y_{x^*, \pi_r \cup g^i_{r_{[1,1]}}} - Y_{x^*, \pi_r \cup g^i_{r_{[1,m]}}})\\ &= \mathrm{Cov}(X_{\pi_l \cup g^i_{l_{[1,j-1]}}} - X_{\pi_l \cup g^i_{l_{[1,j]}}},\\ &\qquad Y_{x^*, \pi_r \cup g^i_{r_{[1,0]}}} - Y_{x^*, \pi_r \cup g^i_{r_{[1,1]}}})\end{aligned}$$

$$+ \operatorname{Cov}(X_{\pi_l \cup g^i_{l_{[1,j-1]}}} - X_{\pi_l \cup g^i_{l_{[1,j]}}},$$
$$Y_{x^*,\pi_r \cup g^i_{r_{[1,1]}}} - Y_{x^*,\pi_r \cup g^i_{r_{[1,2]}}})$$
$$+ \operatorname{Cov}(X_{\pi_l \cup g^i_{l_{[1,j-1]}}} - X_{\pi_l \cup g^i_{l_{[1,j]}}},$$
$$Y_{x^*,\pi_r \cup g^i_{r_{[1,2]}}} - Y_{x^*,\pi_r \cup g^i_{r_{[1,m]}}})$$
$$\vdots$$
$$= \sum_{k=1}^{m} \operatorname{Cov}(X_{\pi_l \cup g^i_{l_{[1,j-1]}}} - X_{\pi_l \cup g^i_{l_{[1,j]}}},$$
$$Y_{x^*,\pi_r \cup g^i_{r_{[1,k-1]}}} - Y_{x^*,\pi_r \cup g^i_{r_{[1,k]}}}).$$

Together, we can obtain

$$\operatorname{Cov}^s_{x^*}(X,Y)_{U_i} = \sum_{j=1}^{n}\sum_{k=1}^{m} \operatorname{Cov}^{ts}_{\tau^i_{j,k}[x^*]}(X,Y)_{\mathcal{L}^s_\pi}$$

where $\tau^i_{j,k} = (g^i_{l_j}, g^i_{r_k})$. Reorganizing the above equation gives:

$$\operatorname{Cov}^s_{x^*}(X,Y)_{U_i} = \sum_{\tau \in \mathcal{T}^s(X,Y;U_i)} \operatorname{Cov}^{ts}_{\tau[x^*]}(X,Y)_{\mathcal{L}^s_\pi}. \square$$

We are now ready to prove Thm. 3

Proof of Theorem 3. By Lem. 3 and 4, we have:

$$\operatorname{Cov}^s_{[x^*]}(X,Y) = \sum_{i=1}^{|U^s|} \sum_{\tau \in \mathcal{T}^s(X,Y;U_i)} \operatorname{Cov}^{ts}_{\tau[x^*]}(X,Y)_{\mathcal{L}^s_\pi}.$$

Reorganizing the above equation gives:

$$\operatorname{Cov}^s_{[x^*]}(X,Y) = \sum_{\tau \in \mathcal{T}^s(X,Y)} \operatorname{Cov}^{ts}_{\tau[x^*]}(X,Y)_{\mathcal{L}^s_\pi}. \quad \square$$

Proof of Lemma 2. **Existence.** We will prove the existence of V_t by proving a stronger statement: in a semi-Markovian model M, for any non-simple path of the form (g_l, g_r) where g_l, g_r share a common source V_i and have sink X and Y respectively, there always exists a most distant recurring node V_t such that V_t is the only shared common node among subpaths $g_l(V_t, X)$ and $g_r(V_t, Y)$. We will prove this statement by induction on the number of recurring node N shared among g_l, g_r.

- **Base Case:** For $N = 1$, then the common source V_i is the most distant recurring node V_t.
- **Induction Case:** Assume that for all non-simple path of the form (g_l, g_r) with N recurring nodes , there always exists a most distant recurring node V_t such that V_t is the only shared common node among subpaths $g_l(V_t, X)$ and $g_r(V_t, Y)$. We will use this assumption to prove that for all non-simple path of the form (g_l, g_r) with $N + 1$ recurring nodes, the most distant recurring node V_t also exists. For a non-simple path (g_l, g_r), we find the next recurring node V_i' of g_l, g_r other than the common source V_i. The subpaths $(g_l(V_i', X), g_r(V_i', Y))$ forms a non-simple path with N recurring nodes. By the assumption, for the non-simple path $(g_l(V_i', X), g_r(V_i', Y))$, there exists a most distant recurring node V_t such that V_t is the only node shared among subpaths $g_l(V_t, X), g_r(V_t, Y)$.

  We will show that the most distant recurring node V_t of $(g_l(V_i', X), g_r(V_i', Y))$ is also satisfied for (g_l, g_r). Suppose V_t is not a most distant recurring node of (g_l, g_r), this means that the subpaths $g_l(V_t, X), g_r(V_t, Y)$ share another common node other than V_t, which contradicts our assumption.

Uniqueness. We will prove this lemma by contradictions. Suppose there are two distinct nodes V_t^0, V_t^1 for a trek $\tau = (g_l, g_r)$ such that for $i = 0, 1$, V_t^i is the only node shared among subpaths $g_l(V_t^i, X)$ and $g_r(V_t^i, Y)$. V_t^0, V_t^1 must satisfy one of the following cases.

1. There exists a causal path from V_t^0 to V_t^1 in g_l, denoted by $(V_t^0 \to V_t^1)_{g_l}$, and a causal path from V_t^0 to V_t^1 in g_r, denoted by $(V_t^0 \to V_t^1)_{g_r}$.
2. $(V_t^1 \to V_t^0)_{g_l}$ and $(V_t^1 \to V_t^0)_{g_r}$.
3. $(V_t^0 \to V_t^1)_{g_l}$ and $(V_t^1 \to V_t^0)_{g_r}$.
4. $(V_t^1 \to V_t^0)_{g_l}$ and $(V_t^0 \to V_t^1)_{g_r}$.

For Case. 1, we must have that V_t^1 is also a common node shared among the subpaths $g_l(V_t^0, X)$ and $g_r(V_t^0, Y)$, which contradicts our assumptions. Similarly, Case. 2 lead to an contradiction, as V_t^0 is also a common node shared among the subpaths $g_l(V_t^1, X)$ and $g_r(V_t^1, Y)$.

For Case. 3, if exists a causal path from V_t^0 to V_t^1 and a causal path from V_t^1 to V_t^0, the causal diagram G of the semi-Markovian model M is not a DAG, which is a contradiction. Similarly, Case. 4 contradicts the assumption that G is a DAG. Since Cases. 1-4 all lead to contradictions, the most distant recurring node V_t is unique for each trek $\tau \in \mathcal{T}^s(X,Y)$. $\square$

Proof of Property 7. For a spurious path $l = (g_l, g_r)$ with the common source V_t, if $l \notin \Pi^s(X,Y)$, then one of the following conditions must hold:

$$g_l \not\in \Pi^c(V_t, X), \quad g_r \not\in \Pi^c(V_t, Y|X).$$

For each $\tau \in \mathcal{T}^s(l)$, g_l, g_r are both its subpaths. This implies that from the above conditions, we must have $\tau \not\in \mathcal{T}^s(X,Y)$. By Prop. 6, we have

$$\mathrm{Cov}^s_{l[x^*]}(X,Y)_\pi = \sum_{\tau \in \mathcal{T}^s(l)} \mathrm{Cov}^{ts}_{\tau[x^*]}(X,Y)_\pi = 0. \square$$

Proof of Theorem 4. Thm. 3 implies

$$\mathrm{Cov}^s_{x^*}(X,Y) = \sum_{\tau \in \mathcal{T}^s(X,Y)} \mathrm{Cov}^{ts}_{\tau[x^*]}(X,Y)_{\mathcal{L}^s_\pi}. \quad (17)$$

Since the mapping $f : \mathcal{T}^s(X,Y) \to \Pi^s(X,Y)$ is a surjective function, $\{\mathcal{T}^s(l) = f^{-1}(l)\}_{l \in \Pi^s(X,Y)}$ is a partition over the set $\mathcal{T}^s(X,Y)$. Eq. 17 could be written as:

$$\begin{aligned} \mathrm{Cov}^s_{x^*}(X,Y) &= \sum_{\tau \in \mathcal{T}^s(X,Y)} \mathrm{Cov}^{ts}_{\tau[x^*]}(X,Y)_{\mathcal{L}^s_\pi} \\ &= \sum_{l \in \Pi^s(X,Y)} \sum_{\tau \in \mathcal{T}^s(l)} \mathrm{Cov}^{ts}_{\tau[x^*]}(X,Y)_\pi \\ &= \sum_{l \in \Pi^s(X,Y)} \mathrm{Cov}^s_{l[x^*]}(X,Y)_\pi. \qquad \square \end{aligned}$$

Proof of Theorem 5. By Thm. 1, we have

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}^c_{x^*}(X,Y) + \mathrm{Cov}^s_{x^*}(X,Y).$$

Applying Thm. 2 and 5 to the above equation gives

$$\begin{aligned} \mathrm{Cov}(X,Y) = &\sum_{l \in \Pi^c(X,Y)} \mathrm{Cov}^c_{l[x^*]}(X,Y)_{\mathcal{L}^c_\pi} \\ &+ \sum_{l \in \Pi^s(X,Y)} \mathrm{Cov}^s_{l[x^*]}(X,Y)_{\mathcal{L}^s_\pi}. \qquad \square \end{aligned}$$

We will next prove Thm. 6. Recall in the standard model of Fig. 1(a), X and Y are connected with causal paths $l_1 : X \to Y$, $l_2 : X \to W \to Y$ and spurious paths $l_3 : X \leftarrow Z \to Y$ and $l_4 : X \leftarrow Z \to W \to Y$. $U^s = \{U_Z\}$ affects the treatment X through a causal path $g_l = U_Z \to Z \to X$ and the outcome Y through causal paths $g_{r_1} = U_Z \to Z \to Y$ and $g_{r_2} = U_Z \to Z \to W \to Y$. To prove Thm. 7, we will introduce following lemmas.

