On statistical models associated with acyclic directed mixed graphs

Qingyuan Zhao[*]

September 30, 2024

Abstract

Causal models in statistics are often described by acyclic directed mixed graphs (ADMGs), which contain directed and bidirected edges and no directed cycles. This article surveys various interpretations of ADMGs, discusses their relations in different sub-classes of ADMGs, and argues that one of them—nonparametric equation system (the E model below)—should be used as the default interpretation. The E model is closely related to but different from the interpretation of ADMGs as directed acyclic graphs (DAGs) with latent vairables that is commonly found in the literature. Our endorsement of the E model is based on two observations. First, in a subclass of ADMGs called unconfounded graphs (which retain most of the good properties of directed acyclic graphs and bidirected graphs), the E model is equivalent to many other interpretations including the global Markov and nested Markov models. Second, the E model for an arbitrary ADMG is exactly the union of that for all unconfounded expansions of that graph. This property is referred to as *completeness*, as it shows that the model does not commit to any specific latent variable explanation. In proving that the E model is nested Markov, we also develop an ADMG-based theory for causality that may be of independent interest.

1 Introduction

Acyclic directed mixed graphs (ADMGs) are first used by Wright (1934) to describe causal relationships between a collection of random variables. They play a central role in the modern statistical theory for causality; see, for example, Pearl (2009) and Richardson, Evans, et al. (2023). ADMGs have two types of edges—directed and bidirected. When interpreting model assumptions encoded by ADMGs, two heuristics are commonly used:

1. A directed edge means direct causal influence and a bidirected edge means exogenous correlation (Wright (1934) calls this "residual correlation").
2. ADMG describes a latent variable model because, in the definition of "latent projection" of ADMGs by Verma and Pearl (1990), the graphical structures

$$V_1 \longleftarrow V_2 \longrightarrow V_3,\ V_1 \longleftrightarrow V_2 \longrightarrow V_3,\ \text{and } V_1 \longleftarrow V_2 \longleftrightarrow V_3$$

   all marginalize to $V_1 \longleftrightarrow V_3$ when we treat V_2 as unobserved.

[*]Statistical Laboratory, University of Cambridge, qyzhao@statslab.cam.ac.uk.

Based on these heuristics, many interpretations of ADMGs have been proposed in the literature (see e.g. Richardson 2003; Peters, Janzing, and Schölkopf 2017; Bareinboim et al. 2022). Unfortunately, these interpretations generally do not agree with each other, and it is notoriously difficult to describe the complicated constraints imposed by the latent variables on the probability distribution of the observed variables. The purpose of this article is to give a survey of those interpretations, discuss their relations, and put forward a case that one of those interpretations—the nonparametric equation system (the E model below)—should be used as the default. A key argument is that the E model is *complete* with respect to certain latent variable explanations, a concept that will be defined shortly.

Before moving to any technincal discussion, it is useful to first consider when an interpretation may be regarded as natural. Generally speaking, a natural or interesting mathematical definition can be found in at least two ways:

Equivalence When many definitions motivated by apparently different considerations are equivalent to each other, we may believe they describe a natural mathematical concept.

Completion When there exists a natural definition for a smaller class of mathematical objects, we may try to find a "completion" of that definition to a larger class of objects.

In fact, the Equivalence argument is regularly used in the graphical models literature. A prominent example is the Hammersley-Clifford theorem, which shows that two statistical models (of distributions with positive densities) associated with a undirected graph—one defined via factorization and another via Markov property—are equivalent. Another familiar example is the equivalence of the factorization model ("Bayesian networks") and the global Markov model associated with directed acyclic graphs (DAGs). However, the Equivalence argument by itself cannot define the "right" interpretation of ADMGs. In fact, we will see shortly that most common interpretations of ADMGs in statistics are genuinely different. Given this, one may be tempted to use the Completion argument instead. We will see below that this is indeed possible but requires a careful definition of "completion".

