Proceedings of Machine Learning Research vol 213:1–36, 2023 2nd Conference on Causal Learning and Reasoning

Factorization of the Partial Covariance in Singly-Connected Path Diagrams

Jose M. Peña JOSE.M.PENA@LIU.SE
STIMA, IDA, Linköping University, Sweden

Editors: Mihaela van der Schaar, Dominik Janzing and Cheng Zhang

Abstract

We extend path analysis by showing that, for a singly-connected path diagram, the partial covariance of two random variables factorizes over the nodes and edges in the path between the variables. This result allows us to determine the contribution of each node and edge to the partial covariance. It also allows us to show that Simpson's paradox cannot occur in singly-connected path diagrams.

Keywords: Path analysis, structural equation models, Simpson's paradox

1. Introduction

To ease interpretation, linear structural equation models are typically represented as path diagrams: Nodes represent random variables, directed edges represent direct causal relationships, and bidirected edges represent confounding, i.e. correlation between error terms. Moreover, each directed edge is annotated with the corresponding coefficient in the linear structural equation model, a.k.a. path coefficient. Likewise, each bidirected edge is annotated with the corresponding error correlation. A path diagram also brings in computational benefits. For instance, the covariance σ_{XY} of two random variables X and Y can be determined from the path diagram. Specifically, σ_{XY} can be expressed as the sum over the paths from X to Y of the product of path coefficients and error covariances of the edges in the path (Wright, 1921; Pearl, 2009). Hence, the covariance factorizes over the edges and nodes in the paths. In this work, we develop a similar factorization for the partial covariance $\sigma_{XY \cdot Z}$ in singly-connected path diagrams, i.e. the underlying undirected graph is a tree and, thus, no undirected cycle exists. While path analysis in a singly-connected path diagram determines the contribution of each node and edge to the covariance, our results determine the contribution of each node and edge to the partial covariance. Moreover, we use our results to show that Simpson's paradox cannot occur in singly-connected path diagrams. For path diagrams, Simpson's paradox can be described as the reversal of the sign of the regression coefficient of a random variable Y on a second variable X upon conditioning on a set of variables Z (Pearl, 2009, 2014).

Some previous works have certainly studied measures of association for singly-connected path diagrams, or for Gaussian random vectors in general. However, none of these works develop a factorization of the measure of association, as we do in this work. For singly-connected path diagrams, Chaudhuri and Richardson (2003) and Chaudhuri (2005) identify sufficient conditional independencies for ordering some squared partial correlation coefficients. Chaudhuri (2014) extends these results to general Gaussian random vectors. Chaudhuri and Tan (2010) report similar general results for absolute values of partial regression coefficients. Finally, Ong (2014) proves similar results for (signed) partial covariances, correlation coefficients and regression coefficients for singly-connected path diagrams and general Gaussian random vectors. In Section 5, we discuss further the work by

© 2023 J.M. Peña.

Ong. Finally, it should be mentioned that Pearl (2014) identifies three singly-connected path diagrams that cannot lead to Simpson's paradox. Our results are stronger, as we show that Simpson's paradox cannot occur in any singly-connected path diagram.

The rest of this work is structured as follows. Section 2 presents our factorization of partial covariances for singly-connected path diagrams with no colliders. Section 3 demonstrates our factorization on some examples. Section 4 extends the factorization to diagrams with colliders. Section 5 shows that our factorization implies that Simpson's paradox cannot occur in singly-connected path diagrams. Section 6 closes with some discussion.

2. Paths without Colliders

In this work, we make extensive use of the following recursive definition of the partial covariance of two random variables X and Y given a set of variables Z and a variable W such that $X, Y \notin Z \cup W$ (Anderson, 2003, Section 2.5.3):

$$\sigma_{XY \cdot ZW} = \sigma_{XY \cdot Z} - \frac{\sigma_{XW \cdot Z}\sigma_{WY \cdot Z}}{\sigma^2_{W \cdot Z}} \tag{1}$$

where, for simplicity, we use juxtaposition to denote union. Note that X and Y may be the same random variable, in which case the expression above corresponds to the partial variance $\sigma^2_{X \cdot ZW} = \sigma_{XX \cdot ZW}$. Recall that the partial (co)variances coincide with the conditional (co)variances for Gaussian random vectors.

We continue by recalling the separation criterion for path diagrams (Pearl, 2009, Section 1.2.3). For simplicity, we do not make any distinction between the nodes in the path diagrams and the random variables that they represent. Given a path π_{XY} from a node X to a node Y in a path diagram, a node C is a collider in π_{XY} if $A \circ\!\!\rightarrow C \leftarrow\!\!\circ B$ is a subpath of π_{XY}, where $\circ\!\!\rightarrow$ means $\rightarrow$ or $\leftrightarrow$. Given a set of nodes Z, π_{XY} is said to be Z-open if

- every collider in π_{XY} is in Z or has some descendant in Z, and
- every non-collider in π_{XY} is outside Z.

If there is no Z-open path from X to Y (which we denote as $X \perp Y|Z$), then we can readily conclude that X and Y are conditionally independent given Z in the joint normal distribution represented by the path diagram and, thus, $\sigma_{XY \cdot Z} = 0$ (Pearl, 2009). If on the other hand there is a Z-open path from X to Y (which we denote as $X \not\perp Y|Z$), we assume in this section that it has no colliders, and defer the case with colliders to the next section.

When $X \not\perp Y|\varnothing$, it is known from path analysis that the covariance σ_{XY} of two standardized random variables X and Y can be expressed as the sum for every $\varnothing$-open path from X to Y of the product of the path coefficients and error covariances of the edges in the path (Wright, 1921; Pearl, 2009). For non-standardized variables, one has to multiply the product associated to each path with the variance of the root variable in the path, i.e. the variable with no incoming edges. A path can have no root variables ($X \leftrightarrow Z \rightarrow \cdots \rightarrow Y$ or $X \leftarrow \cdots \leftarrow Z \leftrightarrow W \rightarrow \cdots \rightarrow Y$) or one root variable ($X \rightarrow \cdots \rightarrow Y$ or $X \leftarrow \cdots \leftarrow Z \rightarrow \cdots \rightarrow Y$).

When $X \not\perp Y|Z$ with $Z \neq \varnothing$, one may think that $\sigma_{XY \cdot Z}$ can be obtained by first applying path analysis to obtain an expression for σ_{XY} and, then, modifying this expression by replacing (co)variances with conditional (co)variances given Z. However, this is incorrect as the following example shows.

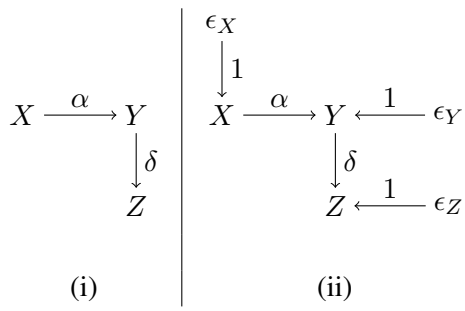

Figure 1: Path diagrams in Example 1.

Example 1 *Consider the path diagram (i) in Figure 1, which corresponds to the following linear structural equation model:*

$$X = \epsilon_X$$
$$Y = \alpha X + \epsilon_Y$$
$$Z = \delta Y + \epsilon_Z.$$

Consider representing the error terms explicitly in the diagram, which results in the path diagram (ii) in Figure 1. Then,

$$\sigma_{XY} = cov(X, \alpha X + \epsilon_Y) = \alpha\sigma_X^2 + cov(X, \epsilon_Y) = \alpha\sigma_X^2$$

where the last equality follows from the fact that $cov(X, \epsilon_Y) = 0$ since $X \perp \epsilon_Y | \varnothing$. However,

$$\sigma_{XY \cdot Z} = cov(X, \alpha X + \epsilon_Y | Z) = \alpha\sigma_{X \cdot Z}^2 + cov(X, \epsilon_Y | Z) \neq \alpha\sigma_{X \cdot Z}^2$$

where the last inequality follows from the fact that $cov(X, \epsilon_Y | Z) \neq 0$ in general, since $X \not\perp \epsilon_Y | Z$ because Z is a descendant of Y, which is a collider in the path from X to ϵ_Y.

For singly-connected path diagrams, the following two theorems show how to obtain $\sigma_{XY \cdot Z}$ from σ_{XY}. Interestingly, $\sigma_{XY \cdot Z}$ can still be written as a product over the nodes and edges in the path. See Appendix A for the proofs. Hereinafter, we use the following notation. The parents of a node X are $Pa(X) = \{Y | Y \to X\}$. The children of X are $Ch(X) = \{Y | X \to Y\}$. The spouses of X are $Sp(X) = \{Y | X \leftrightarrow Y\}$.

Theorem 1 *Let π_{XY} be of the form $X = X_m \leftarrow \cdots \leftarrow X_2 \leftarrow X_1 \to X_{m+1} \to \cdots \to X_{m+n} = Y$ or $X = X_1 \to X_2 \to \cdots \to X_{m+n} = Y$. Let Z^i be a set of nodes such that each is connected to $Pa(X_i) \cup Sp(X_i)$ by a path that does not contain any node in π_{XY}.[1] Let Z_i be a set of nodes such that each is connected to $Ch(X_i)$ by a path that does not contain any node in π_{XY}. Let $Z_i^i = Z_i \cup Z^i$ and $Z_{1:i}^{1:i} = Z_1^1 \cup \cdots \cup Z_i^i$. Then,*

$$\sigma_{XY \cdot Z_{1:m+n}^{1:m+n}} = \sigma_{XY} \frac{\sigma^2_{X_1 \cdot Z_1^1}}{\sigma^2_{X_1}} \prod_{i=2}^{m+n} \frac{\sigma^2_{X_i \cdot Z_{1:i-1}^{1:i-1} Z_i^i}}{\sigma^2_{X_i \cdot Z_{1:i-1}^{1:i-1} Z^i}}$$

where σ_{XY} is obtained by path analysis.

1. It suffices that each node in Z^i is connected to one node in $Pa(X_i) \cup Sp(X_i)$. The connecting path may be of length zero. The path does not need to be open with respect to any set of nodes.

Theorem 2 *Let π_{XY} be of the form $X = X_m \leftarrow \cdots \leftarrow X_2 \leftarrow X_1 \leftrightarrow X_{m+1} \rightarrow \cdots \rightarrow X_{m+n} = Y$ or $X = X_1 \leftrightarrow X_2 \rightarrow \cdots \rightarrow X_{m+n} = Y$. Let Z^i be a set of nodes such that each is connected to $Pa(X_i) \cup Sp(X_i)$ by a path that does not contain any node in π_{XY}. Let Z_i be a set of nodes such that each is connected to $Ch(X_i)$ by a path that does not contain any node in π_{XY}. Let $Z_i^i = Z_i \cup Z^i$ and $Z_{1:i}^{1:i} = Z_1^1 \cup \cdots \cup Z_i^i$. Then,*

$$\sigma_{XY \cdot Z_{1:m+n}^{1:m+n}} = \sigma_{XY} \prod_{i=1}^{m+n} \frac{\sigma^2_{X_i \cdot Z_{1:i-1}^{1:i-1} Z_i^i}}{\sigma^2_{X_i \cdot Z_{1:i-1}^{1:i-1} Z^i}}$$

where σ_{XY} is obtained by path analysis, and $Z_{1:0}^{1:0} = \varnothing$.

We demonstrate the theorems above on some examples in the next section. Before that, note that the numerator and denominator of the partial variance ratio in the theorems above only differ in that the conditioning set of the former is a superset of the conditioning set of the latter. Thus, the ratio is never greater than 1, since conditioning never increases the variance of a random variable. Therefore, the theorems above show that the partial covariance between two nodes can be computed by multiplying the expression for the covariance given by path analysis with a product of partial variance ratios that account for the reduction of the partial variances of the variables in the path between the two nodes. Thus, like the covariance, the partial covariance factorizes over the nodes and edges in the path. Moreover, the partial covariance is never greater than the covariance. However, both share the same sign, i.e. conditioning does not change the sign of the covariance. This implies that if two nodes X and Y are connected by a path of the form $X \rightarrow \cdots \rightarrow Y$, then conditioning does not change the sign of the regression coefficient of Y on X and, thus, of the causal effect of X on Y. This observation will be instrumental in proving in Section 5 that Simpson's paradox does not occur in singly-connected path diagrams. The following corollary is immediate.

Corollary 3 *Let π_{XY} be of the form in Theorems 1 or 2. Moreover, let π_{XY} be open with respect to the sets of nodes U and V. Then, $sign(\sigma_{XY}) = sign(\sigma_{XY \cdot U}) = sign(\sigma_{XY \cdot V})$.*

The expressions in the theorems above can be simplified by removing irrelevant variables from the conditioning set prior to applying the theorems. Specifically, let $T = Z_{1:m+n}^{1:m+n}$, and let $I = \{I_1, \ldots, I_s\}$ denote all the nodes in T such that $X \cup Y \perp I_i | T \setminus I_i$. Then, $X \cup Y \perp I | T \setminus I$ by repeated application of the intersection property (Studený, 2005, Proposition 2.1) and, thus, $\sigma_{XY \cdot T} = \sigma_{XY \cdot T \setminus I}$. In other words, I contains irrelevant nodes. As a matter of fact, I contains all the irrelevant nodes. To see it, assume to the contrary that there exists a second set of nodes $I' \nsubseteq I$ such that $X \cup Y \perp I' | T \setminus I'$. Then, $X \cup Y \perp I'_j | T \setminus I'_j$ for all $I'_j \in I'$ by the weak union property (Studený, 2005, Lemma 2.1), which contradicts the definition of I.

To sum up, the relevance of the theorems above lies in that they somehow complement path analysis: While path analysis in a singly-connected path diagram determines the contribution of each node and edge to the covariance σ_{XY}, the theorems above determine the contribution of each node and edge to the partial covariance $\sigma_{XY \cdot Z_{1:m+n}^{1:m+n}}$. Specifically, the theorems indicate whether the contribution of each node and edge to the covariance changes by conditioning and, if so, by how much. For example, consider Theorem 1 and let π_{XY} be of the form $X = X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_{m+n} = Y$. It follows from the theorem that the contribution of each edge in π_{XY} to the covariance and partial covariance is the same, namely the corresponding path coefficient. It follows from path analysis

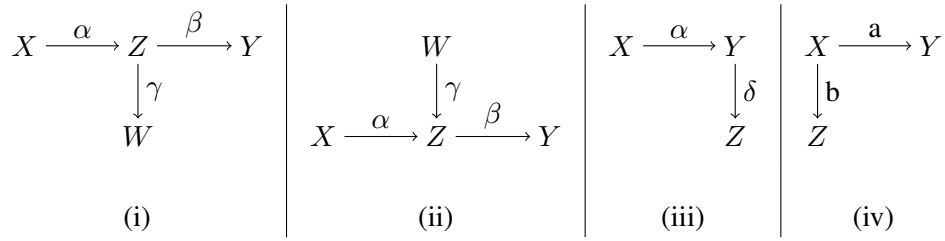

Figure 2: Path diagrams in Examples 2 and 3.

that the contribution of X_1 to the covariance is $\sigma^2_{X_1}$. The theorem shows that this contribution gets reduced by a factor of $\frac{\sigma^2_{X_1 \cdot Z^1_1}}{\sigma^2_{X_1}}$ when conditioning on Z^1_1. Likewise, the contribution of X_2 to the covariance is 1. This contribution gets reduced by a factor of $\frac{\sigma^2_{X_i \cdot Z^1_1 Z^2_2}}{\sigma^2_{X_i \cdot Z^1_1 Z^2}}$ when conditioning on Z^2_2. This indicates that conditioning on Z^2 may change the variance of X_2 but it does not constrain X_2 so as to alter the contribution of X_2. Conditioning on Z_2, on the other hand, has the opposite effect. Likewise for the rest of the nodes in π_{XY}. This fine-grained analysis is not possible with the recursion in Equation 1.

3. Causal Phenomena Explained

In this section, we demonstrate Theorems 1 and 2 on some examples that shed light on some causal phenomena. The examples are borrowed from Pearl (2013), who studied them using Equation 1. The objective of this section is purely illustrative. That is, we do not compare our explanations and those by Pearl (2013), as our theorems and Equation 1 address different problems.