Lemma 5. *In the standard model (Fig. 1(a)), for an order $\mathcal{L}^c : l_1 < l_2$, the path-specific decomposition of the causal covariance $\mathrm{Cov}^c_{x^*}(X,Y)$ (Thm. 2) are identifiable if $P(x, y_{x^*,W})$ and $P(x, y_{x^*})$ are identifiable. Specifically, distributions $P(x, y_{x^*,W})$ and $P(x, y_{x^*})$ can be estimated from the observational distribution $P(x,y,z,w)$ as following:*

$$\begin{aligned} P(x, y_{x^*}) &= \sum_{z,w} P(y|x^*, z, w) P(w|x^*, z) P(x,z), \\ P(x, y_{x^*,W}) &= \sum_{z,w} P(y|x^*, z, w) P(x,z,w). \end{aligned}$$

Proof. By Thm. 2, the causal covariance $\mathrm{Cov}^c_{x^*}(X,Y)$ equates to

$$\mathrm{Cov}^c_{x^*}(X,Y) = \mathrm{Cov}^c_{l_1[x^*]}(X,Y)_{\mathcal{L}^c_\pi} + \mathrm{Cov}^c_{l_2[x^*]}(X,Y)_{\mathcal{L}^c_\pi}.$$

We will show that each quantity on the right-hand side of the above equation is identifiable from $P(x,y,z,w)$. For the order $\mathcal{L}^c : l_1 < l_2$,

$$\begin{aligned} &\mathrm{Cov}^c_{l_1[x^*]}(X,Y)_{\mathcal{L}^c_{1\pi}} = \mathrm{Cov}^{dp}_{x^*}(X,Y) \qquad (18) \\ &= \mathrm{Cov}(X, Y - Y_{x^*,W}) \\ &= \mathrm{Cov}(X,Y) - \mathrm{Cov}(X, Y_{x^*,W}), \\ &\mathrm{Cov}^c_{l_2[x^*]}(X,Y)_{\mathcal{L}^c_{1\pi}} = \mathrm{Cov}^{it}_{x^*}(X,Y) \qquad (19) \\ &= \mathrm{Cov}(X, Y_{x^*,W} - Y_{x^*}) \\ &= \mathrm{Cov}(X, Y_{x^*,W}) - \mathrm{Cov}(X, Y_{x^*}). \end{aligned}$$

It suffices to show that distributions $P(x, y_{x^*,W})$ and $P(x, y_{x^*})$ are identifiable. By expanding on Z, W_{x^*}, $P(x, y_{x^*})$ can be written as:

$$\begin{aligned} &P(x, y_{x^*}) \\ &= \sum_{z,w} P(y_{x^*}|x, z, w_{x^*}) P(w_{x^*}|x, z) P(x,z) \\ &= \sum_{z,w} P(y_{x^*,z,w}|x_z, z, w_{x^*,z}) P(w_{x^*,z}|x_z, z) P(x,z). \end{aligned}$$

The last step holds due to the following reasons: (1) by the exclusion restrictions rule, since Z has no parent node in the model of Fig. 1(a), $Z = Z_{x^*}$ for any x^*; (2) by the composition axiom, we have:

$$\begin{aligned} &Z = z \Rightarrow X = X_z, \\ &Z_{x^*} = z \Rightarrow W_{x^*} = W_{x^*,z}, \\ &Z_{x^*} = z, W_{x^*} = w \Rightarrow Y_{x^*} = Y_{x^*,z,w}. \end{aligned}$$

By the independence exclusions rule, for any x^*, x, z, w,

$$\begin{aligned} W_{x^*,z} &\perp\!\!\!\perp X_z, Z, \qquad (20) \\ Y_{x^*,z,w} &\perp\!\!\!\perp X_z, Z, W_{x,z}. \qquad (21) \end{aligned}$$

We thus have:

$$\begin{aligned} &\sum_{z,w} P(y_{x^*,z,w}|x_z, z, w_{x^*,z}) P(w_{x^*,z}|x_z, z) P(x,z) \\ &= \sum_{z,w} P(y_{x^*,z,w}) P(w_{x^*,z}) P(x,z) \end{aligned}$$

Since the standard model is Markovian,

$$P(w_{x^*,z}) = P(w|x^*, z), \quad (22)$$
$$P(y_{x^*,z,w}) = P(y|x^*, z, w). \quad (23)$$

Thus,

$$P(x, y_{x^*}) = \sum_{z,w} P(y|x^*, z, w)P(w|x^*, z)P(x, z).$$

By expanding on Z, W, $P(x, y_{x^*,W})$ can be written as:

$$\begin{aligned} &P(x, y_{x^*,W}) \\ &= \sum_{z,w} P(y_{x^*,w}|x, z, w)P(x, z, w) \\ &= \sum_{z,w} P(y_{x^*,w}|x, z_{x^*,w}, w_{x,z})P(x, z, w) \end{aligned}$$

The last step holds due to following reasons: (1) By the composition axiom, $W = W_{x,z}$ if $X = x, Z = z$; (2) By the exclusion restrictions rules, $Z = Z_{x^*,w}$ if Z has no parent node. Applying the composition axiom again gives:

$$Z_{x^*,w} = z \Rightarrow Y_{x^*,w} = Y_{x^*,z,w}.$$

We thus have:

$$\begin{aligned} &\sum_{z,w} P(y_{x^*,w}|x, z_{x^*,w}, w_{x,z})P(x, z, w) \\ &= \sum_{z,w} P(y_{x^*,z,w}|x, z_{x^*,w}, w_{x,z})P(x, z, w) \end{aligned}$$

The independence relation 21 gives:

$$\begin{aligned} P(x, y_{x^*,W}) &= \sum_{z,w} P(y_{x^*,z,w}|x, z_{x^*,w}, w_{x,z})P(x, z, w) \\ &= \sum_{z,w} P(y_{x^*,z,w})P(x, z, w) \\ &= \sum_{z,w} P(y|x^*, z, w)P(x, z, w). \end{aligned}$$

The last step holds by Eq. 23. □

Lemma 6. *In the standard model (Fig. 1(a)), for a order $\mathcal{L}^s$ where $\mathcal{L}^s_r : g_{r_1} < g_{r_2}$, the path-specific decomposition of the spurious covariance $\mathrm{Cov}^s_{x^*}(X, Y)$ (Thm. 5) is identifiable if $P(x, y_{x^*}), P(x, y_{x^*,W_{x^*},Z_{U^r_Z}})$ are identifiable. Specifically, distributions $P(x, y_{x^*}), P(x, y_{x^*,W_{x^*},Z_{U^r_Z}})$ are identifiable can be estimated from the observational distribution $P(x, y, z, w)$ as following:*

$$P(x, y_{x^*}) = \sum_{z,w} P(y|x^*, z, w)P(w|x*, z)P(x, z),$$

$$P(x, y_{x^*,W_{x^*},Z_{U^r_Z}}) = \sum_{z,z',w} P(y|x^*, w, z)P(w|x^*, z') \cdot P(x, z')P(z)$$

Proof. By Thm. 5, the spurious covariance $\mathrm{Cov}^s_{x^*}(X, Y)$ equates to

$$\mathrm{Cov}^s_{x^*}(X, Y) = \mathrm{Cov}^s_{l_3[x^*]}(X, Y)_{\mathcal{L}^s_\pi} + \mathrm{Cov}^s_{l_4[x^*]}(X, Y)_{\mathcal{L}^c_\pi}.$$

We will next show that each quantity on the right-hand side of the above equation is identifiable from $P(x, y, z, w)$. In the standard model, Considering the order $\mathcal{L}^s$ where $\mathcal{L}^s_r : g_{r_1} < g_{r_2}$,

$$\begin{aligned} &\mathrm{Cov}^s_{l_3[x^*]}(X, Y)_{\mathcal{L}^s_{1_\pi}} = \mathrm{Cov}(X - X_{g_l}, Y_{x^*} - Y_{x^*,g_{r_1}}) \\ &= \mathrm{Cov}(X - X_{U^l_Z}, Y_{x^*} - Y_{x^*,W_{x^*},Z_{U^r_Z}}) \\ &= \mathrm{Cov}(X, Y_{x^*} - Y_{x^*,W_{x^*},Z_{U^r_Z}}) \quad (24) \\ &= \mathrm{Cov}(X, Y_{x^*}) - \mathrm{Cov}(X, Y_{x^*,W_{x^*},Z_{U^r_Z}}). \end{aligned}$$

The last step holds since $X_{U^l_Z}$ is an independent counterfactual variable: the variable X is function over U_X, U^l_Z; the exogenous variables U_X, U^l_Z are independent of all the other variables in the domain. Similarly,

$$\begin{aligned} &\mathrm{Cov}^s_{l_4[x^*]}(X, Y)_{\mathcal{L}^s_{1_\pi}} = \mathrm{Cov}(X - X_{g_l}, Y_{x^*,g_{r_1}} - Y_{x^*,g_{r_{[1,2]}}}) \\ &= \mathrm{Cov}(X - X_{U^l_Z}, Y_{x^*,W_{x^*},Z_{U^r_Z}} - Y_{x^*,U^r_Z}) \\ &= \mathrm{Cov}(X, Y_{x^*,W_{x^*},Z_{U^r_Z}} - Y_{x^*,U^r_Z}) \quad (25) \\ &= \mathrm{Cov}(X, Y_{x^*,W_{x^*},Z_{U^r_Z}}). \end{aligned}$$

The last two steps holds since $X_{U^l_Z}$ and Y_{x^*,U^r_Z} are independent counterfactual variables. It will suffice to show that the distributions $P(x, y_{x^*}), P(x, y_{x^*,W_{x^*},Z_{U^r_Z}})$ are identifiable. $P(x, y_{x^*})$ can be identified using Lem. 5. By conditioning on U^r_Z, $P(x, y_{x^*,W_{x^*},Z_{U^r_Z}})$ can be written as:

$$P(x, y_{x^*,W_{x^*},Z_{U^r_Z}}) = \sum_{u^r_Z} P(x, y_{x^*,W_{x^*},Z_{u^r_Z}}|u^r_Z)P(u^r_Z)$$

With U^r_Z fixed, variables X and $Y_{x^*,W_{x^*},Z_{u^r_Z}}$ are functions of the exogenous variable U, which is independent of U^r_Z. We thus have the following independence relation

$$U^r_Z \perp\!\!\!\perp X, Y_{x^*,W_{x^*},Z_{u^r_Z}},$$

which gives:

$$P(x, y_{x^*,W_{x^*},Z_{U^r_Z}}) = \sum_{u^r_Z} P(x, y_{x^*,W_{x^*},Z_{u^r_Z}})P(u^r_Z) \quad (26)$$

By expanding on $Z, Z_{U^r_Z}, W_{x^*}$, $P(x, y_{x^*,W_{x^*},Z_{u^r_Z}})$ can be written as:

$$P(x, y_{x^*,W_{x^*},Z_{u^r_Z}}) = \sum_{z,z',w} P(x, y_{x^*,w,z}, w_{x^*}, z', z_{u^r_Z})$$

Since the function f_Z takes only U_Z as an argument, the variables $Z_{u^r_Z}$ are deterministic, i.e.,

$$P(x, y_{x^*,w,z}, w_{x^*}, z', z_{u^r_Z}) \\ = P(x, y_{x^*,w,z}, w_{x^*}, z') I\{Z_{u^r_Z} = z\}$$

where $I\{\cdot\}$ is an indicator function. The above equation, together with Eq. 26, gives:

$$\begin{aligned} &P(x, y_{x^*,W_{x^*},Z_{U^r_Z}}) \\ &= \sum_{z,z',w} P(x, y_{x^*,w,z}, w_{x^*}, z') \sum_{u^r_Z} I\{Z_{u^r_Z} = z\} P(u^r_Z) \\ &= \sum_{z,z',w} P(x, y_{x^*,w,z}, w_{x^*}, z') P(z). \end{aligned} \tag{27}$$