1.1 Directed mixed graphs

A directed mixed graph $\mathrm{G} = (V, \mathcal{D}, \mathcal{B})$ consists of a vertex set V, a set $\mathcal{D} \subseteq V \times V$ of *directed edges*, and a set $\mathcal{B} \subseteq V \times V$ of *bidirected edges* that are required to be symmetric:

$$(V_j, V_k) \in \mathcal{B} \iff (V_k, V_j) \in \mathcal{B}, \text{ for all } V_j, V_k \in V.$$

It is helpful to think about the edges as relations between the vertices and write

$$V_j \longrightarrow V_k \textbf{ in } \mathrm{G} \iff (V_j, V_k) \in \mathcal{D} \quad \text{and} \quad V_j \longleftrightarrow V_k \textbf{ in } \mathrm{G} \iff (V_j, V_k) \in \mathcal{B}\,.$$

The choice of drawing edges in $\mathcal{B}$ as bidirected instead of undirected is intentional and crucial. This is also where the name "directed mixed graph" comes from Richardson (2003). Let $\mathbb{G}(V)$ denote the set of all such graphs. Note that loops, whether bidirected (such as $V_j \longleftrightarrow V_j$) or directed (such as $V_j \longrightarrow V_j$), are allowed. For the remaining of this article, we will assume the graph contains all bidirected loops, that is, $V_j \longleftrightarrow V_j$ **in** G for all $V_j \in V$.[1] Let $\mathbb{G}^*(V)$ denote the collection of all such "canonical" graphs.

Some important subclasses of $\mathbb{G}^*(V)$ include:

[1] Loosely speaking, this means that we allow "random innovations" at each vertex in the corresponding statistical models.

- $\mathbb{G}^*_{\mathrm{B}}(V)$: the class of bidirected graphs (i.e. the directed edge set $\mathcal{D} = \emptyset$);
- $\mathbb{G}^*_{\mathrm{D}}(V)$: the class of directed graphs that contain no bidirected edges other than bidirected loops;
- $\mathbb{G}^*_{\mathrm{A}}(V)$: the class of acyclic directed mixed graphs (ADMGs), where by *acyclic*, we mean there exists no cyclic directed walks like $V_j \longrightarrow \cdots \longrightarrow V_j$ for any $V_j \in V$;
- $\mathbb{G}^*_{\mathrm{DA}}(V) = \mathbb{G}^*_{\mathrm{D}}(V) \cap \mathbb{G}^*_{\mathrm{A}}(V)$: the class of directed acyclic graphs (DAGs).

It is convenient to not actually draw the bidirected loops for graphs in $\mathbb{G}^*(V)$. Indeed, this defines an isomorphism from $\mathbb{G}^*(V)$ that contains *all* bidirected loops to the subclass of $\mathbb{G}(V)$ that contains *no* bidirected loops. For this reason, we will generally not distinguish between these two subclasses in this article.[2]

Let us introduce a new subclass of $\mathbb{G}^*(V)$ that will play an important role in our argument below.

Definition 1. Given $\mathrm{G} \in \mathbb{G}^*(V)$, the set of *exogenous* vertices is defined as

$$E = \{V_j \in V : V_k \not\longrightarrow V_j \text{ for all } V_k \in V\}.$$

We say G is *unconfounded* if for all $V_j, V_k \in V$ such that $V_j \neq V_k$,

$$V_j \longleftrightarrow V_k \textbf{ in } \mathrm{G} \Longrightarrow V_j, V_k \in E. \tag{1}$$

Let $\mathbb{G}^*_{\mathrm{U}}(V)$ denote the set of all such unconfounded graphs with vertex set V and $\mathbb{G}^*_{\mathrm{UA}}(V) = \mathbb{G}^*_{\mathrm{U}}(V) \cap \mathbb{G}^*_{\mathrm{A}}(V)$.