Example 2 *Consider the path diagram (i) in Figure 2. The causal effect of X on Y is given by the regression coefficient $r_{YX} = \alpha\beta$. Since W does not lie on the causal path from X to Y, one may think that the causal effect of X on Y is also given by the partial regression coefficient $r_{YX \cdot W}$, which can be computed from the subpopulation satisfying $W = w$ for any w. However, this is incorrect as shown by Pearl (2013, Section 3.2). We arrive at the same conclusion as Pearl (2013) by applying Theorem 1 with $X_1 = X, X_2 = Z, X_3 = Y, Z^1_1 = Z^2 = Z^3 = Z^3_3 = \varnothing$, and $Z^2_2 = \{W\}$, which gives that*

$$\sigma_{XY \cdot W} = \sigma_{XY} \frac{\sigma^2_X}{\sigma^2_X} \frac{\sigma^2_{Z \cdot W}}{\sigma^2_Z} \frac{\sigma^2_{Y \cdot W}}{\sigma^2_{Y \cdot W}}.$$

Moreover, $\sigma_{XY} = \sigma^2_X \alpha\beta$ by path analysis. Then,

$$r_{YX \cdot W} = \frac{\sigma_{XY \cdot W}}{\sigma^2_{X \cdot W}} = \alpha\beta \frac{\sigma^2_X}{\sigma^2_{X \cdot W}} \frac{\sigma^2_{Z \cdot W}}{\sigma^2_Z}$$

and, thus, $r_{YX \cdot W} \neq \alpha\beta$ unless $\gamma = 0$ or $\alpha = \sigma_Z / \sigma_X$. To see it, note that

$$\sigma^2_{X \cdot W} = \sigma^2_X - \frac{\sigma_{XW}\sigma_{WX}}{\sigma^2_W} = \sigma^2_X - \frac{(\sigma^2_X \alpha\gamma)^2}{\sigma^2_W} = \sigma^2_X \Big(\frac{\sigma^2_W - \sigma^2_X \alpha^2 \gamma^2}{\sigma^2_W} \Big)$$

and, similarly,

$$\sigma^2_{Z\cdot W} = \sigma^2_Z\Big(\frac{\sigma^2_W - \sigma^2_Z\gamma^2}{\sigma^2_W}\Big).$$

Then,

$$\frac{\sigma^2_X}{\sigma^2_{X\cdot W}}\frac{\sigma^2_{Z\cdot W}}{\sigma^2_Z} = \frac{\sigma^2_W - \sigma^2_Z\gamma^2}{\sigma^2_W - \sigma^2_X\alpha^2\gamma^2} = 1$$

if and only if $\gamma = 0$ *or* $\alpha = \sigma_Z/\sigma_X$.[2]

As also shown by Pearl (2013, Section 3.2), no bias is introduced when conditioning on W *in the path diagram (ii) in Figure 2. We arrive at the same conclusion as Pearl (2013) by applying Theorem 1 with* $X_1 = X, X_2 = Z, X_3 = Y, Z^1_1 = Z^3_3 = \varnothing$*, and* $Z^2 = Z^2_2 = \{W\}$*, which gives that*

$$\sigma_{XY\cdot W} = \sigma_{XY}\frac{\sigma^2_X}{\sigma^2_X}\frac{\sigma^2_{Z\cdot W}}{\sigma^2_{Z\cdot W}}\frac{\sigma^2_{Y\cdot W}}{\sigma^2_{Y\cdot W}}.$$

Moreover, $\sigma_{XY} = \sigma^2_X\alpha\beta$ *by path analysis. Then,*

$$r_{YX\cdot W} = \frac{\sigma_{XY\cdot W}}{\sigma^2_{X\cdot W}} = \alpha\beta\frac{\sigma^2_X}{\sigma^2_{X\cdot W}} = \alpha\beta$$

where the last equality follows from the fact that $X \perp W|\varnothing$ *and, thus,* $\sigma^2_X = \sigma^2_{X\cdot W}$.

Example 3 *Consider the path diagram (iii) in Figure 2. The causal effect of* X *on* Y *is given by* $r_{YX} = \alpha$*. As shown by Pearl (2013, Section 3.3), conditioning on* Z *introduces a bias. We arrive at the same conclusion as Pearl (2013) by applying Theorem 1 with* $X_1 = X, X_2 = Y, Z^1_1 = Z^2 = \varnothing$*, and* $Z^2_2 = \{Z\}$*, which gives that*

$$\sigma_{XY\cdot Z} = \sigma_{XY}\frac{\sigma^2_X}{\sigma^2_X}\frac{\sigma^2_{Y\cdot Z}}{\sigma^2_Y}.$$

Moreover, $\sigma_{XY} = \sigma^2_X\alpha$ *by path analysis. Then,*

$$r_{YX\cdot Z} = \frac{\sigma_{XY\cdot Z}}{\sigma^2_{X\cdot Z}} = \alpha\frac{\sigma^2_X}{\sigma^2_{X\cdot Z}}\frac{\sigma^2_{Y\cdot Z}}{\sigma^2_Y}. \tag{2}$$

and, thus, $r_{YX\cdot Z} \neq \alpha$ *unless* $\delta = 0$ *or* $\alpha = \sigma_Y/\sigma_X$ *as shown in Example 2. In summary, the causal effect of* X *on* Y *cannot be computed from the subpopulation satisfying* $Z = z$ *because* $r_{YX\cdot Z} \neq \alpha$*. However, if* σ^2_X *and* σ^2_Z *are known, then the causal effect can be computed from that subpopulation by correcting* $r_{YX\cdot Z}$ *as shown in Equation 2.*

As also shown by Pearl (2013, Section 3.3), no bias is introduced when conditioning on Z *in the path diagram (iv) in Figure 2. We arrive at the same conclusion as Pearl (2013) by applying Theorem 1 with* $X_1 = X, X_2 = Y, Z^1_1 = \{Z\}$*, and* $Z^2 = Z^2_2 = \varnothing$*, which gives that*

$$\sigma_{XY\cdot Z} = \sigma_{XY}\frac{\sigma^2_{X\cdot Z}}{\sigma^2_X}\frac{\sigma^2_{Y\cdot Z}}{\sigma^2_{Y\cdot Z}}.$$

2. The effect of setting $\gamma = 0$ on $r_{YX\cdot W}$ is as follows. Setting $\gamma = 0$ is equivalent to removing the edge $Z \to W$ from the path diagram (i) in Figure 2, which implies that $r_{YX\cdot W} = \alpha\beta$. The effect of setting $\alpha = \sigma_Z/\sigma_X$ on $r_{YX\cdot W}$ is as follows. The path diagram (i) in Figure 2 corresponds to a model that contains the linear structural equation $Z = \alpha X + \epsilon_Z$ with $X \perp \epsilon_Z|\varnothing$. Then, $\sigma^2_Z = \alpha^2\sigma^2_X + var(\epsilon_Z)$ and, thus, $var(\epsilon_Z) = 0$ when $\alpha = \sigma_Z/\sigma_X$, i.e. Z is completely determined by X. In other words, the diagram (i) in Figure 2 reduces to the diagram (iv), which is studied in Example 3.

Moreover, $\sigma_{XY} = \sigma_X^2 a$ *by path analysis. Then,*

$$r_{YX \cdot Z} = \frac{\sigma_{XY \cdot Z}}{\sigma_{X \cdot Z}^2} = a.$$

The examples above show that conditioning on a child of a mediator or on a child of the effect introduces a bias in the estimation of the causal effect of interest. Appendix B illustrates with experiments how this bias may lead to suboptimal decision making. On the other hand, the examples above show that conditioning on a parent of a mediator or on a child of the cause does not introduce any bias, which implies that the causal effect of interest can be computed from the corresponding subpopulation.

For completeness, we show below that conditioning on a parent of the cause or on a parent of the effect does not introduce any bias.

Example 4 *Consider the path diagram (ii) in Figure 2. The causal effect of* Z *on* Y *is given by* $r_{YZ} = \beta$*. We conclude that* $r_{YZ \cdot W} = \beta$ *by applying Theorem 1 with* $X_1 = Z, X_2 = Y, Z_1^1 = \{W\}$*, and* $Z^2 = Z_2^2 = \varnothing$*. Specifically,*

$$\sigma_{ZY \cdot W} = \sigma_{ZY} \frac{\sigma_{Z \cdot W}^2}{\sigma_Z^2} \frac{\sigma_{Y \cdot W}^2}{\sigma_{Y \cdot W}^2}.$$

Moreover, $\sigma_{ZY} = \sigma_Z^2 \beta$ *by path analysis. Then,*

$$r_{YZ \cdot W} = \frac{\sigma_{ZY \cdot W}}{\sigma_{Z \cdot W}^2} = \beta.$$

This result also follows from the first rule of do-calculus (Pearl, 2009, Section 3.4).

Consider again the path diagram (ii) in Figure 2. The causal effect of X *on* Z *is given by* $r_{ZX} = \alpha$*. We conclude that* $r_{ZX \cdot W} = \alpha$ *by applying Theorem 1 with* $X_1 = X, X_2 = Z, Z_1^1 = \varnothing$*, and* $Z^2 = Z_2^2 = \{W\}$*. Specifically,*

$$\sigma_{XZ \cdot W} = \sigma_{XZ} \frac{\sigma_X^2}{\sigma_X^2} \frac{\sigma_{Z \cdot W}^2}{\sigma_{Z \cdot W}^2}.$$

Moreover, $\sigma_{XZ} = \sigma_X^2 \alpha$ *by path analysis. Then,*

$$r_{ZX \cdot W} = \frac{\sigma_{XZ \cdot W}}{\sigma_{X \cdot W}^2} = \alpha \frac{\sigma_X^2}{\sigma_{X \cdot W}^2} = \alpha$$

where the last equality follows from the fact that $X \perp W | \varnothing$ *and, thus,* $\sigma_X^2 = \sigma_{X \cdot W}^2$*. We can arrive at the same conclusion by applying the definition of partial covariance. Specifically,*

$$\sigma_{XZ \cdot W} = \sigma_{XZ} - \frac{\sigma_{XW} \sigma_{WZ}}{\sigma_W^2} = \sigma_{XZ}$$

because $X \perp W | \varnothing$ *implies that* $\sigma_{XW} = 0$*.*

Finally, the example below shows that Theorems 1 and 2 may be of help even when the path diagram at hand is not singly-connected. The example is borrowed from Pearl (2013).

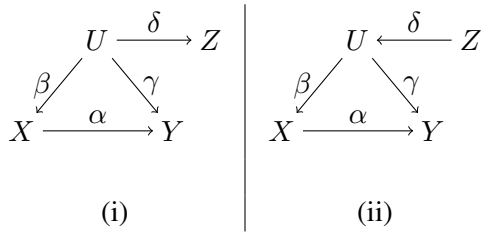

Figure 3: Path diagrams in Example 5.

Example 5 *Consider the path diagram (i) in Figure 3. Let us denote it by G. Moreover, let G^α denote the diagram that results when the edge $X \to Y$ is deleted from G. Since $X \perp Y|U$ holds in G^α, we have that $\alpha = r_{YX\cdot U}$ (Pearl, 2009, Theorem 5.3.1). However, if U is unobserved then $r_{YX\cdot U}$ cannot be computed. Assume that the proxy Z of U is observed and, thus, $r_{YX\cdot Z}$ can be computed. Of course, $\alpha \neq r_{YX\cdot Z}$ because $X \perp Y|Z$ does not hold in G^α. However, Pearl (2013, Section 3.11) shows that the bias introduced by adjusting for Z instead of U vanishes as the correlation between U and Z grows, i.e. when Z is a good proxy of U. The same occurs in the path diagram (ii) in Figure 3.*

Although the path diagrams in Figure 3 are not singly-connected, we can still use our results to reach the same conclusions as Pearl. Since the covariance of X and Y may differ in G and G^α, we use σ_{XY} for the former and σ^α_{XY} for the latter. For the same reason, we distinguish between $\sigma_{XY\cdot Z}$ and $\sigma^\alpha_{XY\cdot Z}$. Since the variance of U is the same in G and G^α, we simply denote it as σ^2_U. For the same reason, we use $\sigma^2_{U\cdot Z}$ to denote the partial variance of U given Z in both G and G^α. Note that checking whether $X \perp Y|Z$ holds in G^α is equivalent to checking whether $\sigma^\alpha_{XY\cdot Z} = 0$ holds. Since G^α is a singly-connected path diagram, we can apply Theorem 1 and conclude that $\sigma^\alpha_{XY\cdot Z} = \sigma^\alpha_{XY}\sigma^2_{U\cdot Z}/\sigma^2_U$. This implies that, although conditioning on Z does not nullify the covariance of X and Y in G^α, it does reduce it. Moreover, the greater the correlation between U and Z, the greater the reduction and, thus, the closer $r_{YX\cdot Z}$ comes to α. We illustrate this with some experiments in Appendix C.

4. Paths with Colliders

In this section, we address the case where π_{XY} has colliders. Specifically, let π_{XY} be Z-open. Given a collider C in π_{XY}, an opener is any node $W \in Z$ such that $C = C_1 \to \cdots \to C_n = W$ and $C_1, \ldots, C_{n-1} \notin Z$. Note that C is an opener if $C \in Z$.

Theorem 4 *Let C be a collider in π_{XY}. Moreover, let π_{XY} be closed with respect to Z but open with respect to $Z \cup Z^{1:n}_{1:n} \cup W_{1:n}$ where (i) $W_1, \ldots, W_n$ are openers for C, (ii) Z^i is a set of nodes such that each is connected to $Pa(W_i) \cup Sp(W_i)$ by a path that does not contain any node in π_{XY} or π_{CW_i}, (iii) Z_i is a set of nodes such that each is connected to $Ch(W_i)$ by a path, and (iv) $Z^{1:j}_{1:i} = Z_1 \cup \cdots \cup Z_i \cup Z^1 \cup \cdots \cup Z^j$. Then,*

$$\sigma_{XY\cdot ZZ^{1:n}_{1:n}W_{1:n}} = -\sum_{i=1}^{n} \frac{\sigma_{XW_i\cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}}\sigma_{W_iY\cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}}}{\sigma^2_{W_i\cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}}} \tag{3}$$

where $Z^{1:1}_{1:0} = Z^1$ and $W_{1:0} = \varnothing$.

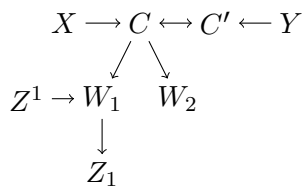

Figure 4: Path diagram in Example 6.

In the theorem above, if π_{XW_i} has some collider then $\sigma_{XW_i \cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}}$ is obtained by recursively applying the theorem to π_{XW_i}. When π_{XW_i} has no colliders, $\sigma_{XW_i \cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}}$ is obtained as shown in Theorems 1 and 2. Likewise for π_{W_iY} and $\sigma_{W_iY \cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}}$. Example 6 demonstrates this recursive procedure. Specifically, let π_{XY} have colliders $C_1, \ldots, C_k$, where C_i has openers $\mathcal{W}_i = \{W_{i1}, \ldots, W_{in_i}\}$. Then, the recursive procedure just described allows us to write $\sigma_{XY \cdot ZZ^{1:n}_{1:n}W_{1:n}}$ as

$$(-1)^k \sum_{O_1 \in \mathcal{W}_1} \cdots \sum_{O_k \in \mathcal{W}_k} \frac{\sigma_{XO_1 \cdot U_{O_1}} \sigma_{O_1O_2 \cdot U_{O_{1:2}}} \cdots \sigma_{O_{k-1}O_k \cdot U_{O_{1:k}}} \sigma_{O_kY \cdot U_{O_{1:k}}}}{\sigma^2_{O_1 \cdot U_{O_1}} \sigma^2_{O_2 \cdot U_{O_{1:2}}} \cdots \sigma^2_{O_{k-1} \cdot U_{O_{1:k}}} \sigma^2_{O_k \cdot U_{O_{1:k}}}} \tag{4}$$

for some sets of nodes $U_{O_1}, U_{O_{1:2}}, \ldots, U_{O_{1:k}}$. In other words, the partial covariance decomposes as a sum over the different ways of opening π_{XY}, and each term in the sum is a product of calls to Theorems 1 and 2. Then, each term in the sum factorizes over the nodes and edges of π_{XY}. This resembles how path analysis on unconstrained path diagrams decomposes the covariance of two random variables over the different $\varnothing$-open paths between them. We demonstrate the theorem above with an example.

Example 6 *Consider the path diagram in Figure 4. Then, the partial covariance $\sigma_{XY \cdot C'Z^1_1W_{1:2}}$ can be computed with the help of Theorem 4 with $Z = \{C'\}$. Specifically,*

$$\sigma_{XY \cdot C'Z^1_1W_{1:2}} = -\frac{\sigma_{XW_1 \cdot C'Z^1} \sigma_{W_1Y \cdot C'Z^1}}{\sigma^2_{W_1 \cdot C'Z^1}} - \frac{\sigma_{XW_2 \cdot C'Z^1_1W_1} \sigma_{W_2Y \cdot C'Z^1_1W_1}}{\sigma^2_{W_2 \cdot C'Z^1_1W_1}}.$$

Moreover, $\sigma_{XW_1 \cdot C'Z^1}$ and $\sigma_{XW_2 \cdot C'Z^1_1W_1}$ can be computed as shown in Theorem 1. On the other hand, $\sigma_{W_1Y \cdot C'Z^1}$ and $\sigma_{W_2Y \cdot C'Z^1_1W_1}$ can be computed by applying Theorem 4 again with $Z = \{Z^1\}$ and $Z = Z^1_1 \cup \{W_1\}$, respectively. Specifically,

$$\sigma_{W_1Y \cdot C'Z^1} = -\frac{\sigma_{W_1C' \cdot Z^1} \sigma_{C'Y \cdot Z^1}}{\sigma^2_{C' \cdot Z^1}}$$

and

$$\sigma_{W_2Y \cdot C'Z^1_1W_1} = -\frac{\sigma_{W_2C' \cdot Z^1_1W_1} \sigma_{C'Y \cdot Z^1_1W_1}}{\sigma^2_{C' \cdot Z^1_1W_1}}$$

where the partial covariances in the numerators can be computed as shown in Theorems 1 and 2. Putting all together, we have that

$$\sigma_{XY \cdot C'Z^1_1W_{1:2}} = \frac{\sigma_{XW_1 \cdot C'Z^1} \sigma_{W_1C' \cdot Z^1} \sigma_{C'Y \cdot Z^1}}{\sigma^2_{W_1 \cdot C'Z^1} \sigma^2_{C' \cdot Z^1}} + \frac{\sigma_{XW_2 \cdot C'Z^1_1W_1} \sigma_{W_2C' \cdot Z^1_1W_1} \sigma_{C'Y \cdot Z^1_1W_1}}{\sigma^2_{W_2 \cdot C'Z^1_1W_1} \sigma^2_{C' \cdot Z^1_1W_1}}$$

which confirms Equation 4 and the discussion thereof.