By the composition axiom and the exclusion restrictions rule [Pearl, 2000, Ch. 7.3], in the model of Fig. 1(a), for any z, x^*,

$$\begin{aligned} &Z = z \Rightarrow X = X_z, \\ &Z = Z_{x^*}, \\ &Z_{x^*} = z \Rightarrow W_{x^*} = W_{x^*,z}, \end{aligned} \tag{28}$$

The above relations imply that:

$$P(x, y_{x^*,w,z}, w_{x^*}, z') = P(x_{z'}, y_{x^*,w,z}, w_{x^*,z'}, z')$$

The independence restrictions rule [Pearl, 2000, Ch. 7.3] implies that in the model of Fig. 1(a), counterfactuals $X_{z'}, Y_{x^*,w,z}, W_{x^*,z'}, Z$ are mutually independent. We thus obtain

$$\begin{aligned} &P(x_{z'}, y_{x^*,w,z}, w_{x^*,z'}, z') \\ &= P(x_{z'}) P(y_{x^*,w,z}) P(w_{x^*,z'}) P(z') \end{aligned} \tag{29}$$

Since the standard model is Markovian,

$$\begin{aligned} &P(x_{z'}) = P(x|z'), \\ &P(y_{x^*,w,z}) = P(y|x^*, w, z), \\ &P(w_{x^*,z'}) = P(w|x^*, z'). \end{aligned} \tag{30}$$

Eqs. 27, 29 and 30 together give

$$\begin{aligned} &P(x, y_{x^*,W_{x^*},Z_{U^r_Z}}) \\ &= \sum_{z,z',w} P(x|z') P(y|x, w, z) P(w|x^*, z') P(z') P(z) \\ &= \sum_{z,z',w} P(y|x, w, z) P(w|x^*, z') P(x, z') P(z). \end{aligned}$$ □

We are now ready to prove Thm. 6.

Proof of Theorem 6. Recall the target path-specific decomposition of $\mathrm{Cov}(X, Y)$ is:

$$\begin{aligned} &\mathrm{Cov}(X, Y - Y_{x^*,W}) + \mathrm{Cov}(X, Y_{x^*,W} - Y_{x^*}) \\ &+ \mathrm{Cov}(X - X_{U^l_Z}, Y_{x^*} - Y_{x^*,W_{x^*} Z_{U^r_Z}}) \\ &+ \mathrm{Cov}(X - X_{U^l_Z}, Y_{x^*,W_{x^*},Z_{U^r_Z}} - Y_{x^*,U^r_Z}). \end{aligned}$$

This decomposition is induced by the order $\mathcal{L}^c : l_1 < l_2$ and $\mathcal{L}^s$ where $\mathcal{L}^s_r : g_{r_1} < g_{r_2}$. Thm. 6 immediately follows from Lems. 5 and 6. □

We next consider the identification of other decompositions of $\mathrm{Cov}(X, Y)$ in the model of Fig. 1(a). Indeed, one could show that the decomposition of $\mathrm{Cov}(X, Y)$ (Thm. 5) are always identifiable in the standard model regardless of the order $\mathcal{L}^c$ and $\mathcal{L}^s$.

Lemma 7. *In the standard model (Fig. 1(a)), for an order $\mathcal{L}^c : l_2 < l_1$, the path-specific decomposition of the causal covariance $\mathrm{Cov}^c_{x^*}(X, Y)$ (Thm. 2) are identifiable if $P(x, y_{x^*})$ and $P(x, y_{W_{x^*}})$ are identifiable. Specifically, distributions $P(x, y_{x^*})$ and $P(x, y_{W_{x^*}})$ can be estimated from the observational distribution $P(x, y, z, w)$ as following:*

$$\begin{aligned} &P(x, y_{x^*}) = \sum_{z,w} P(y|x^*, z, w) P(w|x^* *, z) P(x, z), \\ &P(x, y_{W_{x^*}}) = \sum_{z,w} P(y|x, z, w) P(w|x^*, z) P(x, z). \end{aligned}$$

Proof. Consider the order $\mathcal{L}^c : l_2 < l_1$. The path-specific causal covariance of l_1, l_2 are equal to:

$$\begin{aligned} &\mathrm{Cov}^c_{l_1[x^*]}(X, Y)_{\mathcal{L}^c_{2\pi}} = \mathrm{Cov}^{dt}_{x^*}(X, Y) \\ &= \mathrm{Cov}(X, Y_{W_{x^*}} - Y_{x^*}) \\ &= \mathrm{Cov}(X, Y_{W_{x^*}}) - \mathrm{Cov}(X, Y_{x^*}), \end{aligned} \tag{31}$$

$$\begin{aligned} &\mathrm{Cov}^c_{l_2[x^*]}(X, Y)_{\mathcal{L}^c_{2\pi}} = \mathrm{Cov}^{ip}_{x^*}(X, Y) \\ &= \mathrm{Cov}(X, Y - Y_{W_{x^*}}) \\ &= \mathrm{Cov}(X, Y) - \mathrm{Cov}(X, Y_{W_{x^*}}). \end{aligned} \tag{32}$$

It suffices to show that distributions $P(x, y_{W_{x^*}})$ and $P(x, y_{x^*})$ are identifiable. $P(x, y_{x^*})$ can be identified using Lem. 5. By expanding on Z, W_{x^*}, $P(x, y_{W_{x^*}})$ can be written as:

$$\begin{aligned} &P(x, y_{W_{x^*}}) \\ &= \sum_{z,w} P(y_w|x, z, w_{x^*}) P(w_{x^*}|x, z) P(x, z) \\ &= \sum_{z,w} P(y_w|x_w, z_w, w_{x^*}) P(w_{x^*}|x, z_{x^*}) P(x, z). \end{aligned}$$

In the last step, since Z is a non-descendant node of X, W and X is a non-descendant node of W, we have $Z = Z_{x^*} = Z_w$ and $X = X_w$. By the composition axiom,

$$\begin{aligned} &Z = z \Rightarrow X = X_z, \\ &Z_{x^*} = z \Rightarrow W_{x^*} = W_{x^*,z}, \\ &X_w = x, Z_w = z \Rightarrow Y_w = Y_{x,z,w}. \end{aligned}$$

which gives:

$$
\begin{aligned}
&\sum_{z,w} P(y_w|x_w, z_w, w_{x^*})P(w_{x^*}|x, z_{x^*})P(x,z)\\
&=\sum_{z,w} P(y_{x,z,w}|x_w, z_w, w_{x^*})P(w_{x^*,z}|x, z_{x^*})P(x,z)\\
&=\sum_{z,w} P(y_{x,z,w}|x, z, w_{x^*})P(w_{x^*,z}|x, z)P(x,z)\\
&=\sum_{z,w} P(y_{x,z,w}|x_z, z, w_{x^*})P(w_{x^*,z}|x_z, z)P(x,z)\\
&=\sum_{z,w} P(y_{x,z,w}|x_z, z, w_{x^*,z})P(w_{x^*,z}|x_z, z)P(x,z)
\end{aligned}
$$

The last step holds since $Z = Z_{x^*}$ and $Z_{x^*} = z \Rightarrow W_{x^*} = W_{x^*,z}$. Applying Eqs. 20 and 21 gives:

$$
\begin{aligned}
P(x, y_{W_{x^*}}) &= \sum_{z,w} P(y_{x,z,w})P(w_{x^*,z})P(x,z)\\
&= \sum_{z,w} P(y|x,z,w)P(w|x^*,z)P(x,z).
\end{aligned}
$$

The last step holds by Eqs. 22 and 23. □

Lemma 8. *In the standard model (Fig. 1(a)), for a order $\mathcal{L}^s$ where $\mathcal{L}^s_r : g_{r_2} < g_{r_1}$, the path-specific decomposition of the spurious covariance $\mathrm{Cov}^s_{x^*}(X,Y)$ (Thm. 5) is identifiable if $P(x, y_{x^*}), P(x, y_{x^*,W_{x^*,U^r_Z}})$ are identifiable. Specifically, distributions $P(x, y_{x^*}), P(x, y_{x^*,W_{x^*,U^r_Z}})$ are identifiable can be estimated from the observational distribution $P(x,y,z,w)$ as following:*

$$
P(x, y_{x^*}) = \sum_{z,w} P(y|x^*, z, w)P(w|x*, z)P(x,z),
$$

$$
\begin{aligned}
P(x, y_{x^*,W_{x^*,U^r_Z}}) = \sum_{z,z',w} & P(y|x^*, w, z')P(w|x^*, z)\\
&\cdot P(x,z')P(z).
\end{aligned}
$$

Proof. Considering the order $\mathcal{L}^s$ where $\mathcal{L}^s_r : g_{r_2} < g_{r_1}$,

$$
\begin{aligned}
&\mathrm{Cov}^s_{l_3[x^*]}(X,Y)_{\mathcal{L}^s_{1\pi}} = \mathrm{Cov}(X - X_{g_l}, Y_{x^*,g_{r_2}} - Y_{x^*,g_{r_{[1,2]}}})\\
&= \mathrm{Cov}(X - X_{U^l_Z}, Y_{x^*,W_{x^*,U^r_Z}} - Y_{x^*,U^r_Z})\\
&= \mathrm{Cov}(X, Y_{x^*,W_{x^*,U^r_Z}} - Y_{x^*,U^r_Z}) \qquad (33)\\
&= \mathrm{Cov}(X, Y_{x^*,W_{x^*,U^r_Z}}).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\mathrm{Cov}^s_{l_4[x^*]}(X,Y)_{\mathcal{L}^s_{1\pi}} = \mathrm{Cov}(X - X_{g_l}, Y_{x^*} - Y_{x^*,g_{r_2}})\\
&= \mathrm{Cov}(X - X_{U^l_Z}, Y_{x^*} - Y_{x^*,W_{x^*,U^r_Z}})\\
&= \mathrm{Cov}(X, Y_{x^*} - Y_{x^*,W_{x^*,U^r_Z}}) \qquad (34)\\
&= \mathrm{Cov}(X, Y_{x^*}) - \mathrm{Cov}(X, Y_{x^*,W_{x^*,U^r_Z}}).
\end{aligned}
$$