The semantics of a unconfounded ADMG is simple: the exogenous vertices have some underlying structure as described by the bidirected edges, and they influence the rest of the *endogenous* vertices in a recursive way through the directed edges. The name "unconfounded" is derived from the fact that when such graphs are interpreted causally, all interventional distributions can be identified from the distribution of V because all vertices in the graph are fixable; see Section 4 for more detail. It is obvious that $\mathbb{G}^*_{\mathrm{DA}}(V) \subseteq \mathbb{G}^*_{\mathrm{UA}}(V)$ and $\mathbb{G}^*_{\mathrm{B}}(V) \subseteq \mathbb{G}^*_{\mathrm{UA}}(V)$. In fact, we will see that unconfounded ADMGs share many good properties as DAGs and bidirected graphs.

Note that a similar but different type of graphs is considered by Kiiveri, Speed, and Carlin (1984). There, the exogenous variables are connected by undirected edges and are required to satisfy the global Markov property for undirected graphs, so what they consider is a subclass of the chain graph models (Lauritzen and Wermuth 1989; Frydenberg 1990).

1.2 Statistical models associated with ADMGs and their relations

In graphical statistical models, vertices in the graph are random variables, and there are different ways to interpret edges as relationships between the variables. To formalize such interpretations as statistical models, it is helpful to take the more abstract point of view that a statistical model is a collection of probability distributions. Let $V = (V_1, \ldots, V_d)$ be a random vector that takes values in a product measure space $\mathbb{V} = \mathbb{V}_1 \times \cdots \times \mathbb{V}_d$, and the largest statistical model we will consider is denoted as $\mathbb{P}(\mathbb{V})$, the set of all probability distributions on $\mathbb{V}$ with a density function. With the

[2] One might ask why we do not start with graphs without bidirected loops in the first place. This is mainly because in some problems (not considered here) it is useful to consider graphs in which some vertices have bidirected loops and some do not. One example is the single-world intervention graphs (Richardson and Robins 2013).

possible addition of some regularity (e.g. smoothness) conditions on the density function, this is often referred to as the *nonparametric model* in the statistics literature.

Graphical statistical models associate graphs with subclasses of $\mathbb{P}(\mathbb{V})$; in other words, they are maps from $\mathbb{G}^*_{\mathrm{A}}(V)$ to the power set of $\mathbb{P}(\mathbb{V})$. Let us illustrate this by introducing some ADMG models here:

1. One approach is to associate separations in the graph with conditional independences in the probability distribution. For example, for $\mathrm{G} \in \mathbb{G}^*(V)$, the *global Markov* (GM) model collects all distributions $\mathsf{P} \in \mathbb{P}(\mathbb{V})$ that obeys the global Markov property with respect to G: m-separation in G (this will be defined in Section 2.2) implies conditional independence under P. This is first formally introduced by Richardson (2003) and goes back to earlier investgiations of (cyclic) linear structural equation models.
2. Another approach is to consider certain "expansions" of the graph that have a simpler structure. For example, the *clique expansion* (CE) model expands every clique of bidirected edges with a latent variable, so the resulting graph is a DAG. The *noise expansion* (NE) model associate each vertex in the graph with a latent variable that inherits all its bidirected edges, so the resulting graph is unconfounded. See Figure 2 below for some examples. Many authors take this approach implicitly without fullying defining their model; a more explicit account is given in Richardson, Evans, et al. (2023, Section 4.1).
3. Alternatively, one can consider a system of equations that obey the local structure of the graph. The *nonparametric system of equations* (E) model collects all distribution P of V such that V can be written as (the following event has probability 1 under P):

$$V_j = f_j(V_{\mathrm{pa}(j)}, E_j), \quad \text{for all } V_j \in V,$$

   for some functions $f_1, \dots, f_d$, where $\mathrm{pa}(j) = \{k : V_k \longrightarrow V_j \textbf{ in } \mathrm{G}\}$ is the parent set of V_j in G and, importantly, the distribution of the "noise variables" $E = (E_1, \dots, E_d)$ is global Markov with respect to the bidirecte component of G. This is closely related to the "semi-Markovian" causal model in Pearl (2009, p. 30) and Bareinboim et al. (2022) who leave the distribution of E unspecified.