5. Simpson's Paradox

In this section, we use the theorems developed in the previous sections to show that Simpson's paradox does not occur in singly-connected path diagrams. For path diagrams, Simpson's paradox can be described as the reversal of the sign of the regression coefficient of a random variable Y on a second variable X upon conditioning on a set of variables Z. Note that this is a generalization of the definition by Pearl (2013, 2014), who restricts Z to a singleton. Pearl (2013, Section 3.1) shows that the paradox can well occur for the path diagram $X \rightarrow Y \leftarrow Z \rightarrow X$. Note that the diagram is not singly-connected. Pearl (2014, Section 2.2) argues that the paradox does not occur for the singly-connected path diagrams $Z \leftarrow X \rightarrow Y$, $Z \rightarrow X \rightarrow Y$, and $X \rightarrow Y \leftarrow Z$, because the association between X and Y is collapsible over Z. However, the correctness of this statement depends on the definition of association. To see it, recall from Pearl (2009, Definition 6.5.1) that given a functional $g(p(x,y))$ that measures the association between two random variables Y and X in $p(x,y)$, we say that g is collapsible over a variable Z if

$$E_z[g(p(x,y|z))] = g(p(x,y)).$$

If we now consider the diagram $Z \leftarrow X \rightarrow Y$ and let g be the covariance between Y and X, then collapsibility does not hold since

$$E_z[g(p(x,y|z))] = E_z[cov(X,Y|Z=z)] = cov(X,Y|Z) = \sigma_{XY \cdot Z} \neq \sigma_{XY} = g(p(x,y))$$

where the second equality follows from the fact that the conditional covariance does not depend on the value on which we condition, and the inequality is proven in Example 3. Similarly for the diagram $Z \rightarrow X \rightarrow Y$ as shown in Example 4. For the diagram $X \rightarrow Y \leftarrow Z$, on the other hand, collapsibility does hold as shown in Example 4. If we instead let g be the regression coefficient of Y on X, then collapsibility holds for the three diagrams under consideration, as shown in Examples 3 and 4. Moreover, Pearl (2014) does not discuss if Simpson's paradox can occur for the diagram $X \rightarrow Y \rightarrow Z$. Recall that Example 3 shows that collapsibility does not hold for this diagram, regardless of whether association means covariance or regression coefficient. Pearl does not discuss either the case of singly-connected path diagrams where X and Y are connected by a path of length greater than one with and without colliders, or the case where the conditioning set contains more than one variable. We fill these gaps below.

Note that Simpson's paradox concerns the sign of the regression coefficient of a random variable Y on a random variable X upon conditioning on a set of variables Z or, equivalently, it concerns the sign of the covariance between X and Y upon conditioning on Z. Therefore, we are interested in the collapsibility of the sign rather than in the collapsibility of the regression coefficient or covariance. Corollary 3 implies that conditioning does not change the sign of the covariance for paths without colliders. The following theorem shows that this also holds for paths with colliders. Consequently, Simpson's paradox cannot occur in any singly-connected path diagram.

Theorem 5 *Let π_{XY} be open with respect to the sets of nodes U and V. Then, $sign(\sigma_{XY \cdot U}) = sign(\sigma_{XY \cdot V})$.*

Note that the result above is actually stronger than required to disprove Simpson's paradox, because U may neither include nor be included in V.

It is worth mentioning that Ong (2014, Theorem 6.6) proves the theorem above by other means when U and V only contain descendants of the colliders in π_{XY}. Ong states that the proof can be extended to include other conditionates, as in our theorem. However, since he does not provide the details, we believe that our proof fills some gap. Moreover, Ong does not discuss the relevance of this result for disproving Simpson's paradox. Therefore, our discussion above fills some gap, too.

6. Discussion

In this work, we have extended path analysis by showing that, for a singly-connected path diagram, the partial covariance of two random variables factorizes over the nodes and edges in the path between the variables. This result applies even when the path contains colliders. We find the case where the path has no colliders particularly interesting, since then the partial covariance can be computed by multiplying the expression for the covariance given by path analysis with a product of partial variance ratios that account for the reduction of the partial variances of the variables in the path. Moreover, these results have allowed us to show that Simpson's paradox cannot occur in singly-connected path diagrams. Naturally, we would like in the future to extend our results beyond singly-connected path diagrams. Appendix E presents a first attempt in this direction.

Acknowledgments

We thank the Area Chair and Reviewers for their comments. We acknowledge financial support from the Swedish Research Council (ref. 2019-00245).

References

T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, 2003.

S. Chaudhuri. *Using the Structure of d-Connecting Paths as a Qualitative Measure of the Strength of Dependence*. PhD thesis, University of Washington, 2005.

S. Chaudhuri. Qualitative Inequalities for Squared Partial Correlations of a Gaussian Random Vector. *Annals of the Institute of Statistical Mathematics*, 66:345–367, 2014.

S. Chaudhuri and T. Richardson. Using the Structure of d-Connecting Paths as a Qualitative Measure of the Strength of Dependence. In *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence*, pages 116–123, 2003.

S. Chaudhuri and G. L. Tan. On Qualitative Comparison of Partial Regression Coefficients for Gaussian Graphical Markov Models. In *Algebraic Methods in Statistics and Probability II. Contemporary Mathematics, Vol. 516*, pages 125–133. American Mathematical Society, 2010.

V. M. H. Ong. *Model Selection for Graphical Markov Models*. PhD thesis, National University of Singapore, 2014.

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2009.

J. Pearl. Linear Models: A Useful "Microscope" for Causal Analysis. *Journal of Causal Inference*, 1:155–170, 2013.

J. Pearl. Comment: Understanding Simpson's Paradox. *The American Statistician*, 68:8–13, 2014.

M. Studený. *Probabilistic Conditional Independence Structures*. Springer, 2005.

R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, 2018.

S. Wright. Correlation and Causation. *Journal of Agricultural Research*, 20:557–585, 1921.

Appendix A. Proofs of Sections 2-5

Recall that in all the results in this appendix the path diagram is assumed to be singly-connected.

Lemma 6 *Let S be the root node in a path π_{XY} without colliders, i.e. $A \leftarrow S \rightarrow B$ or $S \rightarrow B$ is a subpath of π_{XY}. Note that $S = X$ or $S = Y$ in the latter case. Let W be a set of nodes such that each is connected to $Pa(S) \cup Ch(S) \cup Sp(S)$ by a path that does not contain any node in π_{XY}. Then,*

$$\sigma_{XY \cdot ZW} = \sigma_{XY \cdot Z} \frac{\sigma^2_{S \cdot ZW}}{\sigma^2_{S \cdot Z}}.$$

Proof Assume that W is a singleton. Consider first the case where $A \leftarrow S \rightarrow B$ is a subpath of π_{XY}. Note that $X \perp W | Z \cup S$. Then,

$$0 = \sigma_{XW \cdot ZS} = \sigma_{XW \cdot Z} - \frac{\sigma_{XS \cdot Z} \sigma_{SW \cdot Z}}{\sigma^2_{S \cdot Z}}$$

which implies that $\sigma_{XW \cdot Z} = \delta_{XS \cdot Z} \sigma_{SW \cdot Z}$ where $\delta_{XS \cdot Z} = \sigma_{XS \cdot Z} / \sigma^2_{S \cdot Z}$. Likewise, $Y \perp W | Z \cup S$ implies that $\sigma_{YW \cdot Z} = \delta_{YS \cdot Z} \sigma_{SW \cdot Z}$ where $\delta_{YS \cdot Z} = \sigma_{YS \cdot Z} / \sigma^2_{S \cdot Z}$. Likewise, $X \perp Y | Z \cup S$ implies that

$$0 = \sigma_{XY \cdot ZS} = \sigma_{XY \cdot Z} - \frac{\sigma_{XS \cdot Z} \sigma_{SY \cdot Z}}{\sigma^2_{S \cdot Z}}$$

which implies that $\sigma_{XY \cdot Z} = \delta_{XS \cdot Z} \delta_{YS \cdot Z} \sigma^2_{S \cdot Z}$. Therefore,

$$\begin{aligned}
\sigma_{XY \cdot ZW} &= \sigma_{XY \cdot Z} - \frac{\sigma_{XW \cdot Z} \sigma_{WY \cdot Z}}{\sigma^2_{W \cdot Z}} \\
&= \delta_{XS \cdot Z} \delta_{YS \cdot Z} \sigma^2_{S \cdot Z} - \frac{\delta_{XS \cdot Z} \sigma_{SW \cdot Z} \delta_{YS \cdot Z} \sigma_{SW \cdot Z}}{\sigma^2_{W \cdot Z}} \\
&= \delta_{XS \cdot Z} \delta_{YS \cdot Z} \left(\sigma^2_{S \cdot Z} - \frac{\sigma_{SW \cdot Z} \sigma_{SW \cdot Z}}{\sigma^2_{W \cdot Z}} \right) \\
&= \delta_{XS \cdot Z} \delta_{YS \cdot Z} \sigma^2_{S \cdot ZW} = \sigma_{XY \cdot Z} \frac{\sigma^2_{S \cdot ZW}}{\sigma^2_{S \cdot Z}}.
\end{aligned}$$

Now, consider the case where $S \rightarrow B$ is a subpath of π_{XY}. Assume without loss of generality that $S = X$. Note that $Y \perp W | Z \cup X$. Then,

$$0 = \sigma_{YW \cdot ZX} = \sigma_{YW \cdot Z} - \frac{\sigma_{YX \cdot Z} \sigma_{XW \cdot Z}}{\sigma^2_{X \cdot Z}}$$

which implies that

$$\sigma_{YW \cdot Z} = \frac{\sigma_{YX \cdot Z}\sigma_{XW \cdot Z}}{\sigma^2_{X \cdot Z}}.$$

Therefore,

$$\begin{aligned}\sigma_{XY \cdot ZW} &= \sigma_{XY \cdot Z} - \frac{\sigma_{XW \cdot Z}\sigma_{WY \cdot Z}}{\sigma^2_{W \cdot Z}} = \sigma_{XY \cdot Z} - \frac{\sigma_{XW \cdot Z}\sigma_{YX \cdot Z}\sigma_{XW \cdot Z}}{\sigma^2_{W \cdot Z}\sigma^2_{X \cdot Z}} \\ &= \sigma_{XY \cdot Z}\Big(1 - \frac{\sigma_{XW \cdot Z}\sigma_{XW \cdot Z}}{\sigma^2_{W \cdot Z}\sigma^2_{X \cdot Z}}\Big) = \frac{\sigma_{XY \cdot Z}}{\sigma^2_{X \cdot Z}}\Big(\sigma^2_{X \cdot Z} - \frac{\sigma_{XW \cdot Z}\sigma_{XW \cdot Z}}{\sigma^2_{W \cdot Z}}\Big) \\ &= \sigma_{XY \cdot Z}\frac{\sigma^2_{X \cdot ZW}}{\sigma^2_{X \cdot Z}}.\end{aligned}$$

Repeated application of the paragraphs above proves the result for when W is a set. Specifically, let $W = \{W_1, \ldots, W_n\}$. Then,

$$\sigma_{XY \cdot ZW_1} = \sigma_{XY \cdot Z}\frac{\sigma^2_{S \cdot ZW_1}}{\sigma^2_{S \cdot Z}}$$

by replacing W with W_1 in the paragraphs above. Likewise,

$$\sigma_{XY \cdot ZW_1W_2} = \sigma_{XY \cdot ZW_1}\frac{\sigma^2_{S \cdot ZW_1W_2}}{\sigma^2_{S \cdot ZW_1}}$$

by replacing Z and W with $Z \cup \{W_1\}$ and W_2, respectively, in the paragraphs above. These last two results imply that

$$\sigma_{XY \cdot ZW_1W_2} = \sigma_{XY \cdot Z}\frac{\sigma^2_{S \cdot ZW_1W_2}}{\sigma^2_{S \cdot Z}}.$$

Continuing with this process for $W_3, \ldots, W_n$ yields the desired result. ■

Lemma 7 *Let S be a non-root node in a path π_{XY} without colliders, i.e. $A \circ\!\!\rightarrow S \rightarrow B$ or $A \circ\!\!\rightarrow S$ is a subpath of π_{XY}. Note that $S = X$ or $S = Y$ in the latter case. Let W be a set of nodes such that each is connected to $Pa(S) \cup Sp(S)$ by a path that does not contain any node in π_{XY}. Then,*

$$\sigma_{XY \cdot ZW} = \sigma_{XY \cdot Z}$$

if Z contains no descendants of S.

Proof Assume that W is a singleton. Then,

$$\sigma_{XY \cdot ZW} = \sigma_{XY \cdot Z} - \frac{\sigma_{XW \cdot Z}\sigma_{WY \cdot Z}}{\sigma^2_{W \cdot Z}}$$

which implies that $\sigma_{XY \cdot ZW} = \sigma_{XY \cdot Z}$ because $\sigma_{XW \cdot Z} = 0$ or $\sigma_{WY \cdot Z} = 0$ since $X \perp W|Z$ or $W \perp Y|Z$.

Repeated application of the paragraph above proves the result for when W is a set. Specifically, let $W = \{W_1, \ldots, W_n\}$. Then,

$$\sigma_{XY \cdot ZW_1} = \sigma_{XY \cdot Z}$$

by replacing W with W_1 in the paragraph above. Likewise,

$$\sigma_{XY \cdot ZW_1W_2} = \sigma_{XY \cdot ZW_1}$$

by replacing Z and W with $Z \cup \{W_1\}$ and W_2, respectively, in the paragraph above. These last two results imply that

$$\sigma_{XY \cdot ZW_1W_2} = \sigma_{XY \cdot Z}.$$

Continuing with this process for $W_3, \ldots, W_n$ yields the desired result. ■

Lemma 8 *Let S be a non-root node in a path π_{XY} without colliders, i.e. $A \leftarrow\!\!\circ S \rightarrow B$ or $A \leftarrow\!\!\circ S$ is a subpath of π_{XY}. Note that $S = X$ or $S = Y$ in the latter case. Let W be a set of nodes such that each is connected to $Ch(S)$ by a path that does not contain any node in π_{XY}. Then,*

$$\sigma_{XY \cdot ZW} = \sigma_{XY \cdot Z} \frac{\sigma^2_{S \cdot ZW}}{\sigma^2_{S \cdot Z}}.$$

Proof Assume that W is a singleton. Consider first the case where $A \leftarrow\!\!\circ S \rightarrow B$ is a subpath of π_{XY}. Note that $X \perp W | Z \cup S$. Then,

$$0 = \sigma_{XW \cdot ZS} = \sigma_{XW \cdot Z} - \frac{\sigma_{XS \cdot Z}\sigma_{SW \cdot Z}}{\sigma^2_{S \cdot Z}}$$

which implies that $\sigma_{XW \cdot Z} = \sigma_{XS \cdot Z}\delta_{SW \cdot Z}$ where $\delta_{SW \cdot Z} = \sigma_{SW \cdot Z}/\sigma^2_{S \cdot Z}$. Note also that $Y \perp W | Z \cup S$. Then,

$$0 = \sigma_{YW \cdot ZS} = \sigma_{YW \cdot Z} - \frac{\sigma_{YS \cdot Z}\sigma_{SW \cdot Z}}{\sigma^2_{S \cdot Z}}$$

which implies that $\sigma_{YW \cdot Z} = \delta_{YS \cdot Z}\sigma_{SW \cdot Z}$ where $\delta_{YS \cdot Z} = \sigma_{YS \cdot Z}/\sigma^2_{S \cdot Z}$. Likewise, $X \perp Y | Z \cup S$ implies that

$$0 = \sigma_{XY \cdot ZS} = \sigma_{XY \cdot Z} - \frac{\sigma_{XS \cdot Z}\sigma_{SY \cdot Z}}{\sigma^2_{S \cdot Z}}$$

which implies that $\sigma_{XY \cdot Z} = \sigma_{XS \cdot Z}\delta_{YS \cdot Z}$. Therefore,

$$\begin{aligned}
\sigma_{XY \cdot ZW} &= \sigma_{XY \cdot Z} - \frac{\sigma_{XW \cdot Z}\sigma_{WY \cdot Z}}{\sigma^2_{W \cdot Z}} = \sigma_{XY \cdot Z} - \frac{\sigma_{XS \cdot Z}\delta_{SW \cdot Z}\delta_{YS \cdot Z}\sigma_{SW \cdot Z}}{\sigma^2_{W \cdot Z}} \\
&= \sigma_{XY \cdot Z}\Big(1 - \frac{\delta_{SW \cdot Z}\sigma_{SW \cdot Z}}{\sigma^2_{W \cdot Z}}\Big) = \frac{\sigma_{XY \cdot Z}}{\sigma^2_{S \cdot Z}}\Big(\sigma^2_{S \cdot Z} - \frac{\sigma^2_{S \cdot Z}\delta_{SW \cdot Z}\sigma_{SW \cdot Z}}{\sigma^2_{W \cdot Z}}\Big) \\
&= \sigma_{XY \cdot Z}\frac{\sigma^2_{S \cdot ZW}}{\sigma^2_{S \cdot Z}}.
\end{aligned}$$

Now, consider the case where $A \leftarrow\!\!\circ S$ is a subpath of π_{XY}. Assume without loss of generality that $S = Y$. Note that $X \perp W | Z \cup Y$. Then,

$$0 = \sigma_{XW \cdot ZY} = \sigma_{XW \cdot Z} - \frac{\sigma_{XY \cdot Z}\sigma_{YW \cdot Z}}{\sigma^2_{Y \cdot Z}}$$

which implies that

$$\sigma_{XW \cdot Z} = \frac{\sigma_{XY \cdot Z}\sigma_{YW \cdot Z}}{\sigma^2_{Y \cdot Z}}.$$