The last step holds since $X_{U^l_Z}$ and Y_{x^*,U^r_Z} are independent counterfactual variables. It will suffice to show that the distributions $P(x, y_{x^*}), P(x, y_{x^*,W_{x^*,U^r_Z}})$ are identifiable. By conditioning on U^r_Z,

$$
P(x, y_{x^*,W_{x^*,U^r_Z}}) = \sum_{u^r_Z} P(x, y_{x^*,W_{x^*,u^r_Z}}|u^r_Z)P(u^r_Z)
$$

With U^r_Z fixed, variables X and $Y_{x^*,W_{x^*,u^r_Z}}$ are functions of the exogenous variable U, which is independent of U^r_Z. We thus have the independence relation

$$
U^r_Z \perp\!\!\!\perp X, Y_{x^*,W_{x^*,u^r_Z}},
$$

which gives:

$$
P(x, y_{x^*,W_{x^*,U^r_Z}}) = \sum_{u^r_Z} P(x, y_{x^*,W_{x^*,u^r_Z}})P(u^r_Z) \qquad (35)
$$

By expanding on $Z, Z_{u^r_Z}, W_{x^*,u^r_Z}$,

$$
P(x, y_{x^*,W_{x^*,u^r_Z}}) = \sum_{z,z',w} P(x, y_{x^*,w}, w_{x^*,u^r_Z}, z', z_{u^r_Z})
$$

By the composition axiom and the exclusion restrictions rule [Pearl, 2000, Ch. 7.3] (treating U_Z as an endogenous variable), for any z, x^*, w, u_Z,

$$
\begin{aligned}
&Z_{u_Z} = Z_{x^*,u_Z},\\
&Z_{x^*,u_Z} = z \Rightarrow W_{x^*,u_Z} = W_{x^*,z,u_Z} = W_{x^*,z},\\
&Z = Z_{x^*,w},\\
&Z_{x^*,w} = z \Rightarrow Y_{x^*,w} = Y_{x^*,w,z}.
\end{aligned} \qquad (36)
$$

Eqs. 28 and 36 imply

$$
\begin{aligned}
&P(x, y_{x^*,w}, w_{x^*,u^r_Z}, z', z_{u^r_Z})\\
&= P(x_{z'}, y_{x^*,w,z'}, w_{x^*,z}, z', z_{u^r_Z})
\end{aligned}
$$

Since the function f_Z takes only U_Z as an argument. The variables $Z_{u^r_Z}$ are thus deterministic, i.e.,

$$
\begin{aligned}
&P(x_{z'}, y_{x^*,w,z'}, w_{x^*,z}, z', z_{u^r_Z})\\
&= P(x_{z'}, y_{x^*,w,z'}, w_{x^*,z}, z')I\{Z_{u^r_Z} = z\}.
\end{aligned}
$$

The above equation, together with Eq.35, gives

$$
\begin{aligned}
&P(x, y_{x^*,W_{x^*,U^r_Z}})\\
&= \sum_{z,z',w} P(x_{z'}, y_{x^*,w,z'}, w_{x^*,z}, z') \sum_{u^r_Z} I\{Z_{u^r_Z} = z\}P(u^r_Z)\\
&= \sum_{z,z',w} P(x_{z'}, y_{x^*,w,z'}, w_{x^*,z}, z')P(z)
\end{aligned}
$$

The independence restrictions rule [Pearl, 2000, Ch. 7.3] implies that in the model of Fig. 1(a), counterfactuals

$X_{z'}, Y_{x^*,w,z'}, W_{x^*,z}, Z$ are mutually independent. Together with Eq. 30, the above equation is equal to:

$$
\begin{aligned}
&P(x, y_{x^*,W_{x^*,U_Z^r}}) \\
&= \sum_{z,z',w} P(x_{z'})P(y_{x^*,w,z'})P(w_{x^*,z})P(z')P(z) \\
&= \sum_{z,z',w} P(x|z')P(y|x^*,w,z')P(w|x^*,z)P(z')P(z) \\
&= \sum_{z,z',w} P(y|x^*,w,z')P(w|x^*,z)P(x,z')P(z). \qquad \square
\end{aligned}
$$

Since Lems. 5-8 cover all possible orders $\mathcal{L}^c$, $\mathcal{L}^s$, the decompositions Thm. 5 are always identifiable in the standard fairness model.

Proof of Theorem 7. By Eqs. 18, 19, 31 and 32, we have for order $\mathcal{L}_1^c : l_1 < l_2$ and $\mathcal{L}_2^c : l_2 < l_1$,

$$
\begin{aligned}
\mathrm{Cov}^c_{l_1[x^*]}(X,Y)_{\mathcal{L}^c_{1_\pi}} &= \mathrm{Cov}^{dp}_{x^*}(X,Y), \\
\mathrm{Cov}^c_{l_1[x^*]}(X,Y)_{\mathcal{L}^c_{2_\pi}} &= \mathrm{Cov}^{dt}_{x^*}(X,Y), \\
\mathrm{Cov}^c_{l_2[x^*]}(X,Y)_{\mathcal{L}^c_{1_\pi}} &= \mathrm{Cov}^{it}_{x^*}(X,Y), \\
\mathrm{Cov}^c_{l_2[x^*]}(X,Y)_{\mathcal{L}^c_{2_\pi}} &= \mathrm{Cov}^{ip}_{x^*}(X,Y).
\end{aligned}
$$

Applying Thm. 1 to the above equations implies that for an arbitrary order $\mathcal{L}^c$ over $l_{[1,2]}$,

$$
\mathrm{Cov}^c_{l_1[x^*]}(X,Y)_{\mathcal{L}^c_\pi} = \alpha_{YX}, \mathrm{Cov}^c_{l_2[x^*]}(X,Y)_{\mathcal{L}^c_\pi} = \alpha_{WX}\alpha_{YW}.
$$

We will next consider the path-specific spurious covariance of l_3, l_4. As for $\mathcal{L}^s$ where $\mathcal{L}^s_r : g_{r_1} < g_{r_2}$, by Eqs. 24-25,

$$
\begin{aligned}
&\mathrm{Cov}^s_{l_3[x^*]}(X,Y)_{\mathcal{L}^s_{1_\pi}} \\
&= \mathrm{Cov}(X, Y_{x^*} - Y_{x^*,W_{x^*},Z_{U_Z^r}}) \\
&= \alpha_{YZ}\mathrm{Cov}(X,Z) + \alpha_{YZ}\mathrm{Cov}(X, Z_{U_Z^r}) \qquad (37)
\end{aligned}
$$

Since the domain is normalized, $\mathrm{Cov}(Z,Z) = \mathrm{Var}(Z) = 1$, $\mathrm{Cov}(X,Z)$ is equal to

$$
\begin{aligned}
&\mathrm{Cov}(X,Z) = \mathrm{Cov}(\alpha_{XZ}Z + U_X, Z) \\
&= \alpha_{XZ}\mathrm{Var}(Z) = \alpha_{XZ} \qquad (38)
\end{aligned}
$$

$\mathrm{Cov}(X, Z_{U_Z^r})$ equates to:

$$
\begin{aligned}
&\mathrm{Cov}(X, Z_{U_Z^r}) = \mathrm{Cov}(\alpha_{XZ}Z + U_X, Z_{U_Z^r}) \\
&= \mathrm{Cov}(\alpha_{XZ}U_Z + U_X, U_Z^r) \\
&= \alpha_{XZ}\mathrm{Cov}(U_Z, U_Z^r) + \mathrm{Cov}(U_X, U_Z^r) = 0. \qquad (39)
\end{aligned}
$$

The last step holds since U_Z, U_X and U_Z^r are mutually independent. Eqs. 37-39 together give:

$$
\mathrm{Cov}^s_{l_3[x^*]}(X,Y)_{\mathcal{L}^s_{1_\pi}} = \alpha_{XZ}\alpha_{YZ}. \qquad (40)
$$

Similarly,

$$
\begin{aligned}
&\mathrm{Cov}^s_{l_4[x^*]}(X,Y)_{\mathcal{L}^s_{1_\pi}} \\
&= \mathrm{Cov}(X, Y_{x^*,W_{x^*},Z_{U_Z^r}} - Y_{x^*,U_Z^r}) \\
&= \alpha_{YW}\mathrm{Cov}(X, W_{x^*}) + \alpha_{YZ}\mathrm{Cov}(X, W_{x^*,U_Z^r}) \\
&= \alpha_{WZ}\alpha_{YW}(\mathrm{Cov}(X,Z) + \mathrm{Cov}(X, Z_{U_Z^r})) \\
&= \alpha_{XZ}\alpha_{WZ}\alpha_{YW}. \qquad (41)
\end{aligned}
$$

We will next consider the order $\mathcal{L}^s$ where $\mathcal{L}^s_r : g_{r_2} < g_{r_1}$. By Eqs. 33-34,

$$
\begin{aligned}
&\mathrm{Cov}^s_{l_3[x^*]}(X,Y)_{\mathcal{L}^s_{1_\pi}} \\
&= \mathrm{Cov}(X, Y_{x^*,W_{x^*,U_Z^r}} - Y_{x^*,U_Z^r}) \\
&= \alpha_{YZ}\mathrm{Cov}(X,Z) + \alpha_{YZ}\mathrm{Cov}(X, Z_{U_Z^r}) \\
&= \alpha_{XZ}\alpha_{YZ} \qquad (42)
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{Cov}^s_{l_4[x^*]}(X,Y)_{\mathcal{L}^s_{1_\pi}} \\
&= \mathrm{Cov}(X, Y_{x^*} - Y_{x^*,W_{x^*,U_Z^r}}) \\
&= \alpha_{YW}\mathrm{Cov}(X, W_{x^*}) + \alpha_{YZ}\mathrm{Cov}(X, W_{x^*,U_Z^r}) \\
&= \alpha_{XZ}\alpha_{WZ}\alpha_{YW} \qquad (43)
\end{aligned}
$$

Eqs. 40 - 43 combined imply that for an arbitrary order $\mathcal{L}^s$,

$$
\begin{aligned}
\mathrm{Cov}^s_{l_3[x^*]}(X,Y)_{\mathcal{L}^s_\pi} &= \alpha_{XZ}\alpha_{YZ}, \\
\mathrm{Cov}^s_{l_4[x^*]}(X,Y)_{\mathcal{L}^s_\pi} &= \alpha_{XZ}\alpha_{WZ}\alpha_{YW}.
\end{aligned}
$$

Specifically, Parameters α can be estimated from the partial regression coefficients [Pearl, 2000, Ch. 5] as following:

$$
\begin{aligned}
&\alpha_{YX} = \gamma_{YX.ZW}, \quad \alpha_{YZ} = \gamma_{YZ.XW}, \quad \alpha_{YW} = \gamma_{YW.XZ}, \\
&\alpha_{WX} = \gamma_{WX.Z}, \quad \alpha_{WZ} = \gamma_{WZ.X}, \quad \alpha_{XZ} = \alpha_{XZ}. \qquad \square
\end{aligned}
$$

2 EXAMPLES

In this section, we will illustrate the results presented in this paper with more detailed examples.