We will formally define the above models and some other interpretations of ADMGs in Section 3.

The next Theorem summarizes the relations between those models. Many results in this Theorem are already obtained in the literature. Among the new claims, the most non-trivial result is that the E/NE model is nested Markov (NM), although this is not totally surprising given that Richardson, Evans, et al. (2023, Section 4.1) have shown that the marginal of any DAG model is nested Markov with respect to the corresponding marginal ADMG (which basically means CE $\Rightarrow$ NM in our terminology). We prove E/NE $\Rightarrow$ NM by considering a causal Markov model associated with ADMGs, and this proof is outlined in Section 4. All other proofs can be found in the Appendix.

Theorem 1. *The relations in Figure 1 hold for all* G *in the corresponding classes of graphs, where* $\Rightarrow$ *(*$\Leftrightarrow$*) should be interpreted as* $\supseteq$ *(*$=$*) for the corresponding graphical statistical models with the same state space* $\mathbb{V}$*. Moreover, all* $\Rightarrow$ *in Figure 1 are strict in the sense that the reverse impliations are not true for all graphs in the corresponding subclass.*

Although we have not introduced most statistical models in Figure 1 yet, some high-level observations can already be made:

$$
\begin{array}{l}
\text{PE} \\
\Downarrow \\
\text{CE} \\
\Downarrow \\
\text{NE} \iff \text{E} \\
\Downarrow \\
\text{NM} \\
\Downarrow \\
\text{LM} \Leftrightarrow \text{GM} \Leftrightarrow \text{A} \\
\Downarrow \\
\text{UM}
\end{array}
$$

(a) G is an ADMG: $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$.

$$
\begin{array}{l}
\text{PE} \\
\Downarrow \\
\text{CE} \\
\Downarrow \\
\text{NE} \Leftrightarrow \text{E} \Leftrightarrow \text{NM} \Leftrightarrow \text{EF} \Leftrightarrow \text{LM} \Leftrightarrow \text{GM} \Leftrightarrow \text{A} \\
\Downarrow \\
\text{UM}
\end{array}
$$

(b) G is an unconfounded ADMG: $\mathrm{G} \in \mathbb{G}^*_{\mathrm{UA}}(V)$.

$$
\begin{array}{l}
\text{PE} \Leftrightarrow \text{CE} \Leftrightarrow \text{NE} \Leftrightarrow \text{E} \Leftrightarrow \text{NM} \Leftrightarrow \text{EF} \Leftrightarrow \text{F} \Leftrightarrow \text{LM} \Leftrightarrow \text{GM} \Leftrightarrow \text{A} \\
\Downarrow \\
\text{UM}
\end{array}
$$

(c) G is a DAG: $\mathrm{G} \in \mathbb{G}^*_{\mathrm{DA}}(V)$.

$$
\begin{array}{l}
\text{PE} \\
\Downarrow \\
\text{CE} \\
\Downarrow \\
\text{NE} \Leftrightarrow \text{E} \Leftrightarrow \text{NM} \Leftrightarrow \text{EF} \Leftrightarrow \text{LM} \Leftrightarrow \text{GM} \Leftrightarrow \text{A} \Leftrightarrow \text{UM}
\end{array}
$$