Therefore,

$$\begin{aligned}\sigma_{XY\cdot ZW} &= \sigma_{XY\cdot Z} - \frac{\sigma_{XW\cdot Z}\sigma_{WY\cdot Z}}{\sigma^2_{W\cdot Z}} = \sigma_{XY\cdot Z} - \frac{\sigma_{XY\cdot Z}\sigma_{YW\cdot Z}\sigma_{WY\cdot Z}}{\sigma^2_{Y\cdot Z}\sigma^2_{W\cdot Z}} \\ &= \sigma_{XY\cdot Z}\left(1 - \frac{\sigma_{YW\cdot Z}\sigma_{WY\cdot Z}}{\sigma^2_{Y\cdot Z}\sigma^2_{W\cdot Z}}\right) = \frac{\sigma_{XY\cdot Z}}{\sigma^2_{Y\cdot Z}}\left(\sigma^2_{Y\cdot Z} - \frac{\sigma_{YW\cdot Z}\sigma_{WY\cdot Z}}{\sigma^2_{W\cdot Z}}\right) \\ &= \sigma_{XY\cdot Z}\frac{\sigma^2_{Y\cdot ZW}}{\sigma^2_{Y\cdot Z}}.\end{aligned}$$

Repeated application of the paragraphs above proves the result for when W is a set. Specifically, let $W = \{W_1, \ldots, W_n\}$. Then,

$$\sigma_{XY\cdot ZW_1} = \sigma_{XY\cdot Z}\frac{\sigma^2_{S\cdot ZW_1}}{\sigma^2_{S\cdot Z}}$$

by replacing W with W_1 in the paragraphs above. Likewise,

$$\sigma_{XY\cdot ZW_1W_2} = \sigma_{XY\cdot ZW_1}\frac{\sigma^2_{S\cdot ZW_1W_2}}{\sigma^2_{S\cdot ZW_1}}$$

by replacing Z and W with $Z \cup \{W_1\}$ and W_2, respectively, in the paragraphs above. These last two results imply that

$$\sigma_{XY\cdot ZW_1W_2} = \sigma_{XY\cdot Z}\frac{\sigma^2_{S\cdot ZW_1W_2}}{\sigma^2_{S\cdot Z}}.$$

Continuing with this process for $W_3, \ldots, W_n$ yields the desired result. ■

Proof of Theorem 1 First, note that

$$\sigma_{X_mX_{m+n}\cdot Z_1^1} = \sigma_{X_mX_{m+n}}\frac{\sigma^2_{X_1\cdot Z_1^1}}{\sigma^2_{X_1}}$$

by Lemma 6. Then, note that

$$\sigma_{X_mX_{m+n}\cdot Z_1^1Z^2} = \sigma_{X_mX_{m+n}\cdot Z_1^1}$$

by Lemma 7. Finally, note that

$$\sigma_{X_mX_{m+n}\cdot Z_1^1Z^2Z_2} = \sigma_{X_mX_{m+n}\cdot Z_1^1Z^2}\frac{\sigma^2_{X_1\cdot Z_1^1Z^2Z_2}}{\sigma^2_{X_1\cdot Z_1^1Z^2}}$$

by Lemma 8. Continuing with this process for the rest of the nodes yields the desired result. ■

Proof of Theorem 2 First, note that

$$\sigma_{X_mX_{m+n}\cdot Z^1} = \sigma_{X_mX_{m+n}}$$

by Lemma 7. Then, note that

$$\sigma_{X_mX_{m+n}\cdot Z^1Z_1} = \sigma_{X_mX_{m+n}\cdot Z^1}\frac{\sigma^2_{X_1\cdot Z^1Z_1}}{\sigma^2_{X_1\cdot Z^1}}$$

by Lemma 8. Continuing with this process for the rest of the nodes yields the desired result. ■

Proof of Theorem 4 First, note that

$$\begin{aligned}\sigma_{XY\cdot ZZ^{1:n}_{1:n}W_{1:n}} &= \sigma_{XY\cdot ZZ^{1:n}_{1:n-1}W_{1:n}} \\ &= \sigma_{XY\cdot ZZ^{1:n}_{1:n-1}W_{1:n-1}} - \frac{\sigma_{XW_n\cdot ZZ^{1:n}_{1:n-1}W_{1:n-1}}\sigma_{W_nY\cdot ZZ^{1:n}_{1:n-1}W_{1:n-1}}}{\sigma^2_{W_n\cdot ZZ^{1:n}_{1:n-1}W_{1:n-1}}} \\ &= \sigma_{XY\cdot ZZ^{1:n-1}_{1:n-1}W_{1:n-1}} - \frac{\sigma_{XW_n\cdot ZZ^{1:n}_{1:n-1}W_{1:n-1}}\sigma_{W_nY\cdot ZZ^{1:n}_{1:n-1}W_{1:n-1}}}{\sigma^2_{W_n\cdot ZZ^{1:n}_{1:n-1}W_{1:n-1}}}\end{aligned}$$

because $X \cup Y \perp Z_n | Z \cup Z^{1:n}_{1:n-1} \cup W_{1:n}$ and $X \cup Y \perp Z^n | Z \cup Z^{1:n-1}_{1:n-1} \cup W_{1:n-1}$. Then, the theorem follows by recursively applying the paragraph above to $\sigma_{XY\cdot ZZ^{1:n-1}_{1:n-1}W_{1:n-1}}$ until $n-1=0$, in which case $\sigma_{XY\cdot ZZ^{1:n-1}_{1:n-1}W_{1:n-1}} = \sigma_{XY\cdot Z} = 0$ because $X \perp Y | Z$. ■

Lemma 9 *Let π_{XY} be a path that is open with respect to a set of nodes U. Then, the sign of $\sigma_{XY\cdot U}$ does not depend on U, i.e. $sign(\sigma_{XY}) = sign(\sigma_{XY\cdot U})$.*

Proof If π_{XY} has no colliders, then the result follows from Corollary 3. Otherwise, consider any collider C in π_{XY}, and split U into $Z \cup Z^{1:n}_{1:n} \cup W_{1:n}$ as indicated in Theorem 4. We prove the result by induction over the number of colliders in π_{XY}. First, assume that C is the only collider in π_{XY}. Consider the following two cases.

Case 1. Suppose that C is an opener for the collider C. Then, C is the only opener by definition and, thus, the summation in Equation 3 reduces to

$$\frac{\sigma_{XW_1\cdot ZZ^1}\sigma_{W_1Y\cdot ZZ^1}}{\sigma^2_{W_1\cdot ZZ^1}}$$

with $W_1 = C$.

Case 2. Suppose that C is not an opener for the collider C. Then, note that $X \perp W_i | Z \cup Z^{1:i}_{1:i-1} \cup W_{1:i-1} \cup C$ for all i. Then,

$$0 = \sigma_{XW_i\cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}C} = \sigma_{XW_i\cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}} - \frac{\sigma_{XC\cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}}\sigma_{CW_i\cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}}}{\sigma^2_{C\cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}}}$$

and thus

$$\sigma_{XW_i\cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}} = \frac{\sigma_{XC\cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}}\sigma_{CW_i\cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}}}{\sigma^2_{C\cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}}}.$$

Likewise for $\sigma_{W_iY\cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}}$. Then, each term in the summation in Equation 3 can be rewritten as

$$\frac{\sigma_{XC\cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}}\sigma_{CY\cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}}\sigma^2_{CW_i\cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}}}{\sigma^2_{W_i\cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}}\sigma^4_{C\cdot ZZ^{1:i}_{1:i-1}W_{1:i-1}}}.$$

In both cases above, π_{XC} and π_{YC} have no colliders and, thus, the signs of $\sigma_{XC \cdot M}$ and $\sigma_{CY \cdot N}$ do not depend on M and N by Corollary 3 for any sets of nodes M and N. Therefore, the sign of $\sigma_{XY \cdot Z Z_{1:n}^{1:n} W_{1:n}}$ is the same in both cases above and, moreover, it does not depend on $Z \cup Z_{1:n}^{1:n} \cup W_{1:n}$.

After proving above that the result holds for any path with zero or one collider, we now assume as induction hypothesis that the result holds for any path with fewer than k colliders. To prove it for k colliders, we simply let C be any collider in π_{XY} and consider the same two cases as above. Note that π_{XC} and π_{CY} may now have colliders. So, Corollary 3 does not apply. However, since π_{XC} and π_{CY} have fewer than k colliders, the induction hypothesis does apply, which leads to the same conclusions as before. ■

Proof of Theorem 5 It follows from Lemma 9. ■

Appendix B. Suboptimal Decision Making I

In this appendix, we show that the bias introduced by conditioning on a child of the effect (recall Example 3) may lead to suboptimal decision making. We do so with the help of the following fictitious but, in our opinion, realistic scenario. Doctor 1 and Doctor 2 both treat a certain disease by administering approximately 5 units of drug X, i.e. $X \sim \mathcal{N}(5, \sigma_X)$. The doctors use different methods to administer the drug, which we suspect affects the effectiveness of the drug. The effectiveness of the drug is assessed by measuring the abundance of Y in blood, which is determined by X, i.e. $Y = \alpha_i X + \epsilon_Y$ for Doctor i and $\epsilon_Y \sim \mathcal{N}(0, \sigma_Y)$. The higher the value of Y the higher the effectiveness of the treatment. Moreover, the doctors also monitor the abundances of Z and W in blood, which are determined by respectively X and Y, specifically $Z = X + \epsilon_Z$ and $W = Y + \epsilon_W$ for both doctors and $\epsilon_Z \sim \mathcal{N}(0, \sigma_Z)$ and $\epsilon_W \sim \mathcal{N}(0, \sigma_W)$. The doctors divide the treatments into ordinary and extraordinary. Specifically, Doctor 1 declares the treatment ordinary when $4 < Z < 6$, and Doctor 2 when $4 < W < 6$. The doctors share with us data only about ordinary treatments. They believe that extraordinary treatments may lead to new findings about the disease at hand and, thus, they are not willing to share them as of today.

The problem above can be rephrased as follows. We want to estimate α_1 in the following path diagram (Doctor 1) from a sample of the subpopulation satisfying $4 < Z < 6$:

$$\begin{tikzcd} X \arrow[r, "\alpha_1"] \arrow[d, "1"] & Y \\ Z & \end{tikzcd}$$

We also want to estimate α_2 in the following path diagram (Doctor 2) from a sample of the subpopulation satisfying $4 < W < 6$:

$$\begin{tikzcd} X \arrow[r, "\alpha_2"] & Y \arrow[d, "1"] \\ & W \end{tikzcd}$$

As discussed in Example 3, the estimate of α_1 will be unbiased, whereas the estimate of α_2 will be biased. This may make us recommend the suboptimal doctor to future patients. We illustrate this below with some experiments.

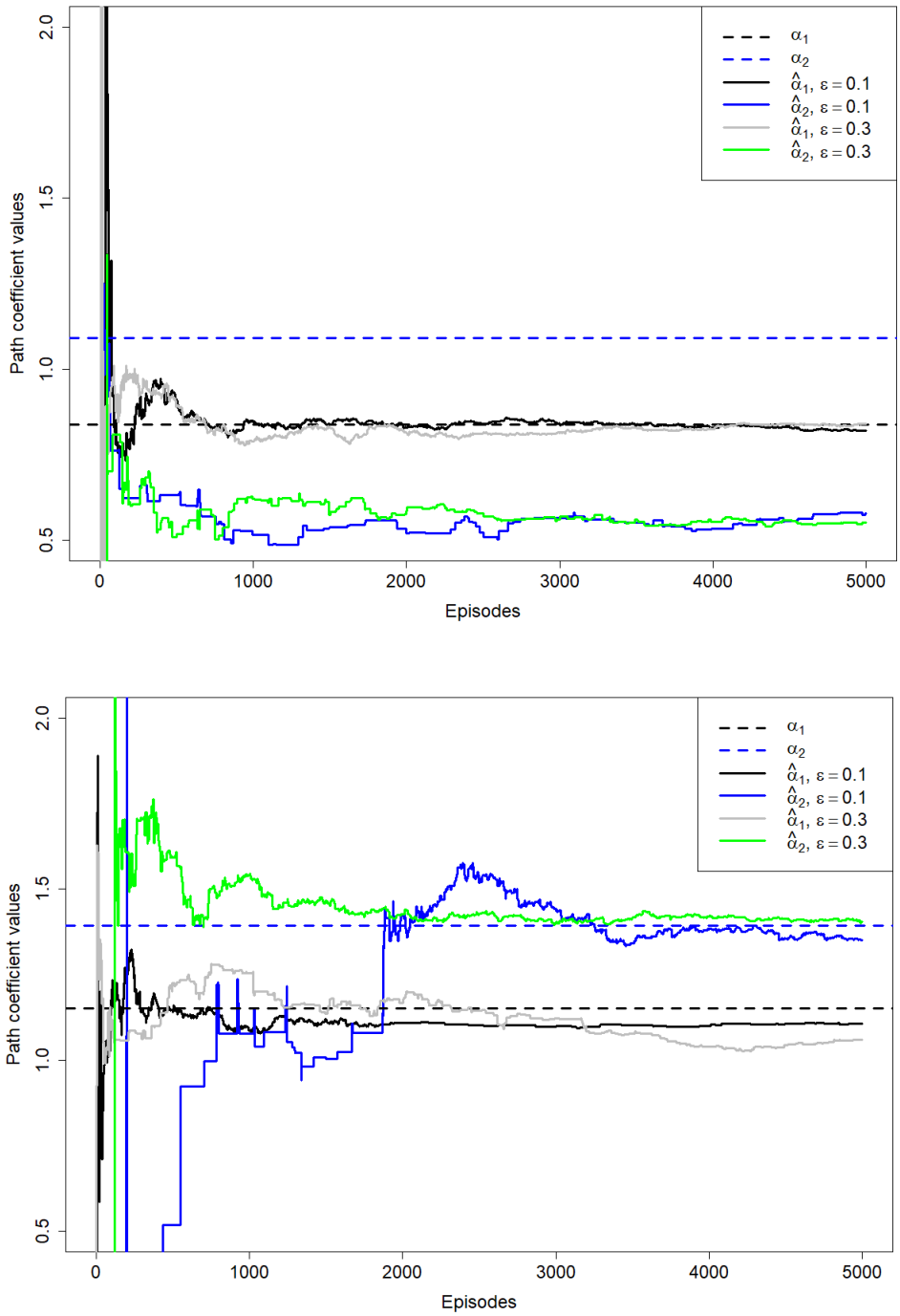

Figure 5: Estimates of α_1 and α_2 obtained by ϵ-greedy without correction (top) and with correction (bottom).

Estimating α_1 and α_2 above can be seen as an instance of the exploration/exploitation dilemma: In order to learn the effectiveness of the treatment administered by a particular doctor, the doctor has to administer the treatment to some patients, which leads to some patients receiving suboptimal treatment. A straightforward solution to this dilemma consists in combining exploration and exploitation as follows: Select the doctor with the highest effectiveness so far (i.e. exploitation) with probability $1 - \epsilon$, otherwise select the doctor at random (i.e. exploration). This strategy is called ϵ-greedy in the reinforcement learning community (Sutton and Barto, 2018).

Figure 5 (top) shows the estimates of α_1 and α_2 (denoted as $\hat{\alpha}_1$ and $\hat{\alpha}_2$) obtained by running ϵ-greedy for 5000 iterations (a.k.a. episodes). Each episode consists in selecting a doctor for treating a patient. The doctor shares the data with us only if the treatment is regarded as ordinary. In our experiments, this means that each episode starts by choosing a doctor, say Doctor 1, according to the ϵ-greedy strategy. Then, a triplet of values (x, y, z) is sampled from the corresponding linear structural equation model.[3] Finally, the triplet is kept if $4 < z < 6$ and discarded otherwise. In the figure, we can clearly see that $\hat{\alpha}_1$ converges to α_1, whereas $\hat{\alpha}_2$ does not converge to α_2. Moreover, $\hat{\alpha}_1$ converges to a larger value than $\hat{\alpha}_2$, which means that Doctor 1 is considered more effective than Doctor 2 and, thus, we should recommend the former. This is suboptimal because, as shown in the figure, α_2 is greater than α_1 and, thus, Doctor 2 should be preferred. This conclusion was consistent across many runs of the experiment. In each run, α_1 and α_2 were sampled uniformly from the intervals $(0.5, 1.5)$ and $(\alpha_1 + 0.15, \alpha_1 + 0.3)$ respectively, i.e. Doctor 2 was more effective than Doctor 1. In each run, $\sigma_X = \sigma_Y = \sigma_Z = \sigma_W = 1$.

As discussed in Example 3, if we can estimate σ_X^2 and σ_Y^2, then we can correct the bias in $\hat{\alpha}_2$. To illustrate this, assume that the doctors do not share with us data about individual extraordinary treatments but they do share aggregated data, in particular some estimates of σ_X^2 and σ_Y^2 (which they can compute from all the ordinary and extraordinary treatments performed). Figure 5 (bottom) shows $\hat{\alpha}_1$ and $\hat{\alpha}_2$ when the correction is applied to the latter. We can appreciate that both path coefficient estimates converge to the true values, and that Doctor 2 is now preferred. Again, this conclusion was consistent across many runs of the experiment.