2.1 PATH-SPECIFIC POTENTIAL RESPONSE

Consider the standard model of Fig. 1(a). Recall the path $g_1 : X \to W_1 \to W_2 \to Y$. We next show, step by step, the derivation of the g_1-specific potential response $Y_{g_1[x^*]}$. Since the edge $X \to Y \notin \{g_1\}$, the set $X_{\pi\to Y} = \emptyset$. We thus have $S = (Pa(Y)_G \cap V) - X_{\pi\to Y} = \{X, W_1, W_2\}$. By Def. 6,

$$
\begin{aligned}
&Y_{g_1[x^*]} \\
&= Y_{X_{\lhd_{X\to Y}(g_1)[x^*]}, W_{1_{\lhd_{W_1\to Y}(g_1)[x^*]}}, W_{2_{\lhd_{W_2\to Y}(g_1)[x^*]}}}.
\end{aligned}
$$

Since the edges $X \to Y$ and $W_1 \to Y$ are not subpaths of g_1,

$$\lhd_{X\to Y}(g_1) = \lhd_{W_1\to Y}(g_1) = \emptyset.$$

By Def. 6, the above equation implies

$$X_{\lhd_{X\to Y}(g_1)[x^*]} = X, \quad W_{1_{\lhd_{W_1\to Y}(g_1)[x^*]}} = W_1.$$

$Y_{g_1[x^*]}$ can thus be written as:

$$Y_{g_1[x^*]} = Y_{X,W_1,W_{2_{\lhd_{W_2\to Y}(g_1)[x^*]}}}.$$

Since $\lhd_{W_2\to Y}(g_1)$ returns the subpath $\{g_1(X,W_2)\}$,

$$Y_{g_1[x^*]} = Y_{X,W_1,W_{2_{g_1(X,W_2)[x^*]}}}, \tag{44}$$

where $W_{2_{g_1(X,W_2)[x^*]}}$ is the path-specific potential response of W_2. Since $X \to Y \notin \{g_1(X,W_2)\}$, the set S for $W_{2_{g_1(X,W_2)[x^*]}}$ is $\{X, W_1\}$. Applying Def. 6 again,

$$\begin{aligned} &W_{2_{g_1(X,W_2)[x^*]}} \\ &= W_{2_{X_{\lhd_{X\to W_2}(g_1(X,W_2))[x^*]}, W_{1_{\lhd_{W_1\to W_2}(g_1(X,W_2))[x^*]}}}}. \end{aligned}$$

Since

$$\begin{aligned} &\lhd_{X\to W_2}(g_1(X,W_2)) = \emptyset, \\ &\lhd_{W_1\to W_2}(g_1(X,W_2)) = \{g_1(X,W_1)\}, \end{aligned}$$

$W_{2_{g_1(X,W_2)}}$ can be written as:

$$W_{2_{g_1(X,W_2)[x^*]}} = W_{2_{X,W_{1_{g_1(X,W_1)[x^*]}}}}, \tag{45}$$

where $W_{1_{g_1(X,W_1)[x^*]}}$ is the path-specific potential response of W_1. Since the edge $X \to W_1 = g_1(X,W_1)$, the set $X_{\pi\to W_1} = \{X\}$ and $S = (Pa(W_1)_G \cap V) - X_{\pi\to W_1} = \emptyset$. By Def. 6,

$$W_{1_{g_1(X,W_1)[x^*]}} = W_{1_{x^*}}. \tag{46}$$

Eqs. 44-46 together give:

$$Y_{g_1[x^*]} = Y_{X,W_1,W_{2_{X,W_{1_{x^*}}}}} = Y_{W_{2_{W_{1_{x^*}}}}}.$$

2.2 DECOMPOSING CAUSAL RELATIONS

We will consider the model in Fig. 6 where causal effects from X and Y are mediated by W_1, W_2, W_3, and all directed edges are confounded. There are eight causal paths from X to Y:

$$\begin{aligned} &g_1 : X \to Y, \\ &g_2 : X \to W_1 \to Y, \\ &g_3 : X \to W_2 \to Y, \\ &g_4 : X \to W_3 \to Y, \\ &g_5 : X \to W_1 \to W_2 \to Y, \\ &g_6 : X \to W_1 \to W_3 \to Y, \\ &g_7 : X \to W_2 \to W_3 \to Y, \\ &g_8 : X \to W_1 \to W_2 \to W_3 \to Y. \end{aligned}$$

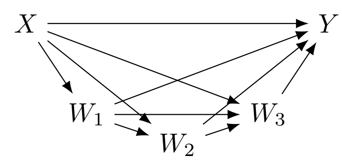


Figure 6: Causal diagram for the three-mediators setting where causal paths from X and Y are mediated by W_1, W_2, W_3.

Let an order $\mathcal{L}^c$ be $g_i < g_j$ if $i < j$. Thm. 2 is applicable and express the causal covariance $\mathrm{Cov}^c_{x^*}(X,Y)$ as:

$$\mathrm{Cov}^c_{x^*}(X,Y) = \sum_{i=1}^{8} \mathrm{Cov}^c_{g_i[x^*]}(X,Y)_{\mathcal{L}^c_\pi}.$$

The path-specific causal covariance $\{\mathrm{Cov}^c_{g_i[x^*]}(X,Y)_{\mathcal{L}^c_\pi}\}_{i=1,\dots,8}$ are equal to:

$$\begin{aligned}
&\mathrm{Cov}^c_{g_1[x^*]}(X,Y)_{\mathcal{L}^c_\pi} = \mathrm{Cov}(X, Y - Y_{g_1[x^*]}) \\
&= \mathrm{Cov}(X, Y - Y_{x^*,W_1,W_2,W_3}), \\
&\mathrm{Cov}^c_{g_2[x^*]}(X,Y)_{\mathcal{L}^c_\pi} = \mathrm{Cov}(X, Y_{g_1[x^*]} - Y_{g_{[1,2]}[x^*]}) \\
&= \mathrm{Cov}(X, Y_{x^*,W_1,W_2,W_3} - Y_{x^*,W_{1_{x^*}},W_2,W_3}), \\
&\mathrm{Cov}^c_{g_3[x^*]}(X,Y)_{\mathcal{L}^c_\pi} = \mathrm{Cov}(X, Y_{g_{[1,2]}[x^*]} - Y_{g_{[1,3]}[x^*]}) \\
&= \mathrm{Cov}(X, Y_{x^*,W_{1_{x^*}},W_2,W_3} - Y_{x^*,W_{1_{x^*}},W_{2_{x^*,W_1}},W_3}), \\
&\mathrm{Cov}^c_{g_4[x^*]}(X,Y)_{\mathcal{L}^c_\pi} = \mathrm{Cov}(X, Y_{g_{[1,3]}[x^*]} - Y_{g_{[1,4]}[x^*]}) \\
&= \mathrm{Cov}(X, Y_{x^*,W_{1_{x^*}},W_{2_{x^*,W_1}},W_3} \\
&\qquad\qquad - Y_{x^*,W_{1_{x^*}},W_{2_{x^*,W_1}},W_{3_{x^*,W_1,W_2}}}), \\
&\mathrm{Cov}^c_{g_5[x^*]}(X,Y)_{\mathcal{L}^c_\pi} = \mathrm{Cov}(X, Y_{g_{[1,4]}[x^*]} - Y_{g_{[1,5]}[x^*]}) \\
&= \mathrm{Cov}(X, Y_{x^*,W_{1_{x^*}},W_{2_{x^*,W_1}},W_{3_{x^*,W_1,W_2}}} \\
&\qquad\qquad - Y_{x^*,W_{1_{x^*}},W_{2_{x^*}},W_{3_{x^*,W_1,W_2}}}), \\
&\mathrm{Cov}^c_{g_6[x^*]}(X,Y)_{\mathcal{L}^c_\pi} = \mathrm{Cov}(X, Y_{g_{[1,5]}[x^*]} - Y_{g_{[1,6]}[x^*]}) \\
&= \mathrm{Cov}(X, Y_{x^*,W_{1_{x^*}},W_{2_{x^*}},W_{3_{x^*,W_1,W_2}}} \\
&\qquad\qquad - Y_{x^*,W_{1_{x^*}},W_{2_{x^*}},W_{3_{x^*,W_{1_{x^*}},W_2}}}), \\
&\mathrm{Cov}^c_{g_7[x^*]}(X,Y)_{\mathcal{L}^c_\pi} = \mathrm{Cov}(X, Y_{g_{[1,6]}[x^*]} - Y_{g_{[1,7]}[x^*]}) \\
&= \mathrm{Cov}(X, Y_{x^*,W_{1_{x^*}},W_{2_{x^*}},W_{3_{x^*,W_{1_{x^*}},W_2}}} \\
&\qquad\qquad - Y_{x^*,W_{1_{x^*}},W_{2_{x^*}},W_{3_{x^*,W_{1_{x^*}},W_{2_{x^*,W_1}}}}}), \\
&\mathrm{Cov}^c_{g_8[x^*]}(X,Y)_{\mathcal{L}^c_\pi} = \mathrm{Cov}(X, Y_{g_{[1,7]}[x^*]} - Y_{g_{[1,8]}[x^*]}) \\
&= \mathrm{Cov}(X, Y_{x^*,W_{1_{x^*}},W_{2_{x^*}},W_{3_{x^*,W_{1_{x^*}},W_{2_{x^*,W_1}}}}} - Y_{x^*}).
\end{aligned}$$

2.3 DECOMPOSING SPURIOUS RELATIONS

We will consider the generalized two-confounders setting described in Fig. 7(a) where X and Y are confounded by Z_1, Z_2. The exogenous variables U_1, U_2

associated with Z_1, Z_2 are represented explicitly in the causal diagram. In the model of Fig. 7(a), $U^s = \{U_1, U_2\}$ which affects the observational X and the counterfactuals Y_{x^*} through causal paths shown in Fig. 7(b). There are thus five spurious treks:

$$\tau_1 = (g^1_{l_1}, g^1_{r_1}), \quad \tau_2 = (g^1_{l_1}, g^1_{r_2}), \quad \tau_3 = (g^1_{l_2}, g^1_{r_1})$$
$$\tau_4 = (g^1_{l_2}, g^1_{r_2}), \quad \tau_5 = (g^2_{l_1}, g^2_{r_1}).$$

The treatment X and the outcome Y are connected through four spurious paths:

$$l_1 : X \leftarrow Z_1 \rightarrow Y, \qquad l_2 : X \leftarrow Z_2 \rightarrow Y,$$
$$l_3 : X \leftarrow Z_1 \rightarrow Z_2 \rightarrow Y, \quad l_4 : X \leftarrow Z_2 \leftarrow Z_1 \rightarrow Y.$$