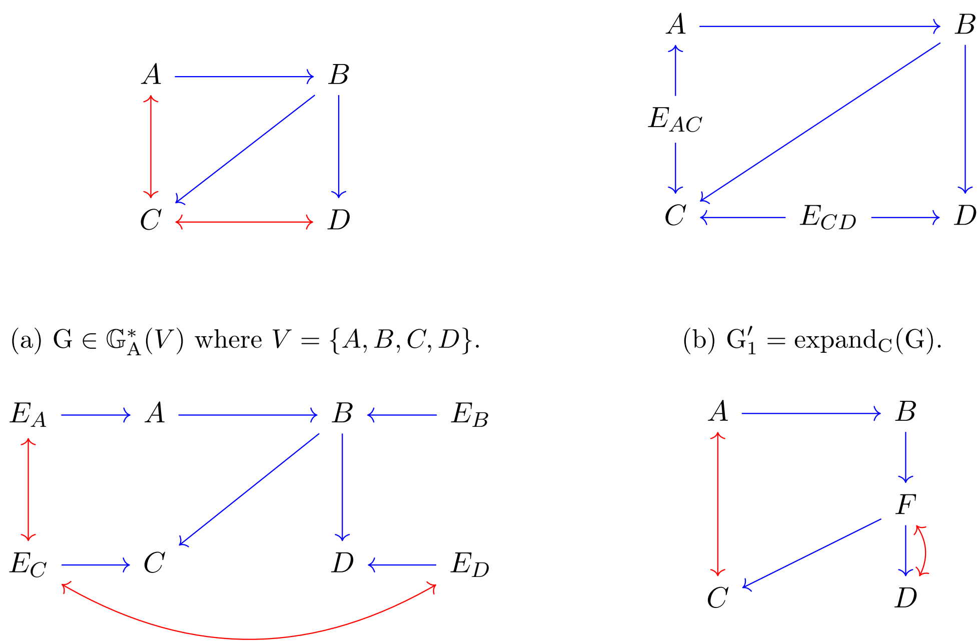
(d) G is a bidirected graph: $\mathrm{G} \in \mathbb{G}^*_{\mathrm{B}}(V)$.

Figure 1: Relations between some statistical models associated with (subclasses of) ADMGs that are formally defined in Section 3. (A: Augmentation; CE: Clique Expansion; E: Nonparametric Equations; EF: Exogenous Factorization; F: Factorization; GM: Global Markov; LM: Local Markov; NE: Noise Expansion; NM: Nested Markov; PE: Pairwise Expansion; UM: Unconditional Markov.)

1. Unconfounded ADMGs share the equivalences of statistical models that are found for DAGs and bidirected graphs. For example, E $\Leftrightarrow$ GM is true for unconfounded ADMGs (and thus DAGs and bidirected graphs) but not all ADMGs. For this reason, unconfounded ADMGs may be considered as the natural generalization of DAGs and bidirected graphs.
2. A general ADMG is associated with many "tiers" of statistical models that are generally not equivalent with each other. Thus, the Equivalence argument does not give a natural definition of statistical model for all ADMGs.

1.3 Graph expansion and complete models

We will now turn to the Completion argument and define what we mean by "complete". To this end, let margin_V denote the (overloaded) "marginalization" operator on ADMGs (that maps $\mathbb{G}^*_{\mathrm{A}}(V')$ to $\mathbb{G}^*_{\mathrm{A}}(V)$ for some $V' \supseteq V$) and probability distributions (that maps $\mathbb{P}(V')$ to $\mathbb{P}(V)$); these will be



(a) $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$ where $V = \{A, B, C, D\}$.

(b) $\mathrm{G}'_1 = \mathrm{expand}_{\mathrm{C}}(\mathrm{G})$.

(c) $\mathrm{G}'_2 = \mathrm{expand}_{\mathrm{N}}(\mathrm{G})$.

(d) $\mathrm{G}'_3 \in \mathrm{expand}_{V'}(\mathrm{G})$ where $V' = V \cup \{F\}$.

Figure 2: Examples of graph expansion (all bidirected loops are omitted).

formally in Section 2.3. Let $\mathrm{expand}_{V'}$ denote the inverse image of margin_V, that is,

$$\mathrm{expand}_{V'}(\mathrm{G}) = \left\{ \mathrm{G}' \in \mathbb{G}^*_{\mathrm{A}}(V') : \mathrm{margin}_V(\mathrm{G}') = \mathrm{G} \right\}.$$