Of course, ϵ-greedy is not the only way of solving the problem above. Alternative solutions include Thompson sampling, upper confidence bound (UCB) or directly performing a randomized controlled trial. However, the conclusions should not differ essentially from the ones presented above. The code for our experiments is publicly available at

`https://www.dropbox.com/s/hawshrihhgr5uvi/MAB.zip?dl=0`

Appendix C. Suboptimal Decision Making II

In this appendix, we show that the bias introduced by adjusting for a faithful proxy of a confounder is negligible for decision making (recall Example 5). However, the bias may be substantial when adjusting for a proxy of a non-confounder in a confounding path, which may lead to suboptimal decision making. We do so with the help of the following fictitious but, in our opinion, realistic scenario. Doctor 1 and Doctor 2 both treat a certain disease by administering a dose of drug X. The dose is determined by the abundance of U in blood. The doctors use different methods to administer the drug, which we suspect affects the effectiveness of the drug. The effectiveness of the drug is assessed by measuring the abundance of Y in blood, which is determined by X and U. The lower

3. Since negative abundance values do not make sense, if x, y or z are negative then a new triplet is sampled. This seldom happens, anyway.

the value of Y the higher the effectiveness of the treatment. Unwilling to disclose further details about the treatment, the doctors do not share with us any measurements of U. However, they do provide us with measurements of two proxies of U. Specifically, Doctor 1 provides us with the abundance of Z in blood, whereas Doctor 2 provides us with the abundance of W in blood. The former is known to be caused by U, whereas the latter is known to cause U.[4]

In the language of path diagrams, the problem above can be stated as follows. We want to estimate α_1 in the following path diagram (Doctor 1) from a sample for X, Y and Z:

$$\begin{tikzcd} & U \arrow[r, "1"] \arrow[dl, "1"'] \arrow[dr, "1"] & Z \\ X \arrow[rr, "\alpha_1"] & & Y \end{tikzcd}$$

We also want to estimate α_2 in the following diagram (Doctor 2) from a sample for X, Y and W:

$$\begin{tikzcd} & U \arrow[dl, "1"'] \arrow[dr, "1"] & W \arrow[l, "1"'] \\ X \arrow[rr, "\alpha_2"] & & Y \end{tikzcd}$$

Recall that U is unobserved. As discussed in Example 5, if Z and W are faithful proxies of U, then the estimates of $r_{YX \cdot Z}$ and $r_{YX \cdot W}$ should be close to α_1 and α_2, respectively, which implies that we should be able to identify the optimal doctor. We illustrate this below with some experiments.

As in Appendix B, we use ϵ-greedy to solve the problem above. We consider $\epsilon = 0.2$ and 5000 episodes. In each run of ϵ-greedy, α_1 and α_2 are sampled uniformly from the intervals $(0.5, 1.5)$ and $(\alpha_1 - 0.3, \alpha_1 - 0.15)$ respectively, i.e. Doctor 2 is more effective than Doctor 1. The standard deviations of the error terms are all equal to 1, with the exception of the term corresponding to Z for Doctor 1 and the term corresponding to U for Doctor 2. Specifically, $\sigma_Z, \sigma_U = 0.1, 0.5, 1$. The smaller the values of σ_Z and σ_U the better Z and W are as proxies of U. Figure 6 (top) shows a representative run of the many that we performed. We can see that Doctor 2 is preferred if Z and W are equally good proxies of U, i.e. $\sigma_Z = \sigma_U$. Moreover, both $\hat{\alpha}_1$ and $\hat{\alpha}_2$ converge to the true values when Z and W are faithful proxies of U, i.e. $\sigma_Z = \sigma_U = 0.1$.

The experiments above may lead one to conclude that blocking a confounding path by adjusting for a proxy does not bias much the estimate of a causal effect as long as the proxy is a good one. However, this is not true. To illustrate it, we repeat the experiments above after replacing the confounding path from X to Y in the path diagrams with the confounding path $X \leftarrow U' \rightarrow U \rightarrow Y$. Additional parameters are all set to 1. Figure 6 (bottom) shows a representative run of the new experiments. We can clearly see that $\hat{\alpha}_1$ converges to a smaller value than $\hat{\alpha}_2$ for every combination of σ_Z and σ_U, i.e. no matter how good Z and W are as proxies of U. This means that Doctor 1 is considered more effective than Doctor 2 and, thus, that we should recommend the former. This is suboptimal because, as shown in the figure, α_2 is smaller than α_1 and, thus, Doctor 2 should be preferred. Note also that $\hat{\alpha}_1$ converges to α_1 when Z is almost a perfect proxy of U. On the other hand, $\hat{\alpha}_2$ behaves bad no matter how good W is as a proxy of U. In summary, on the negative side, we wrongly recommend Doctor 1 but, on the positive side, we can estimate her effectiveness accurately if Z is a faithful proxy of U. To get further insight into these results, we can repeat the

4. Some authors would call Z a proxy and W a driver.

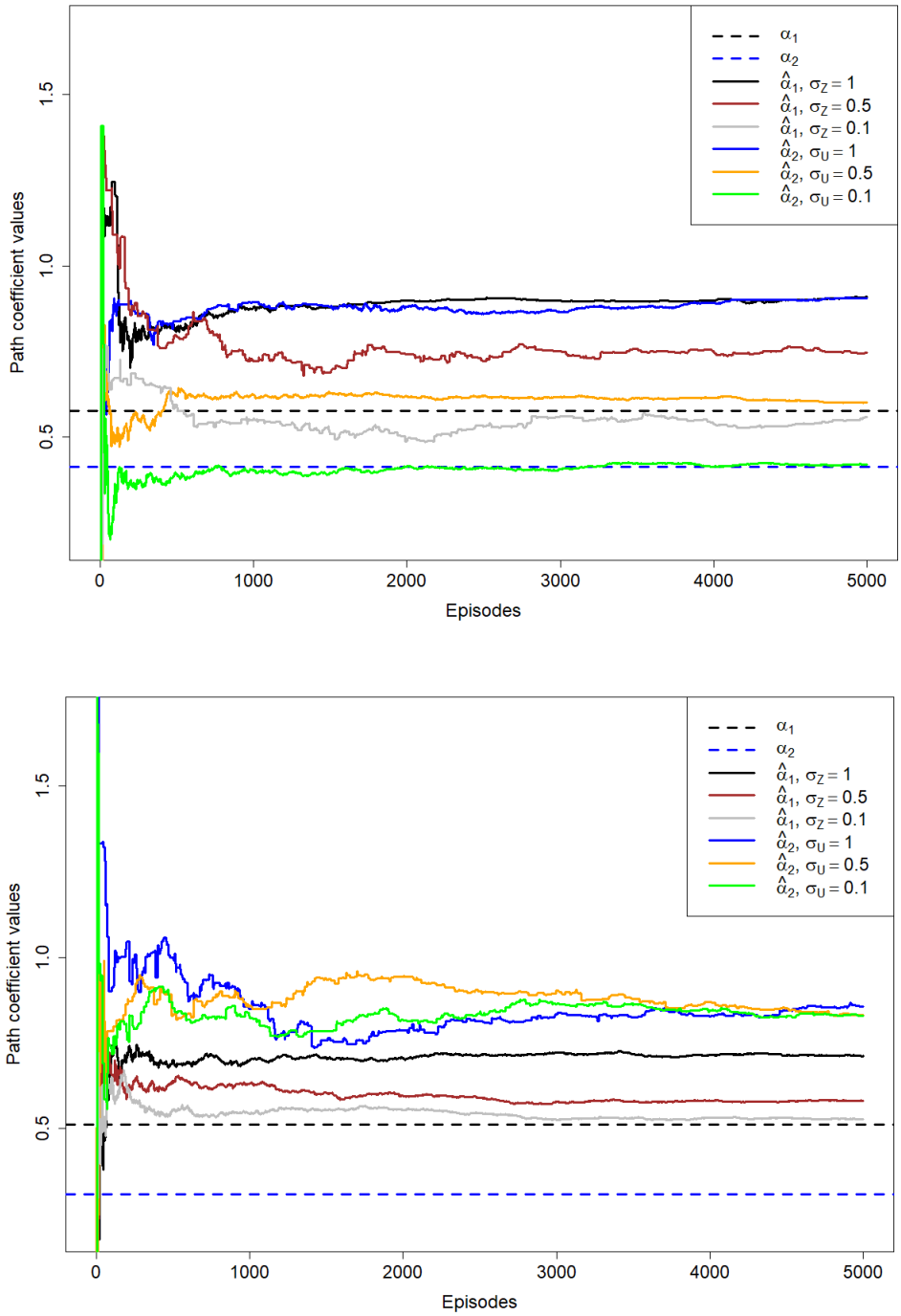

Figure 6: Estimates of α_1 and α_2 obtained by ϵ-greedy when adjusting for a proxy of a confounder (top) and for a proxy of a non-confounder (bottom).

reasoning in Example 5 now for the path diagrams of Doctor 1 and Doctor 2 after replacing the confounding path from X to Y with the confounding path $X \leftarrow U' \rightarrow U \rightarrow Y$. Let G denote the path diagram of Doctor 1. Since G^{α_1} is singly-connected, we can apply Theorem 1 and conclude that $\sigma^{\alpha_1}_{XY \cdot Z} = \sigma^{\alpha_1}_{XY} \sigma^2_{U \cdot Z} / \sigma^2_U$. This implies that conditioning on Z reduces the covariance between X and Y. Moreover, the greater the correlation between U and Z, the greater the reduction and, thus, the closer $r_{YX \cdot Z}$ comes to α_1. Let G now denote the path diagram of Doctor 2. Applying Theorem 1 to G^{α_2} gives that $\sigma^{\alpha_2}_{XY \cdot W} = \sigma^{\alpha_2}_{XY}$. In other words, conditioning on W leaves the covariance of X and Y unchanged. Moreover, $X \perp W | \varnothing$ in G and, thus, $\sigma_{X \cdot W} = \sigma_X$ and, thus, $r_{YX \cdot W} = r_{YX}$. In other words, adjusting for W does not solve our problem, even if W is almost a perfect proxy of U. In summary, the effectiveness of adjusting for a proxy depends on the type of confounding path, the type of causal relation between the proxy and the unobserved variable, and the correlation between them.

Appendix D. Beyond Singly-Connected Path Diagrams

In this appendix, we extend Theorems 1 and 2 from singly-connected diagrams to a superclass thereof. Note that we then only consider paths without colliders. The extension to path with colliders seems complicated.

In Section 2, we defined the separation criterion for path diagrams in terms of paths. For some of the results in this appendix, it is more convenient to define it in terms of routes. Recall that whereas all the nodes in a path must be different, the nodes in a route do not need to be so. Given a route $\rho_{X:Y}$ from a node X to a node Y in a path diagram, a node C is a collider in $\rho_{X:Y}$ if $A \leftrightarrow C \leftarrow B$ is a subroute of $\rho_{X:Y}$. Note that A and B may be the same node. Given a set of nodes Z, $\rho_{X:Y}$ is said to be Z-open if

- every collider in $\rho_{X:Y}$ is in Z, and
- ever non-collider in $\rho_{X:Y}$ is outside Z.

Note that there is a Z-open route from X to Y if and only if there is a Z-open path from X to Y (see Lemma 13 in Appendix E). When such a path or route exists, we say that X and Y are Z-connected.

Before presenting the results in this appendix, we define the operation of conditioning a path diagram on a node A as replacing every edge $A \rightarrow B$ with an edge $A_B \rightarrow B$, where A_B is a new node. Note that A is not removed. In terms of the associated system of linear equations, this implies (i) adding a new equation $A_B = \epsilon_{A_B}$ where ϵ_{A_B} is normally distributed with arbitrary mean and variance, and (ii) replacing every equation $B = \alpha^T(A, Pa(B) \setminus A) + \epsilon_B$ with an equation $B = \alpha^T(A_B, Pa(B) \setminus A) + \epsilon_B$. Note that, after conditioning, we have that $Ch(A) = \varnothing$ whereas $Pa(A_B) \cup Sp(A_B) = \varnothing$ and $Ch(A_B) = B$. See Figure 7 for an illustration. Let V denote all the nodes in the path diagram at hand, and consider the distribution $p(V \setminus A, A = a)$ defined by the system of equations before conditioning on A. This is the unnormalized conditional distribution of $V \setminus A$ given A. Let A' denote the new nodes created by conditioning on A, and consider the distribution $p(V \setminus A, A = a, A' = a)$ defined by the system of equations after conditioning on A. This is the unnormalized conditional distribution of $V \setminus A$ given $A \cup A'$. Note that both unnormalized conditional distributions coincide for all a, i.e. $p(V \setminus A = x, A = a) = p(V \setminus A = x, A = a, A' = a)$ for all x and a. Thus, their normalized versions coincide as well. So, computing partial covariances in either of them gives the same result, since partial covariances coincide with conditional covariances for Gaussian random vectors. In other words, the partial covariance $\sigma_{XY \cdot A}$ in the original path

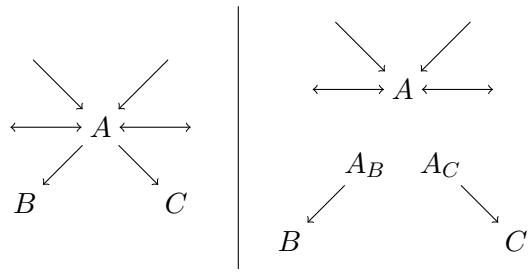

Figure 7: Conditioning the path diagram to the left on the node A results in the diagram to the right.

diagram is equal to $\sigma_{XY \cdot AA'}$ in the conditional diagram. Finally, we define conditioning on a set of nodes S as conditioning on each node in S. By the previous reasoning, the partial covariance $\sigma_{XY \cdot S}$ in the original path diagram is equal to $\sigma_{XY \cdot SS'}$ in the conditional diagram, where S' denotes the new nodes created by conditioning on S. The following theorems show how to compute the latter. See the Appendix E for the proofs. We illustrate the theorems through some examples afterwards.

Theorem 10 *Consider a path diagram conditioned on a set of nodes S. Let $Z = S \cup S'$. Let $\Pi_{X:Y}$ denote all the Z-open paths from X to Y. Suppose that no path in $\Pi_{X:Y}$ has colliders. Suppose that all the paths in $\Pi_{X:Y}$ have a subpath $X_m \leftarrow \cdots \leftarrow X_2 \leftarrow X_1 \rightarrow X_{m+1} \rightarrow \cdots \rightarrow X_{m+n}$ or $X_1 = X \rightarrow X_2 \rightarrow \cdots \rightarrow X_{m+n}$. Suppose that there is no Z-open route $X_i \rightarrow A \circ\!\!-\!\!\circ \cdots \circ\!\!-\!\!\circ B \circ\!\!\rightarrow X_i$ with $i > 1$. Moreover, let $Z_i^i = Z_i \cup Z^i$ and $Z_{1:a}^{1:b} = Z_1 \cup \cdots \cup Z_a \cup Z^1 \cup \cdots \cup Z^b$ where*

- *$Z^i = \{W_1, W_2, \ldots\}$ is a subset of $Z \setminus Z_{1:i-1}^{1:i-1}$ such that each W_j is $(Z_{1:i-1}^{1:i-1} \cup W_{1:j-1})$-connected to X_i through $Pa(X_i) \cup Sp(X_i)$ by a path that does not contain any node that is in some path in $\Pi_{X:Y}$ except X_i, and*
- *$Z_i = \{W_1, W_2, \ldots\}$ is a subset of $Z \setminus Z_{1:i-1}^{1:i}$ such that each W_j is $(Z_{1:i-1}^{1:i} \cup W_{1:j-1})$-connected to X_i through $Ch(X_i)$ by a path that does not contain any node that is in some path in $\Pi_{X:Y}$ except X_i.*

Finally, let $Z \setminus Z_{1:m+n}^{1:m+n} = \{W_1, W_2, \ldots\}$ and suppose that $X \perp W_j | Z_{1:m+n}^{1:m+n} \cup W_{1:j-1}$ or $Y \perp W_j | Z_{1:m+n}^{1:m+n} \cup W_{1:j-1}$. Then,

$$\sigma_{XY \cdot Z} = \sigma_{XY} \frac{\sigma^2_{X_1 \cdot Z_1^1}}{\sigma^2_{X_1}} \prod_{i=2}^{m+n} \frac{\sigma^2_{X_i \cdot Z_{1:i}^{1:i}}}{\sigma^2_{X_i \cdot Z_{1:i-1}^{1:i}}}.$$

Theorem 11 *Consider the same assumptions as in Theorem 10 with the only difference that all the paths in $\Pi_{X:Y}$ have now a subpath $X_m \leftarrow \cdots \leftarrow X_2 \leftarrow X_1 \leftrightarrow X_{m+1} \rightarrow \cdots \rightarrow X_{m+n}$ or $X_1 \leftrightarrow X_2 \rightarrow \cdots \rightarrow X_{m+n}$ or $\circ\!\!\rightarrow X_1 \rightarrow \cdots \rightarrow X_{m+n}$.[5] Then,*

$$\sigma_{XY \cdot Z} = \sigma_{XY} \prod_{i=1}^{m+n} \frac{\sigma^2_{X_i \cdot Z_{1:i}^{1:i}}}{\sigma^2_{X_i \cdot Z_{1:i-1}^{1:i}}}.$$

5. In the third subpath type, the predecessor of X_1 does not have to be the same in every path in $\Pi_{X:Y}$. It just has to reach X_1 through an edge $\rightarrow$ or $\leftrightarrow$ in every path in $\Pi_{X:Y}$.

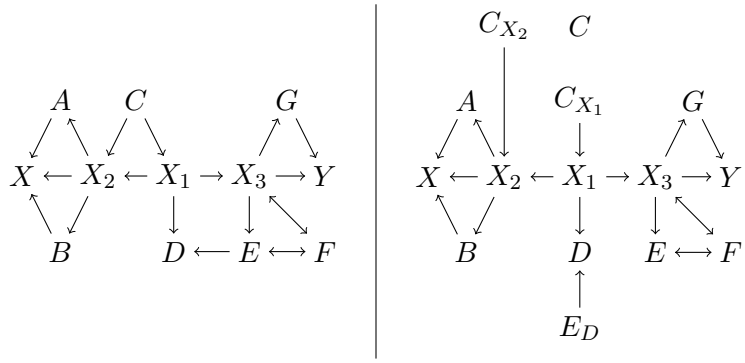

Figure 8: Left: Path diagram where Theorem 10 can be applied to compute $\sigma_{XY \cdot CDE}$. Right: The path diagram to the left conditioned on $\{C, D, E\}$.