Let an order $\mathcal{L}^s_u$ be $U_1 < U_2$. Let an order $L^s_{r_i}$ be $g^i_{r_j} < g^i_{r_k}$ if $j < k$. The order L_{l_i} is similarly defined. Thm. 3 decomposes the spurious covariance $\mathrm{Cov}^s_{x^*}(X, Y)$ over the spurious paths $l_1, \dots, l_4$:

$$\mathrm{Cov}^s_{x^*}(X, Y) = \sum_{i=1}^{4} \mathrm{Cov}^s_{l_i[x^*]}(X, Y)_{\mathcal{L}^s_\pi}.$$

The path-specific spurious covariance $\left\{\mathrm{Cov}^s_{l_i[x^*]}(X, Y)_{\mathcal{L}^s_\pi}\right\}_{i=1,\dots,4}$ are equal to:

$$\begin{aligned}
&\mathrm{Cov}^s_{l_1[x^*]}(X, Y)_{\mathcal{L}^s_\pi} = \mathrm{Cov}^{ts}_{\tau_1[x^*]}(X, Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}(X - X_{g^1_{l_1}}, Y_{x^*} - Y_{x^*, g^1_{r_1}}) \\
&= \mathrm{Cov}(X - X_{{Z_1}_{U_1^l}, Z_2}, Y_{x^*} - Y_{x^*, {Z_1}_{U_1^r}, Z_2}),
\end{aligned}$$

$$\begin{aligned}
&\mathrm{Cov}^s_{l_2[x^*]}(X, Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}^{ts}_{\tau_4[x^*]}(X, Y)_{\mathcal{L}^s_\pi} + \mathrm{Cov}^{ts}_{\tau_5[x^*]}(X, Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}(X_{g^1_{l_1}} - X_{g^1_{l_{[1,2]}}}, Y_{x^*, g^1_{r_1}} - Y_{x^*, g^1_{r_{[1,2]}}}) \\
&+ \mathrm{Cov}(X_{g^1_{l_{[1,2]}} \cup g^2_{l_1}}, Y_{x^*, g^1_{r_{[1,2]}}} - Y_{x^*, g^1_{r_{[1,2]}} \cup g^2_{r_1}}) \\
&= \mathrm{Cov}(X_{{Z_1}_{U_1^l}, Z_2} - X_{U_1^l}, Y_{x^*, {Z_1}_{U_1^r}, Z_2} - Y_{x^*, U_1^r}) \\
&+ \mathrm{Cov}(X_{U_1^l} - X_{U^l_{[1,2]}}, Y_{x^*, U_1^r} - Y_{x^*, U^r_{[1,2]}}), \\
&\mathrm{Cov}^s_{l_3[x^*]}(X, Y)_{\mathcal{L}^s_\pi} = \mathrm{Cov}^{ts}_{\tau_2[x^*]}(X, Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}(X - X_{g^1_{l_1}}, Y_{x^*, g^1_{r_1}} - Y_{x^*, g^1_{r_{[1,2]}}}) \\
&= \mathrm{Cov}(X - X_{{Z_1}_{U_1^l}, Z_2}, Y_{x^*, {Z_1}_{U_1^r}, Z_2} - Y_{x^*, U_1^r}), \\
&\mathrm{Cov}^s_{l_4[x^*]}(X, Y)_{\mathcal{L}^s_\pi} = \mathrm{Cov}^{ts}_{\tau_3[x^*]}(X, Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}(X_{g^1_{l_1}} - X_{g^1_{l_{[1,2]}}}, Y_{x^*} - Y_{x^*, g^1_{r_1}}) \\
&= \mathrm{Cov}(X_{{Z_1}_{U_1^l}, Z_2} - X_{U_1^l}, Y_{x^*} - Y_{x^*, {Z_1}_{U_1^r}, Z_2}).
\end{aligned}$$

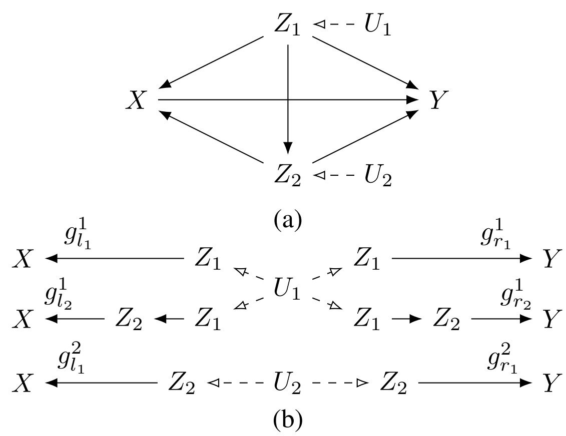


Figure 7: (a) Causal diagram for the two-confounders setting where X to Y are confounded by Z_1, Z_2; (b) Causal paths through which the exogenous variables U_1, U_2 affect X and Y_{x^*} in the two-confounders setting.

2.4 PATH-SPECIFIC DECOMPOSITION

Considering the model of Fig. 8, the treatment X and the outcome Y are connected by the causal paths:

$$l_1 : X \rightarrow Y, \qquad l_2 : X \rightarrow W_1 \rightarrow Y,$$
$$l_3 : X \rightarrow W_2 \rightarrow Y, \quad l_4 : X \rightarrow W_1 \rightarrow W_2 \rightarrow Y,$$

and the spurious paths:

$$\begin{aligned}
&l_5 : X \leftarrow Z_1 \rightarrow Y, \\
&l_6 : X \leftarrow Z_1 \rightarrow W_1 \rightarrow Y, \\
&l_7 : X \leftarrow Z_1 \rightarrow W_2 \rightarrow Y, \\
&l_8 : X \leftarrow Z_1 \rightarrow W_1 \rightarrow W_2 \rightarrow Y, \\
&l_9 : X \leftarrow Z_1 \rightarrow Z_2 \rightarrow Y, \\
&l_{10} : X \leftarrow Z_1 \rightarrow Z_2 \rightarrow W_1 \rightarrow Y, \\
&l_{11} : X \leftarrow Z_1 \rightarrow Z_2 \rightarrow W_2 \rightarrow Y, \\
&l_{12} : X \leftarrow Z_1 \rightarrow Z_2 \rightarrow W_1 \rightarrow W_2 \rightarrow Y, \\
&l_{13} : X \leftarrow Z_2 \leftarrow Z_1 \rightarrow Y, \\
&l_{14} : X \leftarrow Z_2 \leftarrow Z_1 \rightarrow W_1 \rightarrow Y, \\
&l_{15} : X \leftarrow Z_2 \leftarrow Z_1 \rightarrow W_2 \rightarrow Y, \\
&l_{16} : X \leftarrow Z_2 \leftarrow Z_1 \rightarrow W_1 \rightarrow W_2 \rightarrow Y, \\
&l_{17} : X \leftarrow Z_2 \rightarrow Y, \\
&l_{18} : X \leftarrow Z_2 \rightarrow W_1 \rightarrow Y, \\
&l_{19} : X \leftarrow Z_2 \rightarrow W_2 \rightarrow Y, \\
&l_{20} : X \leftarrow Z_2 \rightarrow W_1 \rightarrow W_2 \rightarrow Y.
\end{aligned}$$

Let U_1, U_2 denote the independent errors associated with the confounders Z_1, Z_2 respectively. In this model, $U^s = \{U_1, U_2\}$ where the causal paths $\Pi^c(U_1, X)$ and

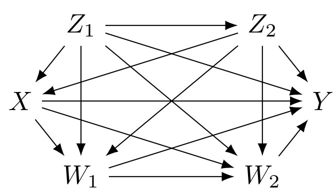


Figure 8: Causal diagram for the two-mediators-two-confounders setting where X to Y are confounded by Z_1, Z_2 and mediated by W_1, W_2.

$\Pi^c(U_1, X|Y)$ are:

$$\begin{aligned}
g^1_{l_1} &: U_1 \to Z_1 \to X,\\
g^1_{l_2} &: U_1 \to Z_1 \to Z_2 \to X,\\
g^1_{r_1} &: U_1 \to Z_1 \to Y,\\
g^1_{r_2} &: U_1 \to Z_1 \to W_1 \to Y,
\end{aligned}$$

$$\begin{aligned}
g^1_{r_3} &: U_1 \to Z_1 \to W_2 \to Y,\\
g^1_{r_4} &: U_1 \to Z_1 \to W_1 \to W_2 \to Y,\\
g^1_{r_5} &: U_1 \to Z_1 \to Z_2 \to Y,\\
g^1_{r_6} &: U_1 \to Z_1 \to Z_2 \to W_1 \to Y,\\
g^1_{r_7} &: U_1 \to Z_1 \to Z_2 \to W_2 \to Y,\\
g^1_{r_8} &: U_1 \to Z_1 \to Z_2 \to W_1 \to W_2 \to Y,
\end{aligned}$$

and the causal paths $\Pi^c(U_2, X)$ and $\Pi^c(U_2, X|Y)$ are:

$$\begin{aligned}
g^2_{l_1} &: U_2 \to Z_2 \to X,\\
g^2_{r_1} &: U_1 \to Z_2 \to Y,\\
g^2_{r_2} &: U_1 \to Z_2 \to W_1 \to Y,\\
g^2_{r_3} &: U_1 \to Z_2 \to W_2 \to Y,\\
g^2_{r_4} &: U_1 \to Z_2 \to W_1 \to W_2 \to Y.
\end{aligned}$$

There are thus twenty spurious treks from X to Y:

$$\begin{aligned}
&\tau_1 = (g^1_{l_1}, g^1_{r_1}), \quad \tau_2 = (g^1_{l_1}, g^1_{r_2}), \quad \tau_3 = (g^1_{l_1}, g^1_{r_3}),\\
&\tau_4 = (g^1_{l_1}, g^1_{r_4}), \quad \tau_5 = (g^1_{l_1}, g^1_{r_5}), \quad \tau_6 = (g^1_{l_1}, g^1_{r_6}),\\
&\tau_7 = (g^1_{l_1}, g^1_{r_7}), \quad \tau_8 = (g^1_{l_1}, g^1_{r_8}), \quad \tau_9 = (g^1_{l_2}, g^1_{r_1}),\\
&\tau_{10} = (g^1_{l_2}, g^1_{r_2}), \quad \tau_{11} = (g^1_{l_2}, g^1_{r_3}), \quad \tau_{12} = (g^1_{l_2}, g^1_{r_4}),\\
&\tau_{13} = (g^1_{l_2}, g^1_{r_5}), \quad \tau_{14} = (g^1_{l_2}, g^1_{r_6}), \quad \tau_{15} = (g^1_{l_2}, g^1_{r_7}),\\
&\tau_{16} = (g^1_{l_2}, g^1_{r_8}), \quad \tau_{17} = (g^2_{l_1}, g^2_{r_1}), \quad \tau_{18} = (g^2_{l_1}, g^2_{r_2}),\\
&\tau_{19} = (g^2_{l_1}, g^2_{r_3}), \quad \tau_{20} = (g^2_{l_1}, g^2_{r_4}).
\end{aligned}$$