Like Theorems 1 and 2, the two theorems above show that the partial covariance of two random variables can be computed by correcting their covariance with the product of some partial variance ratios. This implies that the partial covariance inherits the salient feature of factorizing over the nodes and edges in the paths between the two variables of interest. The two theorems above also imply that conditioning does not change the sign of the covariance, as stated in the following immediate corollary.

Corollary 12 *Suppose that two sets of nodes S_1 and S_2 satisfy the assumptions in Theorem 10 or 11. Then, $sign(\sigma_{XY}) = sign(\sigma_{XY \cdot S_1}) = sign(\sigma_{XY \cdot S_2})$.*

For the path diagrams that satisfy the conditions in Theorem 10 or 11, the corollary above implies that conditioning does not change the sign of the causal effect of X on Y, and that Simpson's paradox cannot occur.

We illustrate Theorem 10 with the following example.

Example 7 *Consider the path diagram to the left in Figure 8. Say that we want to compute $\sigma_{XY \cdot S}$ with $S = \{C, D, E\}$. The path diagram conditioned on S can be seen to the right in Figure 8. Then, $S' = \{C_{X_1}, C_{X_2}, E_D\}$ and $Z = S \cup S' = \{C, D, E, C_{X_1}, C_{X_2}, E_D\}$. As discussed before, $\sigma_{XY \cdot S}$ in the original diagram coincides with $\sigma_{XY \cdot Z}$ in the conditional diagram. Now, note that $Z^1 = \{C_{X_1}\}$, $Z_1 = \{D, E_D\}$, $Z^2 = \{C_{X_2}\}$, $Z_2 = \emptyset$, $Z^3 = \emptyset$, and $Z_3 = \{E\}$. Then, Theorem 10 gives*

$$\sigma_{XY \cdot Z} = \sigma_{XY} \frac{\sigma^2_{X_1 \cdot C_{X_1} D E_D}}{\sigma^2_{X_1}} \frac{\sigma^2_{X_2 \cdot C_{X_1} D E_D C_{X_2}}}{\sigma^2_{X_2 \cdot C_{X_1} D E_D C_{X_2}}} \frac{\sigma^2_{X_3 \cdot C_{X_1} D E_D C_{X_2} E}}{\sigma^2_{X_3 \cdot C_{X_1} D E_D C_{X_2}}}.$$

We illustrate Theorem 11 with the following example.

Example 8 *Consider the path diagram to the left in Figure 9. Say that we want to compute $\sigma_{XY \cdot S}$ with $S = \{C, D\}$. The path diagram conditioned on S can be seen to the right in Figure 9. Then, $S' = \{C_{X_1}, C_{X_2}, D_E\}$ and $Z = S \cup S' = \{C, D, C_{X_1}, C_{X_2}, D_E, D_Y\}$. As discussed before, $\sigma_{XY \cdot S}$ in*

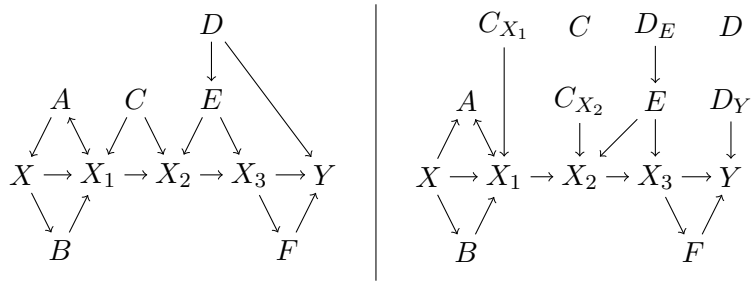

Figure 9: Left: Path diagram where Theorem 11 can be applied to compute $\sigma_{XY\cdot CD}$. Right: The path diagram to the left conditioned on $\{C, D\}$.

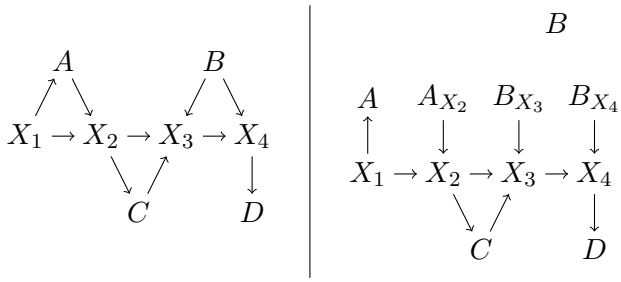

Figure 10: Left: Path diagram where Theorems 10 and 11 can be combined to compute $\sigma_{X_1X_4\cdot ABD}$. Left: The path diagram to the left conditioned on $\{A, B, D\}$.

the original diagram coincides with $\sigma_{XY\cdot Z}$ in the conditional diagram. Now, note that $Z^1 = \{C_{X_1}\}$, $Z^2 = \{C_{X_2}, D_E\}$, and $Z_1 = Z_2 = Z_3 = Z^3 = \varnothing$. Then, Theorem 11 gives

$$\sigma_{XY\cdot Z} = \sigma_{XY}\frac{\sigma^2_{X_1\cdot C_{X_1}}}{\sigma^2_{X_1\cdot C_{X_1}}}\frac{\sigma^2_{X_2\cdot C_{X_1}C_{X_2}D_E}}{\sigma^2_{X_2\cdot C_{X_1}C_{X_2}D_E}}\frac{\sigma^2_{X_3\cdot C_{X_1}C_{X_2}D_E}}{\sigma^2_{X_3\cdot C_{X_1}C_{X_2}D_E}}.$$

Theorems 10 and 11 can be extended to when all the paths from X to Y in the conditional path diagram share more than one subpath. For instance, consider the path diagram to the left in Figure 10. Say that we want to compute $\sigma_{X_1X_4\cdot S}$ with $S = \{A, B, D\}$. The path diagram conditioned on S can be seen to the right in Figure 10. Then, $S' = \{A_{X_2}, B_{X_3}, B_{X_4}\}$ and $Z = S \cup S' = \{A, B, D, A_{X_2}, B_{X_3}, B_{X_4}\}$. As discussed before, $\sigma_{X_1X_4\cdot S}$ in the original diagram coincides with $\sigma_{X_1X_4\cdot Z}$ in the conditional diagram. Now, note that $Z_1 = \{A\}$, $Z^2 = \{A_{X_2}\}$, $Z^3 = \{B_{X_3}\}$, $Z^4 = \{B_{X_4}\}$, $Z_4 = \{D\}$, and $Z^1 = Z_2 = Z_3 = \varnothing$. Note also that the conditional diagram has two Z-open paths from X_1 to X_4, which share two subpaths: $X_1 \to X_2$ and $\to X_3 \to X_4$. Therefore, neither Theorem 10 nor 11 applies. However, applying Theorem 10 followed by Theorem 11 gives

$$\sigma_{X_1X_4\cdot Z} = \sigma_{X_1X_4}\frac{\sigma^2_{X_1\cdot A}}{\sigma^2_{X_1}}\frac{\sigma^2_{X_2\cdot AA_{X_2}}}{\sigma^2_{X_2\cdot AA_{X_2}}}\frac{\sigma^2_{X_3\cdot AA_{X_2}B_{X_3}}}{\sigma^2_{X_3\cdot AA_{X_2}B_{X_3}}}\frac{\sigma^2_{X_4\cdot AA_{X_2}B_{X_3}B_{X_4}D}}{\sigma^2_{X_4\cdot AA_{X_2}B_{X_3}B_{X_4}}}. \tag{5}$$

The proof of correctness of the previous expression is simply a concatenation of the proofs of Theorems 10 and 11. We omit the details. An alternative way of answering the previous query is by first absorbing the subpath $X_2 \rightarrow C \rightarrow X_3$ into the subpath $X_2 \rightarrow X_3$. Now, there is only one shared subpath in the path diagram conditioned on S, namely $X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4$. Then, Theorem 10 gives Equation 5. This absorption trick is always possible when there are several shared subpaths. We omit the details.

Appendix E. Proofs of Appendix D

We start with some lemmas stating some auxiliary results. Recall from Footnote 5 that when we say that every path has a subpath $A \circ\!\!\rightarrow$, we do not mean that the successor of A is the same in every path. We mean that the successor is reached through an edge $\rightarrow$ or $\leftrightarrow$ in every path. Given a route $\rho_{X:Y}$ from X to Y, we let $\rho_{X:A}$ denote the subroute of $\rho_{X:Y}$ from X to A. Given two routes $\rho_{X:A}$ and $\rho_{A:Y}$, we let $\rho_{X:A} \cup \rho_{A:Y}$ denote the route from X to Y resulting from concatenating $\rho_{X:A}$ and $\rho_{A:Y}$. Finally, the path corresponding to a Z-open route from X to Y is a Z-open path from X to Y whose edges are a subset of the edges in the route. Such a path always exists by Lemma 13.

Lemma 13 *There is a Z-open route from X to Y if and only if there is a Z-open path from X to Y. Moreover, the path and the route can be chosen such that the edges in the former are a subset of the edges in the latter.*

Proof Let $\pi_{X:Y}$ be a Z-open path from X to Y. For every subpath $A \circ\!\!\rightarrow C \leftarrow\!\!\circ B$ of $\pi_{X:Y}$ such that $C \notin Z$, do the following. First, find a path $C \rightarrow X_1 \rightarrow \cdots \rightarrow X_n$ with $X_n \in Z$ and $X_1, \ldots, X_{n-1} \notin Z$. Such a path must exist for $\pi_{X:Y}$ to be Z-open. Second, replace $A \circ\!\!\rightarrow C \leftarrow\!\!\circ B$ with $A \circ\!\!\rightarrow C \rightarrow X_1 \rightarrow \cdots \rightarrow X_n \leftarrow \cdots \leftarrow X_1 \leftarrow C \leftarrow\!\!\circ B$. The result is the desired route.

Let $\rho_{X:Y}$ be a Z-open route from X to Y. Repeat the following two steps while possible. The result is the desired path. First, choose a node A that occurs several times in $\rho_{X:Y}$. Let A_1 and A_2 denote the first and the last occurrences of A in $\rho_{X:Y}$. Assume without loss of generality that no node in $\rho_{X:A_1}$ occurs in $\rho_{A_2:Y}$ except A. Second, consider the following cases.

- If $\rho_{X:Y}$ is $X \circ\!\!-\!\!\circ \cdots \leftarrow A_1 \circ\!\!-\!\!\circ \cdots \circ\!\!-\!\!\circ A_2 \circ\!\!-\!\!\circ \cdots \circ\!\!-\!\!\circ Y$, then replace it with $\rho_{X:A_1} \cup \rho_{A_2:Y}$.
- If $\rho_{X:Y}$ is $X \circ\!\!-\!\!\circ \cdots \circ\!\!\rightarrow A_1 \circ\!\!-\!\!\circ \cdots \circ\!\!-\!\!\circ A_2 \rightarrow \cdots \circ\!\!-\!\!\circ Y$, then replace it with $\rho_{X:A_1} \cup \rho_{A_2:Y}$.
- If $\rho_{X:Y}$ is $X \circ\!\!-\!\!\circ \cdots \circ\!\!\rightarrow A_1 \circ\!\!-\!\!\circ \cdots \circ\!\!-\!\!\circ A_2 \leftarrow\!\!\circ \cdots \circ\!\!-\!\!\circ Y$, then replace it with $\rho_{X:A_1} \cup \rho_{A_2:Y}$. Note that A or some descendant of A must be in Z for the original $\rho_{X:Y}$ to be Z-open.

■

Lemma 14 *Consider a path diagram. Let X, Y, R and W be nodes and Z a set of nodes. If $X \perp W | Z \cup R$ and $Y \perp W | Z \cup R$ and $X \perp Y | Z \cup R$, then*

$$\sigma_{XY \cdot ZW} = \sigma_{XY \cdot Z} \frac{\sigma^2_{R \cdot ZW}}{\sigma^2_{R \cdot Z}}.$$

Proof Note that $X \perp W | Z \cup R$ implies that

$$0 = \sigma_{XW \cdot ZR} = \sigma_{XW \cdot Z} - \frac{\sigma_{XR \cdot Z}\sigma_{RW \cdot Z}}{\sigma^2_{R \cdot Z}}$$

which implies that $\sigma_{XW \cdot Z} = \delta_{XR \cdot Z}\sigma_{RW \cdot Z}$ where $\delta_{XR \cdot Z} = \sigma_{XR \cdot Z}/\sigma^2_{R \cdot Z}$. Likewise, $Y \perp W | Z \cup R$ implies that $\sigma_{YW \cdot Z} = \delta_{YR \cdot Z}\sigma_{RW \cdot Z}$ where $\delta_{YR \cdot Z} = \sigma_{YR \cdot Z}/\sigma^2_{R \cdot Z}$. Likewise, $X \perp Y | Z \cup R$ implies that $\sigma_{XY \cdot Z} = \delta_{XR \cdot Z}\delta_{YR \cdot Z}\sigma^2_{R \cdot Z}$. Therefore,

$$\begin{aligned}
\sigma_{XY \cdot ZW} &= \sigma_{XY \cdot Z} - \frac{\sigma_{XW \cdot Z}\sigma_{WY \cdot Z}}{\sigma^2_{W \cdot Z}} \\
&= \delta_{XR \cdot Z}\delta_{YR \cdot Z}\sigma^2_{R \cdot Z} - \frac{\delta_{XR \cdot Z}\sigma_{RW \cdot Z}\delta_{YR \cdot Z}\sigma_{RW \cdot Z}}{\sigma^2_{W \cdot Z}} \\
&= \delta_{XR \cdot Z}\delta_{YR \cdot Z}\Big(\sigma^2_{R \cdot Z} - \frac{\sigma_{RW \cdot Z}\sigma_{RW \cdot Z}}{\sigma^2_{W \cdot Z}}\Big) \\
&= \delta_{XR \cdot Z}\delta_{YR \cdot Z}\sigma^2_{R \cdot ZW} = \sigma_{XY \cdot Z}\frac{\sigma^2_{R \cdot ZW}}{\sigma^2_{R \cdot Z}}.
\end{aligned}$$

■

Lemma 15 *Consider a path diagram. Let X, Y and W be nodes and Z a set of nodes. If $Y \perp W | Z \cup X$, then*

$$\sigma_{XY \cdot ZW} = \sigma_{XY \cdot Z}\frac{\sigma^2_{X \cdot ZW}}{\sigma^2_{X \cdot Z}}.$$

Proof Note that $Y \perp W | Z \cup X$ implies that

$$0 = \sigma_{YW \cdot ZX} = \sigma_{YW \cdot Z} - \frac{\sigma_{YX \cdot Z}\sigma_{XW \cdot Z}}{\sigma^2_{X \cdot Z}}$$

which implies that

$$\sigma_{YW \cdot Z} = \frac{\sigma_{YX \cdot Z}\sigma_{XW \cdot Z}}{\sigma^2_{X \cdot Z}}.$$

Therefore,

$$\begin{aligned}
\sigma_{XY \cdot ZW} &= \sigma_{XY \cdot Z} - \frac{\sigma_{XW \cdot Z}\sigma_{WY \cdot Z}}{\sigma^2_{W \cdot Z}} = \sigma_{XY \cdot Z} - \frac{\sigma_{XW \cdot Z}\sigma_{YX \cdot Z}\sigma_{XW \cdot Z}}{\sigma^2_{W \cdot Z}\sigma^2_{X \cdot Z}} \\
&= \sigma_{XY \cdot Z}\Big(1 - \frac{\sigma_{XW \cdot Z}\sigma_{XW \cdot Z}}{\sigma^2_{W \cdot Z}\sigma^2_{X \cdot Z}}\Big) = \frac{\sigma_{XY \cdot Z}}{\sigma^2_{X \cdot Z}}\Big(\sigma^2_{X \cdot Z} - \frac{\sigma_{XW \cdot Z}\sigma_{XW \cdot Z}}{\sigma^2_{W \cdot Z}}\Big) \\
&= \sigma_{XY \cdot Z}\frac{\sigma^2_{X \cdot ZW}}{\sigma^2_{X \cdot Z}}.
\end{aligned}$$

■

Lemma 16 *Consider a path diagram. Let X, Y and W be nodes and Z a set of nodes. If $X \perp W | Z$ or $Y \perp W | Z$, then $\sigma_{XY \cdot ZW} = \sigma_{XY \cdot Z}$.*

Proof Note that $X \perp W|Z$ and $Y \perp W|Z$ imply $\sigma_{XW \cdot Z} = 0$ and $\sigma_{YW \cdot Z} = 0$, respectively. Thus, if $X \perp W|Z$ or $Y \perp W|Z$ then

$$\sigma_{XY \cdot ZW} = \sigma_{XY \cdot Z} - \frac{\sigma_{XW \cdot Z}\sigma_{WY \cdot Z}}{\sigma^2_{W \cdot Z}} = \sigma_{XY \cdot Z}.$$

■

Lemma 17 *Consider a path diagram. Let $\Pi_{X:Y}$ denote all the Z-open paths from X to Y. Suppose that no path in $\Pi_{X:Y}$ has colliders. Suppose that all the paths in $\Pi_{X:Y}$ have a subpath $\leftarrow R \rightarrow$ or $R = X \rightarrow$. Let W be a node that is Z-connected to R by a path that does not contain any node that is in some path in $\Pi_{X:Y}$ except R. If $Pa(W) \cup Sp(W) = \emptyset$, then*

$$\sigma_{XY \cdot ZW} = \sigma_{XY \cdot Z}\frac{\sigma^2_{R \cdot ZW}}{\sigma^2_{R \cdot Z}}.$$

Proof Consider first the case where $\leftarrow R \rightarrow$ is a subpath of every path in $\Pi_{X:Y}$. Assume to the contrary that $X \not\perp W|Z \cup R$ and let $\rho_{X:W}$ be a $(Z \cup R)$-open **route**. Follow $\rho_{X:W}$ until reaching R or W, and let $\pi_{X:Y} \in \Pi_{X:Y}$.

- If R is reached first, then consider the first occurrence of R in $\rho_{X:W}$ and note that $\rho_{X:R} \cup \pi_{R:Y}$ is a Z-open route. Moreover, it does not contain any edge $\leftarrow R$ that is in some path in $\Pi_{X:Y}$ because, otherwise, $\rho_{X:W}$ contains the edge ($\pi_{R:Y}$ cannot by definition) and, thus, it is not $(Z \cup R)$-open. Then, the path corresponding to $\rho_{X:R} \cup \pi_{R:Y}$ contradicts the assumptions in the lemma.

- If W is reached first, then let $\varrho_{R:W}$ denote a Z-open path that does not contain any node that is in some path in $\Pi_{X:Y}$ except R. Such a path exists by the assumptions in the lemma. Thus, $\rho_{X:W} \cup \varrho_{W:R} \cup \pi_{R:Y}$ is a Z-open route, because neither R nor W is a collider in it. The latter follows from the assumption that $Pa(W) \cup Sp(W) = \emptyset$. Moreover, the route does not contain any edge $\leftarrow R$ that is in some path in $\Pi_{X:Y}$ because, otherwise, $\rho_{X:W}$ contains the edge ($\varrho_{W:R}$ and $\pi_{R:Y}$ cannot by definition) and thus it reaches R first. Then, the path corresponding to $\rho_{X:W} \cup \varrho_{W:R} \cup \pi_{R:Y}$ contradicts the assumptions in the lemma.

Consequently, $X \perp W|Z \cup R$. We can analogously prove that $Y \perp W|Z \cup R$. Now, assume to the contrary that $X \not\perp Y|Z \cup R$ and let $\rho_{X:Y}$ be a $(Z \cup R)$-open **route**. Note that R must be in $\rho_{X:Y}$ because, otherwise, $\rho_{X:Y}$ is Z-open and, thus, its corresponding path contradicts the assumptions in the lemma. Consider the first occurrence of R in $\rho_{X:Y}$, and let $\pi_{X:Y} \in \Pi_{X:Y}$. Then, $\rho_{X:R} \cup \pi_{R:Y}$ is a Z-open route. Moreover, the route does not contain any edge $\leftarrow R$ that is in some path in $\Pi_{X:Y}$ because, otherwise, $\rho_{X:Y}$ contains the edge ($\pi_{R:Y}$ cannot by definition) and thus it is not $(Z \cup R)$-open. Then, the path corresponding to $\rho_{X:R} \cup \pi_{R:Y}$ contradicts the assumptions in the lemma. Consequently, $X \perp Y|Z \cup R$. Therefore, the desired result follows from Lemma 14.

Finally, consider the case where $R = X \rightarrow$ is a subpath of every path in $\Pi_{X:Y}$. Assume to the contrary that $Y \not\perp W|Z \cup X$ and let $\rho_{Y:W}$ be a $(Z \cup X)$-open **route**. Follow $\rho_{Y:W}$ until reaching X or W.

- If X is reached first, then note that $\rho_{Y:X}$ does not contain any edge $\leftarrow X$ that is in some path in $\Pi_{Y:X}$ because, otherwise, $\rho_{Y:W}$ is not $(Z \cup X)$-open. Then, the path corresponding to $\rho_{X:Y}$ contradicts the assumptions in the lemma.

- If W is reached first, let $\varrho_{X:W}$ denote a Z-open path that does not contain any node that is in some path in $\Pi_{X:Y}$ except X. Such a path exists by the assumptions in the lemma. Then, $\varrho_{X:W} \cup \rho_{W:Y}$ is a Z-open route, because W is not a collider in it due to the assumption that $Pa(W) \cup Sp(W) = \varnothing$. Moreover, the route does not contain any edge $X \rightarrow$ that is in some path in $\Pi_{X:Y}$ because, otherwise, $\rho_{W:Y}$ contains the edge ($\varrho_{X:W}$ cannot by definition) and thus $\rho_{Y:W}$ reaches X first. Then, the path corresponding to $\varrho_{X:W} \cup \rho_{W:Y}$ contradicts the assumptions in the lemma.

Consequently, $Y \perp W | Z \cup X$ and, thus, the desired result follows from Lemma 15. ■

Lemma 18 *Consider a path diagram. Let $\Pi_{X:Y}$ denote all the Z-open paths from X to Y. Suppose that no path in $\Pi_{X:Y}$ has colliders. Suppose that all the paths in $\Pi_{X:Y}$ have a subpath $\leftarrow R \rightarrow$ or $R = X \rightarrow$. Let W be a node that is Z-connected to R by a path that does not contain any node that is in some path in $\Pi_{X:Y}$ except R. If $Ch(W) = \varnothing$ and $\Pi_{X:Y}$ are all the $(Z \cup W)$-open paths from X to Y, then*

$$\sigma_{XY \cdot ZW} = \sigma_{XY \cdot Z} \frac{\sigma^2_{R \cdot ZW}}{\sigma^2_{R \cdot Z}}.$$

Proof Consider first the case where $\leftarrow R \rightarrow$ is a subpath of every path in $\Pi_{X:Y}$. Assume to the contrary that $X \not\perp W | Z \cup R$ and let $\rho_{X:W}$ be a $(Z \cup R)$-open **route**. Follow $\rho_{X:W}$ until reaching R or W, and let $\pi_{X:Y} \in \Pi_{X:Y}$.

- If R is reached first, then consider the first occurrence of R in $\rho_{X:W}$ and note that $\rho_{X:R} \cup \pi_{R:Y}$ is a Z-open route. Moreover, it does not contain any edge $\leftarrow R$ that is in some path in $\Pi_{X:Y}$ because, otherwise, $\rho_{X:W}$ contains the edge ($\pi_{R:Y}$ cannot by definition) and, thus, it is not $(Z \cup R)$-open. Then, the path corresponding to $\rho_{X:R} \cup \pi_{R:Y}$ contradicts the assumptions in the lemma.

- If W is reached first, then let $\varrho_{R:W}$ denote a Z-open path that does not contain any node that is in some path in $\Pi_{X:Y}$ except R. Such a path exists by the assumptions in the lemma. Thus, $\rho_{X:W} \cup \varrho_{W:R} \cup \pi_{R:Y}$ is a $(Z \cup W)$-open route from X to Y, because W is a collider in it whereas R is not. The former follows from the assumption that $Ch(W) = \varnothing$. Moreover, the route does not contain any edge $\leftarrow R$ that is in some path in $\Pi_{X:Y}$ because, otherwise, $\rho_{X:W}$ contains the edge ($\varrho_{W:R}$ and $\pi_{R:Y}$ cannot by definition) and thus it reaches R first. Then, the path corresponding to $\rho_{X:W} \cup \varrho_{W:R} \cup \pi_{R:Y}$ contradicts the assumptions in the lemma.

Consequently, $X \perp W | Z \cup R$. We can analogously prove that $Y \perp W | Z \cup R$. Now, assume to the contrary that $X \not\perp Y | Z \cup R$ and let $\rho_{X:Y}$ be a $(Z \cup R)$-open **route**. Note that R must be in $\rho_{X:Y}$ because, otherwise, $\rho_{X:Y}$ is Z-open and, thus, its corresponding path contradicts the assumptions in the lemma. Consider the first occurrence of R in $\rho_{X:Y}$, and let $\pi_{X:Y} \in \Pi_{X:Y}$. Then, $\rho_{X:R} \cup \pi_{R:Y}$ is a Z-open route. Moreover, the route does not contain any edge $\leftarrow R$ that is in some path in $\Pi_{X:Y}$ because, otherwise, $\rho_{X:Y}$ contains the edge ($\pi_{R:Y}$ cannot by definition) and thus it is not $(Z \cup R)$-open. Then, the path corresponding to $\rho_{X:R} \cup \pi_{R:Y}$ contradicts the assumptions in the lemma. Consequently, $X \perp Y | Z \cup R$. Therefore, the desired result follows from Lemma 14.

Finally, consider the case where $R = X \rightarrow$ is a subpath of every path in $\Pi_{X:Y}$. Assume to the contrary that $Y \not\perp W | Z \cup X$ and let $\rho_{Y:W}$ be a $(Z \cup X)$-open **route**. Follow $\rho_{Y:W}$ until reaching X or W.

- If X is reached first, then note that $\rho_{Y:X}$ does not contain any edge $\leftarrow X$ that is in some path in $\Pi_{Y:X}$ because, otherwise, $\rho_{Y:W}$ is not $(Z \cup X)$-open. Then, the path corresponding to $\rho_{X:Y}$ contradicts the assumptions in the lemma.

- If W is reached first, let $\varrho_{X:W}$ denote a Z-open path that does not contain any node that is in some path in $\Pi_{X:Y}$ except X. Such a path exists by the assumptions in the lemma. Then, $\varrho_{X:W} \cup \rho_{W:Y}$ is a $(Z \cup W)$-open route, because W is a collider in it due to the assumption that $Ch(W) = \emptyset$. Moreover, the route does not contain any edge $X \rightarrow$ that is in some path in $\Pi_{X:Y}$ because, otherwise, $\rho_{W:Y}$ contains the edge ($\varrho_{X:W}$ cannot by definition) and thus $\rho_{Y:W}$ reaches X first. Then, the path corresponding to $\varrho_{X:W} \cup \rho_{W:Y}$ contradicts the assumptions in the lemma.

Consequently, $Y \perp W|Z \cup X$ and, thus, the desired result follows from Lemma 15. ■

Lemma 19 *Consider a path diagram. Let $\Pi_{X:Y}$ denote all the Z-open paths from X to Y. Suppose that no path in $\Pi_{X:Y}$ has colliders. Suppose that all the paths in $\Pi_{X:Y}$ have a subpath $\circ\!\!\rightarrow R \rightarrow$ or $\circ\!\!\rightarrow Y = R$. Let W be a node that is Z-connected to R through $Pa(R) \cup Sp(R)$ by a path that does not contain any node that is in some path in $\Pi_{X:Y}$ except R. If $Pa(W) \cup Sp(W) = \emptyset$ and W is not Z-connected to R through $Ch(R)$, then*

$$\sigma_{XY \cdot ZW} = \sigma_{XY \cdot Z}.$$

Proof Consider first the case where $\circ\!\!\rightarrow R \rightarrow$ is a subpath of every path in $\Pi_{X:Y}$. Assume to the contrary that $X \not\perp W|Z$ and let $\rho_{X:W}$ be a Z-open **route**. Follow $\rho_{X:W}$ until reaching R or W, and let $\pi_{X:Y} \in \Pi_{X:Y}$.

- If R is reached first, then consider the first occurrence of R in $\rho_{X:W}$ and note that $\rho_{X:R}$ ends with an edge $\circ\!\!\rightarrow R$ because, otherwise, $\rho_{X:R} \cup \pi_{R:Y}$ is a Z-open route that has a subroute $\leftarrow R \rightarrow$ and, thus, its corresponding path contradicts the assumptions in the lemma. Then, $\rho_{R:W}$ must start with an edge $R \rightarrow$ for $\rho_{X:W}$ to be Z-open. However, this contradicts the assumption that W is not Z-connected to R through $Ch(R)$.

- If W is reached first, then let $\varrho_{R:W}$ denote a Z-open path that does not contain any node that is in some path in $\Pi_{X:Y}$ except R. Such a path exists by the assumptions in the lemma. Thus, $\rho_{X:W} \cup \varrho_{W:R} \cup \pi_{R:Y}$ is a Z-open route, because neither R nor W is a collider in it. The latter follows from the assumption that $Pa(W) \cup Sp(W) = \emptyset$. Moreover, the route does not contain any edge $\circ\!\!\rightarrow R$ that is in some path in $\Pi_{X:Y}$ because, otherwise, $\rho_{X:W}$ contains the edge ($\varrho_{W:R}$ and $\pi_{R:Y}$ cannot by definition) and thus it reaches R first. Then, the path corresponding to $\rho_{X:W} \cup \varrho_{W:R} \cup \pi_{R:Y}$ contradicts the assumptions in the lemma.

Consequently, $X \perp W|Z$. When $\circ\!\!\rightarrow Y = R$ is a subpath of every path in $\Pi_{X:Y}$, we can prove that $X \perp W|Z$ much in the same way. Consequently, $X \perp W|Z$ in either case and, thus, the desired result follows from Lemma 16. ■

Lemma 20 *Consider a path diagram. Let $\Pi_{X:Y}$ denote all the Z-open paths from X to Y. Suppose that no path in $\Pi_{X:Y}$ has colliders. Suppose that all the paths in $\Pi_{X:Y}$ have a subpath $\circ\!\!\rightarrow R \rightarrow$*

or $\circ\!\!\rightarrow Y = R$. *Let* W *be a node that is* Z*-connected to* R *through* $Pa(R) \cup Sp(R)$ *by a path that does not contain any node that is in some path in* $\Pi_{X:Y}$ *except* R. *If* $Ch(W) = \varnothing$, *and* W *is not* Z*-connected to* R *through* $Ch(R)$, *and* $\Pi_{X:Y}$ *are all the* $(Z \cup W)$*-open paths from* X *to* Y, *then*

$$\sigma_{XY \cdot ZW} = \sigma_{XY \cdot Z}.$$

Proof Consider first the case where $\circ\!\!\rightarrow R \rightarrow$ is a subpath of every path in $\Pi_{X:Y}$. Assume to the contrary that $X \not\perp W|Z$ and let $\rho_{X:W}$ be a Z-open **route**. Follow $\rho_{X:W}$ until reaching R or W, and let $\pi_{X:Y} \in \Pi_{X:Y}$.

- If R is reached first, then consider the first occurrence of R in $\rho_{X:W}$ and note that $\rho_{X:R}$ ends with an edge $\circ\!\!\rightarrow R$ because, otherwise, $\rho_{X:R} \cup \pi_{R:Y}$ is a Z-open route that has a subroute $\leftarrow R \rightarrow$ and, thus, its corresponding path contradicts the assumptions in the lemma. Then, $\rho_{R:W}$ must start with an edge $R \rightarrow$ for $\rho_{X:W}$ to be Z-open. However, this contradicts the assumption that W is not Z-connected to R through $Ch(R)$.

- If W is reached first, then let $\varrho_{R:W}$ denote a Z-open path that does not contain any node that is in some path in $\Pi_{X:Y}$ except R. Such a path exists by the assumptions in the lemma. Thus, $\rho_{X:W} \cup \varrho_{W:R} \cup \pi_{R:Y}$ is a $(Z \cup W)$-open route, because W is a collider in it whereas R is not. The former follows from the assumption that $Ch(W) = \varnothing$. Moreover, the route does not contain any edge $\circ\!\!\rightarrow R$ that is in some path in $\Pi_{X:Y}$ because, otherwise, $\rho_{X:W}$ contains the edge ($\varrho_{W:R}$ and $\pi_{R:Y}$ cannot by definition) and thus it reaches R first. Then, the path corresponding to $\rho_{X:W} \cup \varrho_{W:R} \cup \pi_{R:Y}$ contradicts the assumptions in the lemma.

Consequently, $X \perp W|Z$. When $\circ\!\!\rightarrow Y = R$ is a subpath of every path in $\Pi_{X:Y}$, we can prove that $X \perp W|Z$ much in the same way. Consequently, $X \perp W|Z$ in either case and, thus, the desired result follows from Lemma 16. ■

Lemma 21 *Consider a path diagram. Let* $\Pi_{X:Y}$ *denote all the* Z*-open paths from* X *to* Y. *Suppose that no path in* $\Pi_{X:Y}$ *has colliders. Suppose that all the paths in* $\Pi_{X:Y}$ *have a subpath* $\circ\!\!\rightarrow R \rightarrow$ *or* $\circ\!\!\rightarrow Y = R$. *Let* W *be a node that is* Z*-connected to* R *through* $Ch(R)$ *by a path that does not contain any node that is in some path in* $\Pi_{X:Y}$ *except* R. *If* $Pa(W) \cup Sp(W) = \varnothing$ *and* W *is not* Z*-connected to* R *through* $Pa(R) \cup Sp(R)$, *then*

$$\sigma_{XY \cdot ZW} = \sigma_{XY \cdot Z} \frac{\sigma^2_{R \cdot ZW}}{\sigma^2_{R \cdot Z}}.$$

Proof Consider first the case where $\circ\!\!\rightarrow R \rightarrow$ is a subpath of every path in $\Pi_{X:Y}$. Assume to the contrary that $X \not\perp W|Z \cup R$ and let $\rho_{X:W}$ be a $(Z \cup R)$-open **route**. Follow $\rho_{X:W}$ until reaching R or W, and let $\pi_{X:Y} \in \Pi_{X:Y}$.

- If R is reached first, then note R must be a collider in $\rho_{X:W}$ for this to be $(Z \cup R)$-open. However, the last occurrence of R in $\rho_{X:W}$ contradicts the assumption that W is not Z-connected to R through $Pa(R) \cup Sp(R)$.

- If W is reached first, then let $\varrho_{R:W}$ denote a Z-open path that does not contain any node that is in some path in $\Pi_{X:Y}$ except R. Such a path exists by the assumptions in the lemma.

Thus, $\rho_{X:W} \cup \varrho_{W:R} \cup \pi_{R:Y}$ is a Z-open route, because neither R nor W is a collider in it. The latter follows from the assumption that $Pa(W) \cup Sp(W) = \varnothing$. Moreover, the route does not contain any edge $\circ\!\!\rightarrow R$ that is in some path in $\Pi_{X:Y}$ because, otherwise, $\rho_{X:W}$ contains the edge ($\varrho_{W:R}$ and $\pi_{R:Y}$ cannot by definition) and thus it reaches R first. Then, the path corresponding to $\rho_{X:W} \cup \varrho_{W:R} \cup \pi_{R:Y}$ contradicts the assumptions in the lemma.