The set $\{\mathcal{T}^s(l_i)\}_{i=5,\dots,20}$ is a partition over the spurious treks $\tau_{[1,20]}$:

$$\begin{aligned}
&\mathcal{T}^s(l_5) = \{\tau_1\}, && \mathcal{T}^s(l_6) = \{\tau_2\},\\
&\mathcal{T}^s(l_7) = \{\tau_3\}, && \mathcal{T}^s(l_8) = \{\tau_4\},\\
&\mathcal{T}^s(l_9) = \{\tau_5\}, && \mathcal{T}^s(l_{10}) = \{\tau_6\},\\
&\mathcal{T}^s(l_{11}) = \{\tau_7\}, && \mathcal{T}^s(l_{12}) = \{\tau_8\},\\
&\mathcal{T}^s(l_{13}) = \{\tau_9\}, && \mathcal{T}^s(l_{14}) = \{\tau_{10}\},\\
&\mathcal{T}^s(l_{15}) = \{\tau_{11}\}, && \mathcal{T}^s(l_{16}) = \{\tau_{12}\},\\
&\mathcal{T}^s(l_{17}) = \{\tau_{13}, \tau_{17}\}, && \mathcal{T}^s(l_{18}) = \{\tau_{14}, \tau_{18}\},\\
&\mathcal{T}^s(l_{19}) = \{\tau_{15}, \tau_{19}\}, && \mathcal{T}^s(l_{20}) = \{\tau_{16}, \tau_{20}\}.
\end{aligned}$$

Let an order $\mathcal{L}^c$ be $l_1 < l_2 < l_3 < l_4$, and an order $\mathcal{L}^s_u$ be $U_1 < U_2$. Let an order $\mathcal{L}_{l_i}$ ($\mathcal{L}^s_{r_i}$) follow the rule $g^i_{l_j} < g^i_{l_k}$ ($g^i_{r_j} < g^i_{r_k}$) if $j < k$. Thm. 5 is applicable and decomposes the covariance $\text{Cov}(X, Y)$ over paths $l_{[1,20]}$:

$$\begin{aligned}
\text{Cov}(X, Y) = &\sum_{i=1}^{4} \text{Cov}^c_{l_i[x^*]}(X, Y)_{\mathcal{L}^c_\pi}\\
&+ \sum_{i=5}^{20} \text{Cov}^s_{l_i[x^*]}(X, Y)_{\mathcal{L}^s_\pi}.
\end{aligned}$$

The path-specific causal covariance $\left\{\text{Cov}^c_{l_i[x^*]}(X, Y)_{\mathcal{L}^c_\pi}\right\}_{i=1,\dots,4}$ are equal to:

$$\begin{aligned}
&\text{Cov}^c_{l_1[x^*]}(X, Y)_{\mathcal{L}^c_\pi} = \text{Cov}(X, Y - Y_{l_1[x^*]})\\
&= \text{Cov}(X, Y - Y_{x^*, W_1, W_2}),\\
&\text{Cov}^c_{l_2[x^*]}(X, Y)_{\mathcal{L}^c_\pi} = \text{Cov}(X, Y_{l_1[x^*]} - Y_{l_{[1,2]}[x^*]})\\
&= \text{Cov}(X, Y_{x^*, W_1, W_2} - Y_{x^*, W_{1_{x^*}}, W_2}),\\
&\text{Cov}^c_{l_3[x^*]}(X, Y)_{\mathcal{L}^c_\pi} = \text{Cov}(X, Y_{l_{[1,2]}[x^*]} - Y_{l_{[1,3]}[x^*]})\\
&= \text{Cov}(X, Y_{x^*, W_{1_{x^*}}, W_2} - Y_{x^*, W_{1_{x^*}}, W_{2_{x^*, W_1}}}),\\
&\text{Cov}^c_{l_3[x^*]}(X, Y)_{\mathcal{L}^c_\pi} = \text{Cov}(X, Y_{l_{[1,3]}[x^*]} - Y_{l_{[1,4]}[x^*]})\\
&= \text{Cov}(X, Y_{x^*, W_{1_{x^*}}, W_{2_{x^*, W_1}}} - Y_{x^*}).
\end{aligned}$$

The path-specific spurious covariance $\left\{\text{Cov}^s_{l_i[x^*]}(X, Y)_{\mathcal{L}^s_\pi}\right\}_{i=5,\dots,20}$ are equal to:

$$\begin{aligned}
&\text{Cov}^s_{l_5[x^*]}(X, Y)_{\mathcal{L}^s_\pi} = \text{Cov}^{ts}_{\tau_1[x^*]}(X, Y)_{\mathcal{L}^s_\pi}\\
&= \text{Cov}(X - X_{g^1_{l_1}}, Y_{x^*} - Y_{x^*, g^1_{r_1}})\\
&= \text{Cov}(X - X_{Z_{1_{U^l_1}}, Z_2}, Y_{x^*} - Y_{x^*, W_{1_{x^*}}, W_{2_{x^*}}, Z_{1_{U^r_1}}, Z_2}),\\
&\text{Cov}^s_{l_6[x^*]}(X, Y)_{\mathcal{L}^s_\pi} = \text{Cov}^{ts}_{\tau_2[x^*]}(X, Y)_{\mathcal{L}^s_\pi}\\
&= \text{Cov}(X - X_{g^1_{l_1}}, Y_{x^*, g^1_{r_1}} - Y_{x^*, g^1_{r_{[1,2]}}})\\
&= \text{Cov}(X - X_{Z_{1_{U^l_1}}, Z_2},\\
&\qquad Y_{x^*, W_{1_{x^*}}, W_{2_{x^*}}, Z_{1_{U^r_1}}, Z_2}\\
&\qquad - Y_{x^*, W_{1_{x^*, Z_{1_{U^r_1}}, Z_2}}, W_{2_{x^*}}, Z_{1_{U^r_1}}, Z_2}),
\end{aligned}$$