Consequently, $X \perp W | Z \cup R$. Now, assume to the contrary that $Y \not\perp W | Z \cup R$ and let $\rho_{Y:W}$ be a $(Z \cup R)$-open **route**. Follow $\rho_{Y:W}$ until reaching R or W, and let $\pi_{X:Y} \in \Pi_{X:Y}$.

- If R is reached first, then note R must be a collider in $\rho_{Y:W}$ for this be $(Z \cup R)$-open. However, the last occurrence of R in $\rho_{Y:W}$ contradicts the assumption that W is not Z-connected to R through $Pa(R) \cup Sp(R)$.

- If W is reached first, then let $\varrho_{R:W}$ denote a Z-open path that leaves R through $Ch(R)$ and that does not contain any node that is in some path in $\Pi_{X:Y}$ except R. Such a path exists by the assumptions in the lemma. Thus, $\pi_{X:R} \cup \varrho_{R:W} \cup \rho_{W:Y}$ is a Z-open route, because neither R nor W is a collider in it. The latter follows from the assumption that $Pa(W) \cup Sp(W) = \varnothing$. Moreover, the route does not contain any edge $R \rightarrow$ that is in some path in $\Pi_{X:Y}$ because, otherwise, $\rho_{W:Y}$ contains the edge ($\pi_{X:R}$ and $\varrho_{R:W}$ cannot by definition) and thus $\rho_{Y:W}$ reaches R first. Then, the path corresponding to $\pi_{X:R} \cup \varrho_{R:W} \cup \rho_{W:Y}$ contradicts the assumptions in the lemma.

Consequently, $Y \perp W | Z \cup R$. Now, assume to the contrary that $X \not\perp Y | Z \cup R$ and let $\rho_{X:Y}$ be a $(Z \cup R)$-open **path**. Note that R must be a collider or a descendant of a collider in $\rho_{X:Y}$ because, otherwise, R is not in $\rho_{X:Y}$ and, thus, $\rho_{X:Y}$ is Z-open, which contradicts the assumptions in the lemma. Note also that the assumption that W is Z-connected to R through $Ch(R)$ implies that some descendant of R is in $Z \cup W$. Actually, some descendant of R must be in Z due to the assumption that $Pa(W) \cup Sp(W) = \varnothing$. Then, $\rho_{X:Y}$ is Z-open, which contradicts the assumptions in the lemma. Consequently, $X \perp Y | Z \cup R$. Therefore, the desired result follows from Lemma 14.

Finally, consider the case where $\circ\!\!\rightarrow Y = R$ is a subpath of every path in $\Pi_{X:Y}$. Assume to the contrary that $X \not\perp W | Z \cup Y$ and let $\rho_{X:W}$ be a $(Z \cup Y)$-open **route**. Follow $\rho_{X:W}$ until reaching Y or W.

- If Y is reached first, then note Y must be a collider in $\rho_{X:W}$ for this to be $(Z \cup Y)$-open. However, the last occurrence of Y in $\rho_{X:W}$ contradicts the assumption that W is not Z-connected to R through $Pa(Y) \cup Sp(Y)$.

- If W is reached first, let $\varrho_{Y:W}$ denote a Z-open path that does not contain any node that is in some path in $\Pi_{X:Y}$ except Y. Such a path exists by the assumptions in the lemma. Then, $\varrho_{X:W} \cup \rho_{W:Y}$ is a Z-open route, because W is not a collider in it due to the assumption that $Pa(W) \cup Sp(W) = \varnothing$. Moreover, the route does not contain any edge $\circ\!\!\rightarrow Y$ that is in some path in $\Pi_{X:Y}$ because, otherwise, $\rho_{X:W}$ contains the edge ($\varrho_{W:Y}$ cannot by definition) and thus it reaches Y first. Then, the path corresponding to $\varrho_{X:W} \cup \rho_{W:Y}$ contradicts the assumptions in the lemma.

Consequently, $X \perp W | Z \cup Y$ and, thus, the desired result follows from Lemma 15. ■

Lemma 22 *Consider a path diagram. Let $\Pi_{X:Y}$ denote all the Z-open paths from X to Y. Suppose that no path in $\Pi_{X:Y}$ has colliders. Suppose that all the paths in $\Pi_{X:Y}$ have a subpath $\circ\!\!\rightarrow R \rightarrow$ or $\circ\!\!\rightarrow Y = R$. Let W be a node that is Z-connected to R through $Ch(R)$ by a path that does not contain any node that is in some path in $\Pi_{X:Y}$ except R. If $Ch(W) = \varnothing$, and W is not Z-connected to R by any path that reaches R through $Pa(R) \cup Sp(R)$, and $\Pi_{X:Y}$ are all the $(Z \cup W)$-open paths from X to Y, then*

$$\sigma_{XY \cdot ZW} = \sigma_{XY \cdot Z} \frac{\sigma^2_{R \cdot ZW}}{\sigma^2_{R \cdot Z}}.$$

Proof Consider first the case where $\circ\!\!\rightarrow R \rightarrow$ is a subpath of every path in $\Pi_{X:Y}$. Assume to the contrary that $X \not\perp W | Z \cup R$ and let $\rho_{X:W}$ be a $(Z \cup R)$-open **route**. Follow $\rho_{X:W}$ until reaching R or W, and let $\pi_{X:Y} \in \Pi_{X:Y}$.

- If R is reached first, then note that R must be a collider in $\rho_{X:W}$ for this to be $(Z \cup R)$-open. However, the last occurrence of R in $\rho_{X:W}$ contradicts the assumption that W is not Z-connected to R through $Pa(R) \cup Sp(R)$.

- If W is reached first, then let $\varrho_{R:W}$ denote a Z-open path that does not contain any node that is in some path in $\Pi_{X:Y}$ except R. Such a path exists by the assumptions in the lemma. Thus, $\rho_{X:W} \cup \varrho_{W:R} \cup \pi_{R:Y}$ is a $(Z \cup W)$-open route, because W is a collider in it whereas R is not. The former follows from the assumption that $Ch(W) = \varnothing$. Moreover, the route does not contain any edge $\circ\!\!\rightarrow R$ that is in some path in $\Pi_{X:Y}$ because, otherwise, $\rho_{X:W}$ contains the edge ($\varrho_{W:R}$ and $\pi_{R:Y}$ cannot by definition) and thus it reaches R first. Then, the path corresponding to $\rho_{X:W} \cup \varrho_{W:R} \cup \pi_{R:Y}$ contradicts the assumptions in the lemma.

Consequently, $X \perp W | Z \cup R$. Now, assume to the contrary that $Y \not\perp W | Z \cup R$ and let $\rho_{Y:W}$ be a $(Z \cup R)$-open **route**. Follow $\rho_{Y:W}$ until reaching R or W, and let $\pi_{X:Y} \in \Pi_{X:Y}$.

- If R is reached first, then note that R must be a collider in $\rho_{Y:W}$ for this to be $(Z \cup R)$-open. However, the last occurrence of R in $\rho_{Y:W}$ contradicts the assumption that W is not Z-connected to R through $Pa(R) \cup Sp(R)$.

- If W is reached first, then let $\varrho_{R:W}$ denote a Z-open path that leaves R through $Ch(R)$ and that does not contain any node that is in some path in $\Pi_{X:Y}$ except R. Such a path exists by the assumptions in the lemma. Thus, $\pi_{X:R} \cup \varrho_{R:W} \cup \rho_{W:Y}$ is a $(Z \cup W)$-open route, because W is a collider in it whereas R is not. The former follows from the assumption that $Ch(W) = \varnothing$. Moreover, the route does not contain any edge $R \rightarrow$ that is in some path in $\Pi_{X:Y}$ because, otherwise, $\rho_{W:Y}$ contains the edge ($\pi_{X:R}$ and $\varrho_{R:W}$ cannot by definition) and thus $\rho_{Y:W}$ reaches R first. Then, the path corresponding to $\pi_{X:R} \cup \varrho_{R:W} \cup \rho_{W:Y}$ contradicts the assumptions in the lemma.

Consequently, $Y \perp W | Z \cup R$. Now, assume to the contrary that $X \not\perp Y | Z \cup R$ and let $\rho_{X:Y}$ be a $(Z \cup R)$-open **path**. Note that R must be a collider or a descendant of a collider in $\rho_{X:Y}$ because, otherwise, R is not in $\rho_{X:Y}$ and, thus, $\rho_{X:Y}$ is Z-open, which contradicts the assumptions in the lemma. Note also that the assumption that W is Z-connected to R through $Ch(R)$ implies that some descendant of R is in $Z \cup W$. Then, $\rho_{X:Y}$ is $(Z \cup W)$-open, which contradicts the assumptions in the lemma. Consequently, $X \perp Y | Z \cup R$. Therefore, the desired result follows from Lemma 14.

Finally, consider the case where $\circ\!\!\rightarrow Y = R$ is a subpath of every path in $\Pi_{X:Y}$. Assume to the contrary that $X \not\perp W|Z \cup Y$ and let $\rho_{X:W}$ be a $(Z \cup Y)$-open **route**. Follow $\rho_{X:W}$ until reaching Y or W.

- If Y is reached first, then note Y must be a collider in $\rho_{X:W}$ for this to be $(Z \cup Y)$-open. However, the last occurrence of Y in $\rho_{X:W}$ contradicts the assumption that W is not Z-connected to R through $Pa(Y) \cup Sp(Y)$.
- If W is reached first, let $\varrho_{Y:W}$ denote a Z-open path that does not contain any node that is in some path in $\Pi_{X:Y}$ except Y. Such a path exists by the assumptions in the lemma. Then, $\varrho_{X:W} \cup \rho_{W:Y}$ is a $(Z \cup W)$-open route, because W is a collider in it due to the assumption that $Ch(W) = \varnothing$. Moreover, the route does not contain any edge $\circ\!\!\rightarrow Y$ that is in some path in $\Pi_{X:Y}$ because, otherwise, $\rho_{X:W}$ contains the edge ($\varrho_{Y:W}$ cannot by definition) and thus it reaches Y first. Then, the path corresponding to $\varrho_{X:W} \cup \rho_{W:Y}$ contradicts the assumptions in the lemma.

Consequently, $X \perp W|Z \cup Y$ and, thus, the desired result follows from Lemma 15. ■

Proof of Theorem 10 Consider hereinafter the path diagram conditioned on S. We first compute $\sigma_{XY \cdot Z^1}$ from σ_{XY} by adding the nodes in Z^1 to the conditioning set in the order $W_1, W_2, \ldots$. The assumption that $\Pi_{X:Y}$ are all the Z-open paths from X to Y implies that $\Pi_{X:Y}$ are all the $(W_{1:j-1})$-open paths from X to Y because, otherwise, any other path cannot be closed afterwards which contradicts the assumption. To see it, note that a node $W \in Z$ does not close any open path, since there is no subpath $\circ\!\!\rightarrow W \rightarrow$ or $\leftarrow W \rightarrow$ due to the conditioning operation. Then,

$$\sigma_{XY \cdot W_1} = \sigma_{XY} \frac{\sigma^2_{X_1 \cdot W_1}}{\sigma^2_{X_1}}$$

by Lemma 17 if $Pa(W_1) \cup Sp(W_1) = \varnothing$, or Lemma 18 if $Ch(W_1) = \varnothing$. Likewise,

$$\sigma_{XY \cdot W_1 W_2} = \sigma_{XY \cdot W_1} \frac{\sigma^2_{X_1 \cdot W_1 W_2}}{\sigma^2_{X_1 \cdot W_1}}$$

by Lemma 17 or 18. Combining the last two equations gives

$$\sigma_{XY \cdot W_1 W_2} = \sigma_{XY} \frac{\sigma^2_{X_1 \cdot W_1 W_2}}{\sigma^2_{X_1}}.$$

Continuing with this process for the rest of the nodes in Z^1 gives

$$\sigma_{XY \cdot Z^1} = \sigma_{XY} \frac{\sigma^2_{X_1 \cdot Z^1}}{\sigma^2_{X_1}}. \tag{6}$$

Now, we compute $\sigma_{XY \cdot Z^1_1}$ from $\sigma_{XY \cdot Z^1}$ by adding the nodes in Z_1 to the conditioning set in the order $W_1, W_2, \ldots$. Recall from above that $\Pi_{X:Y}$ are all the $(Z^1 \cup W_{1:j-1})$-open paths from X to Y. Then,

$$\sigma_{XY \cdot Z^1 Z_1} = \sigma_{XY \cdot Z^1} \frac{\sigma^2_{X_1 \cdot Z^1 Z_1}}{\sigma^2_{X_1 \cdot Z^1}}$$

by repeating the reasoning that led to Equation 6. Moreover, combining the last two equations yields

$$\sigma_{XY \cdot Z_1^1} = \sigma_{XY} \frac{\sigma^2_{X_1 \cdot Z_1^1}}{\sigma^2_{X_1}}. \tag{7}$$

Now, we compute $\sigma_{XY \cdot Z_1^1 Z^2}$ from $\sigma_{XY \cdot Z_1^1}$ by adding the nodes in Z^2 to the conditioning set in the order $W_1, W_2, \ldots$. The assumption that there is no Z-open route $X_2 \rightarrow A \multimap \cdots \multimap B \rightarrowtail X_2$ implies that there is no $(Z_1^1 \cup W_{1:j-1})$-open path from W_j to X_2 through $Ch(X_2)$. To see it, assume the opposite. Then, there are $(Z_1^1 \cup W_{1:j-1})$-open paths from W_j to X_2 through both $Ch(X_2)$ and $Pa(X_2) \cup Sp(X_2)$. This implies that there is a route $X_2 \rightarrow A \multimap \cdots \multimap B \rightarrowtail X_2$ that contains W_j, and the route is $(Z_1^1 \cup W_{1:j-1})$-open or $(Z_1^1 \cup W_{1:j})$-open. Then, there is a path $\rightarrow A \multimap \cdots \multimap B \rightarrowtail$ that is $(Z_1^1 \cup W_{1:j-1})$-open or $(Z_1^1 \cup W_{1:j})$-open by Lemma 13. However, this path cannot be closed afterwards which contradicts the assumption. To see it, note that a node $W \in Z$ does not close any open path, since there is no subpath $\rightarrowtail W \rightarrow$ or $\leftarrow W \rightarrow$ due to the conditioning operation. Consequently, there is no $(Z_1^1 \cup W_{1:j-1})$-open path from W_j to X_2 through $Ch(X_2)$. Recall also from above that $\Pi_{X:Y}$ are all the $(Z_1^1 \cup W_{1:j-1})$-open paths from X to Y. Then,

$$\sigma_{XY \cdot Z_1^1 W_1} = \sigma_{XY \cdot Z_1^1}.$$

by Lemma 19 if $Pa(W_1) \cup Sp(W_1) = \varnothing$, or Lemma 20 if $Ch(W_1) = \varnothing$. Likewise,

$$\sigma_{XY \cdot Z_1^1 W_1 W_2} = \sigma_{XY \cdot Z_1^1 W_1}.$$

by Lemma 19 or 20. Combining the last two equations gives

$$\sigma_{XY \cdot W_1 W_2} = \sigma_{XY \cdot Z_1^1}.$$

Continuing with this process for the rest of the nodes in Z^2 gives

$$\sigma_{XY \cdot Z_1^1 Z^2} = \sigma_{XY \cdot Z_1^1}. \tag{8}$$

Now, we compute $\sigma_{XY \cdot Z_1^1 Z_2^2}$ from $\sigma_{XY \cdot Z_1^1 Z^2}$ by adding the nodes in Z_2 to the conditioning set in the order $W_1, W_2, \ldots$. Recall from above that $\Pi_{X:Y}$ are all the $(Z_1^{1:2} \cup W_{1:j-1})$-open paths from X to Y. Recall also from above that the assumption that there is no Z-open route $X_2 \rightarrow A \multimap \cdots \multimap B \rightarrowtail X_2$ implies that there is no $(Z_1^{1:2} \cup W_{1:j-1})$-open path from W_j to X_2 through $Pa(X_2) \cup Sp(X_2)$. Then,

$$\sigma_{XY \cdot Z_1^1 Z_2^2} = \sigma_{XY \cdot Z_1^1 Z^2} \frac{\sigma^2_{X_2 \cdot Z_1^1 Z_2^2}}{\sigma^2_{X_2 \cdot Z_1^1 Z^2}}$$

by repeating the reasoning that led to Equation 6 but using Lemmas 21 and 22 instead. Moreover, combining the last equation with Equations 7 and 8 gives

$$\sigma_{XY \cdot Z_1^1 Z_2^2} = \sigma_{XY} \frac{\sigma^2_{X_1 \cdot Z_1^1}}{\sigma^2_{X_1}} \frac{\sigma^2_{X_2 \cdot Z_1^1 Z_2^2}}{\sigma^2_{X_2 \cdot Z_1^1 Z^2}}.$$

Finally, continuing with the process above for $X_3, \ldots, X_{m+n}$ yields

$$\sigma_{XY \cdot Z_{1:m+n}^{1:m+n}} = \sigma_{XY} \frac{\sigma^2_{X_1 \cdot Z_1^1}}{\sigma^2_{X_1}} \prod_{i=2}^{m+n} \frac{\sigma^2_{X_i \cdot Z_{1:i}^{1:i}}}{\sigma^2_{X_i \cdot Z_{1:i-1}^{1:i}}}$$

which implies the desired result by repeated application of Lemma 16. ■

Proof of Theorem 11 The proof is analogous to that of Theorem 10. ■