$$
\begin{aligned}
&\mathrm{Cov}^{s}_{l_7[x^*]}(X,Y)_{\mathcal{L}^s_\pi} = \mathrm{Cov}^{ts}_{\tau_3[x^*]}(X,Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}(X - X_{g^1_{l_1}}, Y_{x^*,g^1_{r_{[1,2]}}} - Y_{x^*,g^1_{r_{[1,3]}}}) \\
&= \mathrm{Cov}(X - X_{Z_{1_{U^l_1}},Z_2}, \\
&\qquad Y_{x^*,W_{1_{x^*,Z_{1_{U^r_1}},Z_2}},W_{2_{x^*}},Z_{1_{U^r_1}},Z_2} \\
&\qquad - Y_{x^*,W_{1_{x^*,Z_{1_{U^r_1}},Z_2}},W_{2_{x^*,W_{1_{x^*}},Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_2}),
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Cov}^{s}_{l_8[x^*]}(X,Y)_{\mathcal{L}^s_\pi} = \mathrm{Cov}^{ts}_{\tau_4[x^*]}(X,Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}(X - X_{g^1_{l_1}}, Y_{x^*,g^1_{r_{[1,3]}}} - Y_{x^*,g^1_{r_{[1,4]}}}) \\
&= \mathrm{Cov}(X - X_{Z_{1_{U^l_1}},Z_2}, \\
&\qquad Y_{x^*,W_{1_{x^*,Z_{1_{U^r_1}},Z_2}},W_{2_{x^*,W_{1_{x^*}},Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_2} \\
&\qquad - Y_{x^*,W_{1_{x^*,Z_{1_{U^r_1}},Z_2}},W_{2_{x^*,Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_2}),
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Cov}^{s}_{l_9[x^*]}(X,Y)_{\mathcal{L}^s_\pi} = \mathrm{Cov}^{ts}_{\tau_5[x^*]}(X,Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}(X - X_{g^1_{l_1}}, Y_{x^*,g^1_{r_{[1,4]}}} - Y_{x^*,g^1_{r_{[1,5]}}}) \\
&= \mathrm{Cov}(X - X_{Z_{1_{U^l_1}},Z_2}, \\
&\qquad Y_{x^*,W_{1_{x^*,Z_{1_{U^r_1}},Z_2}},W_{2_{x^*,Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_2} \\
&\qquad - Y_{x^*,W_{1_{x^*,Z_{1_{U^r_1}},Z_2}},W_{2_{x^*,Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_{2_{U^r_1}}}),
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Cov}^{s}_{l_{10}[x^*]}(X,Y)_{\mathcal{L}^s_\pi} = \mathrm{Cov}^{ts}_{\tau_6[x^*]}(X,Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}(X - X_{g^1_{l_1}}, Y_{x^*,g^1_{r_{[1,5]}}} - Y_{x^*,g^1_{r_{[1,6]}}}) \\
&= \mathrm{Cov}(X - X_{Z_{1_{U^l_1}},Z_2}, \\
&\qquad Y_{x^*,W_{1_{x^*,Z_{1_{U^r_1}},Z_2}},W_{2_{x^*,Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_{2_{U^r_1}}} \\
&\qquad - Y_{x^*,W_{1_{x^*,U^r_1}},W_{2_{x^*,Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_{2_{U^r_1}}}),
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Cov}^{s}_{l_{11}[x^*]}(X,Y)_{\mathcal{L}^s_\pi} = \mathrm{Cov}^{ts}_{\tau_7[x^*]}(X,Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}(X - X_{g^1_{l_1}}, Y_{x^*,g^1_{r_{[1,6]}}} - Y_{x^*,g^1_{r_{[1,7]}}}) \\
&= \mathrm{Cov}(X - X_{Z_{1_{U^l_1}},Z_2}, \\
&\quad Y_{x^*,W_{1_{x^*,U^r_1}},W_{2_{x^*,Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_{2_{U^r_1}}} \\
&\quad - Y_{x^*,W_{1_{x^*,U^r_1}},W_{2_{x^*,W_{1_{Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_{2_{U^r_1}}}},Z_{1_{U^r_1}},Z_{2_{U^r_1}}}),
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Cov}^{s}_{l_{12}[x^*]}(X,Y)_{\mathcal{L}^s_\pi} = \mathrm{Cov}^{ts}_{\tau_8[x^*]}(X,Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}(X - X_{g^1_{l_1}}, Y_{x^*,g^1_{r_{[1,7]}}} - Y_{x^*,g^1_{r_{[1,8]}}}) \\
&= \mathrm{Cov}(X - X_{Z_{1_{U^l_1}},Z_2}, \\
&\quad Y_{x^*,W_{1_{x^*,U^r_1}},W_{2_{x^*,W_{1_{Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_{2_{U^r_1}}}},Z_{1_{U^r_1}},Z_{2_{U^r_1}}} \\
&\quad - Y_{x^*,U^r_1}),
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Cov}^{s}_{l_{13}[x^*]}(X,Y)_{\mathcal{L}^s_\pi} = \mathrm{Cov}^{ts}_{\tau_9[x^*]}(X,Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}(X_{g^1_{l_1}} - X_{g^1_{l_{[1,2]}}}, Y_{x^*} - Y_{x^*,g^1_{r_1}}) \\
&= \mathrm{Cov}(X_{Z_{1_{U^l_1}},Z_2} - X_{U^l_1}, Y_{x^*} - Y_{x^*,W_{1_{x^*}},W_{2_{x^*}},Z_{1_{U^r_1}},Z_2}),
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Cov}^{s}_{l_{14}[x^*]}(X,Y)_{\mathcal{L}^s_\pi} = \mathrm{Cov}^{ts}_{\tau_{10}[x^*]}(X,Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}(X_{g^1_{l_1}} - X_{g^1_{l_{[1,2]}}}, Y_{x^*,g^1_{r_1}} - Y_{x^*,g^1_{r_{[1,2]}}}) \\
&= \mathrm{Cov}(X_{Z_{1_{U^l_1}},Z_2} - X_{U^l_1}, \\
&Y_{x^*,W_{1_{x^*}},W_{2_{x^*}},Z_{1_{U^r_1}},Z_2} - Y_{x^*,W_{1_{x^*,Z_{1_{U^r_1}},Z_2}},W_{2_{x^*}},Z_{1_{U^r_1}},Z_2}),
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Cov}^{s}_{l_{15}[x^*]}(X,Y)_{\mathcal{L}^s_\pi} = \mathrm{Cov}^{ts}_{\tau_{11}[x^*]}(X,Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}(X_{g^1_{l_1}} - X_{g^1_{l_{[1,2]}}}, Y_{x^*,g^1_{r_{[1,2]}}} - Y_{x^*,g^1_{r_{[1,3]}}}) \\
&= \mathrm{Cov}(X_{Z_{1_{U^l_1}},Z_2} - X_{U^l_1}, \\
&\qquad Y_{x^*,W_{1_{x^*,Z_{1_{U^r_1}},Z_2}},W_{2_{x^*}},Z_{1_{U^r_1}},Z_2} \\
&\qquad - Y_{x^*,W_{1_{x^*,Z_{1_{U^r_1}},Z_2}},W_{2_{x^*,W_{1_{x^*}},Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_2}),
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Cov}^{s}_{l_{16}[x^*]}(X,Y)_{\mathcal{L}^s_\pi} = \mathrm{Cov}^{ts}_{\tau_{12}[x^*]}(X,Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}(X_{g^1_{l_1}} - X_{g^1_{l_{[1,2]}}}, Y_{x^*,g^1_{r_{[1,3]}}} - Y_{x^*,g^1_{r_{[1,4]}}}) \\
&= \mathrm{Cov}(X_{Z_{1_{U^l_1}},Z_2} - X_{U^l_1}, \\
&\qquad Y_{x^*,W_{1_{x^*,Z_{1_{U^r_1}},Z_2}},W_{2_{x^*,W_{1_{x^*}},Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_2} \\
&\qquad - Y_{x^*,W_{1_{x^*,Z_{1_{U^r_1}},Z_2}},W_{2_{x^*,Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_2}),
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Cov}^{s}_{l_{17}[x^*]}(X,Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}^{ts}_{\tau_{13}[x^*]}(X,Y)_{\mathcal{L}^s_\pi} + \mathrm{Cov}^{ts}_{\tau_{17}[x^*]}(X,Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}(X_{g^1_{l_1}} - X_{g^1_{l_{[1,2]}}}, Y_{x^*,g^1_{r_{[1,4]}}} - Y_{x^*,g^1_{r_{[1,5]}}}) \\
&+ \mathrm{Cov}(X_{g^1_{l_{[1,2]}}} - X_{g^1_{l_{[1,2]}} \cup g^2_{l_1}}, Y_{x^*,g^1_{r_{[1,8]}}} - Y_{x^*,g^1_{r_{[1,8]}} \cup g^2_{r_1}}) \\
&= \mathrm{Cov}(X_{Z_{1_{U^l_1}},Z_2} - X_{U^l_1}, \\
&\qquad Y_{x^*,W_{1_{x^*,Z_{1_{U^r_1}},Z_2}},W_{2_{x^*,Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_2} \\
&\qquad - Y_{x^*,W_{1_{x^*,Z_{1_{U^r_1}},Z_2}},W_{2_{x^*,Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_{2_{U^r_1}}}) \\
&+ \mathrm{Cov}(X_{U^l_1} - X_{U^l_{[1,2]}}, Y_{x^*,U^r_1} \\
&\qquad - Y_{x^*,U^r_1,W_{1_{x^*,U^r_1}},W_{2_{x^*,U^r_1}},Z_{2_{U^r_{[1,2]}}}}),
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Cov}^{s}_{l_{18}[x^*]}(X,Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}^{ts}_{\tau_{14}[x^*]}(X,Y)_{\mathcal{L}^s_\pi} + \mathrm{Cov}^{ts}_{\tau_{18}[x^*]}(X,Y)_{\mathcal{L}^s_\pi} \\
&= \mathrm{Cov}(X_{g^1_{l_1}} - X_{g^1_{l_{[1,2]}}}, Y_{x^*,g^1_{r_{[1,5]}}} - Y_{x^*,g^1_{r_{[1,6]}}}) \\
&+ \mathrm{Cov}(X_{g^1_{l_{[1,2]}}} - X_{g^1_{l_{[1,2]}} \cup g^2_{l_1}}, Y_{x^*,g^1_{r_{[1,8]}} \cup g^2_{r_1}} - Y_{x^*,g^1_{r_{[1,8]}} \cup g^2_{r_{[1,2]}}}) \\
&= \mathrm{Cov}(X_{Z_{1_{U^l_1}},Z_2} - X_{U^l_1}, \\
&\qquad Y_{x^*,W_{1_{x^*,Z_{1_{U^r_1}},Z_2}},W_{2_{x^*,Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_{2_{U^r_1}}} \\
&\qquad - Y_{x^*,W_{1_{x^*,U^r_1}},W_{2_{x^*,Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_{2_{U^r_1}}}) \\
&+ \mathrm{Cov}(X_{U^l_1} - X_{U^l_{[1,2]}}, \\
&\qquad Y_{x^*,U^r_1,W_{1_{x^*,U^r_1}},W_{2_{x^*,U^r_1}},Z_{2_{U^r_{[1,2]}}}} \\
&\qquad - Y_{x^*,U^r_1,W_{1_{x^*,U^r_{[1,2]}}},W_{2_{x^*,U^r_1}},Z_{2_{U^r_{[1,2]}}}}),
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Cov}^{s}_{l_{19}[x^*]}(X,Y)_{\mathcal{L}^s_\pi}\\
&= \mathrm{Cov}^{ts}_{\tau_{15}[x^*]}(X,Y)_{\mathcal{L}^s_\pi} + \mathrm{Cov}^{ts}_{\tau_{19}[x^*]}(X,Y)_{\mathcal{L}^s_\pi}\\
&= \mathrm{Cov}(X_{g^1_{l_1}} - X_{g^1_{l_{[1,2]}}}, Y_{x^*,g^1_{r_{[1,6]}}} - Y_{x^*,g^1_{r_{[1,7]}}})\\
&+ \mathrm{Cov}(X_{g^1_{l_{[1,2]}}} - X_{g^1_{l_{[1,2]}} \cup g^2_{l_1}}, Y_{x^*,g^1_{r_{[1,8]}} \cup g^2_{r_{[1,2]}}} - Y_{x^*,g^1_{r_{[1,8]}} \cup g^2_{r_{[1,3]}}})\\
&= \mathrm{Cov}(X_{Z_{1_{U^l_1}},Z_2} - X_{U^l_1},\\
&\qquad Y_{x^*,W_{1_{x^*,U^r_1}},W_{2_{x^*,Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_{2_{U^r_1}}}\\
&\qquad - Y_{x^*,W_{1_{x^*,U^r_1}},W_{2_{x^*,W_{1_{Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_{2_{U^r_1}}}},Z_{1_{U^r_1}},Z_{2_{U^r_1}}})\\
&+ \mathrm{Cov}(X_{U^l_1} - X_{U^l_{[1,2]}},\\
&\qquad Y_{x^*,U^r_1,W_{1_{x^*,U^r_{[1,2]}}},W_{2_{x^*,U^r_1}},Z_{2_{U^r_{[1,2]}}}}\\
&\qquad - Y_{x^*,U^r_1,W_{1_{x^*,U^r_{[1,2]}}},W_{2_{x^*,U^r_{[1,2]}},W_{1_{x^*,U^r_1}}},Z_{2_{U^r_{[1,2]}}}}),\\
&\mathrm{Cov}^{s}_{l_{20}[x^*]}(X,Y)_{\mathcal{L}^s_\pi}\\
&= \mathrm{Cov}^{ts}_{\tau_{16}[x^*]}(X,Y)_{\mathcal{L}^s_\pi} + \mathrm{Cov}^{ts}_{\tau_{20}[x^*]}(X,Y)_{\mathcal{L}^s_\pi}\\
&= \mathrm{Cov}(X_{g^1_{l_1}} - X_{g^1_{l_{[1,2]}}}, Y_{x^*,g^1_{r_{[1,7]}}} - Y_{x^*,g^1_{r_{[1,8]}}})\\
&+ \mathrm{Cov}(X_{g^1_{l_{[1,2]}}} - X_{g^1_{l_{[1,2]}} \cup g^2_{l_1}}, Y_{x^*,g^1_{r_{[1,8]}} \cup g^2_{r_{[1,3]}}} - Y_{x^*,g^1_{r_{[1,8]}} \cup g^2_{r_{[1,4]}}})\\
&= \mathrm{Cov}(X_{Z_{1_{U^l_1}},Z_2} - X_{U^l_1},\\
&Y_{x^*,W_{1_{x^*,U^r_1}},W_{2_{x^*,W_{1_{Z_{1_{U^r_1}},Z_2}},Z_{1_{U^r_1}},Z_{2_{U^r_1}}}},Z_{1_{U^r_1}},Z_{2_{U^r_1}}} - Y_{x^*,U^r_1})\\
&+ \mathrm{Cov}(X_{U^l_1} - X_{U^l_{[1,2]}},\\
&Y_{x^*,U^r_1,W_{1_{x^*,U^r_{[1,2]}}},W_{2_{x^*,U^r_{[1,2]}},W_{1_{x^*,U^r_1}}},Z_{2_{U^r_{[1,2]}}}} - Y_{x^*,U^r_{[1,2]}}).
\end{aligned}
$$

References

D. Galles and J. Pearl. Axioms of causal relevance. *Artificial Intelligence*, 97(1-2):9–43, 1997.

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, 2000. 2nd edition, 2009.