Definition 2. For every possible vertex set V, let $\mathcal{G}_0(V) \subseteq \mathbb{G}^*_{\mathrm{A}}(V)$ be a given subclass of (canonical) ADMGs. A collection of statistical models $\mathbb{P}(\mathrm{G})$ for different ADMGs G is said to be *complete* (with respect to expansions in $\mathcal{G}_0$) if it is equal to the union of the V-marginal of all $\mathcal{G}_0$-"expanded" models, that is,

$$\mathbb{P}(\mathrm{G}) = \bigcup_{V' \supset V} \bigcup_{\mathrm{G}'} \mathrm{margin}_V(\mathbb{P}(\mathrm{G}')), \tag{2}$$

where the second union is over $\mathrm{G}' \in \mathrm{expand}_{V'}(\mathrm{G}) \cap \mathcal{G}_0(V')$.

Equation (2) is essentially a way to extend the "base model"—statistical models for a smaller class of graphs—to a larger class of graphs. This heuristic can be widely used in the literature to interpret ADMGs; for example, Pearl (2009, p. 76) writes "... especially true in semi-Markovian models (i.e., DAGs involving unmeasured variables)". This intuitive "latent DAG" interpretation is formalized in Richardson, Evans, et al. (2023, Section 4.1) who use DAGs as the base model (i.e. $\mathcal{G}_0(V) = \mathbb{G}^*_{\mathrm{DA}}(V)$). Theorem 2 below further shows that this latent DAG interpretation is equivalent to the clique expansion (CE) model. However, the CE model implies complicated inequality constraints on the probability distribution (Fritz 2012; Evans 2016). Conceptually, this difficulty arises because the latent DAG interpretation does not treat bidirected graphs as an interesting primitive structure.

We propose to use unconfounded ADMGs as the base model (i.e. $\mathcal{G}_0(V) = \mathbb{G}^*_{\mathrm{UA}}(V)$) in Definition 2. As argued below Theorem 1, unconfounded ADMGs are natural generalizations of DAGs and bidirected graphs. With this choice, a complete ADMG model uses a unconfounded graph

$$
\begin{tikzcd}
\mathrm{PE} \arrow[d, Rightarrow] & & & \mathrm{PE} \arrow[d, Rightarrow] \\
\mathrm{CE} \arrow[d, Rightarrow] & & & \mathrm{CE} \arrow[lll, "\text{Completion}"'] \arrow[d, Rightarrow] \\
\mathrm{NE} \arrow[d, Rightarrow] \arrow[r, Leftrightarrow] & \mathrm{E} & & \mathrm{NE} \arrow[ll, "\text{Completion}"'] \arrow[ddd, Rightarrow] \arrow[r, Leftrightarrow] & \mathrm{E} \arrow[r, Leftrightarrow] & \mathrm{EF} \arrow[r, Leftrightarrow] & \mathrm{NM} \arrow[r, Leftrightarrow] & \mathrm{LM} \arrow[r, Leftrightarrow] & \mathrm{GM} \arrow[r, Leftrightarrow] & \mathrm{A} \\
\mathrm{NM} \arrow[d, Rightarrow] & & & \\
\mathrm{LM} \arrow[d, Rightarrow] \arrow[r, Leftrightarrow] & \mathrm{GM} \arrow[r, Leftrightarrow] & \mathrm{A} & \\
\mathrm{UM} & & & \mathrm{UM} \arrow[lll, "\text{Completion}"']
\end{tikzcd}
$$

General ADMGs ($\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$) | Unconfounded ADMGs ($\mathrm{G} \in \mathbb{G}^*_{\mathrm{UA}}(V)$)

Figure 3: Completion of ADMG models.

expansion with "latent" variables to explain an ADMG but does not commit to a particular unconfounded expansion. This is a form of agnostic reasoning: a complete ADMG model does not try to tell us *why* two variables are related. For instance, a bidirected edge intuitively means that two variables are correlated in an exogenous way, possibly due to one or multiple latent common causes. However, the nature of that exogenous correlation is not part of a complete model. Similarly, when the ADMG is interpreted as a causal model, a directed edge is usually interpreted as a direct causal effect not through other variables in the graph. It is entirely possible that such a direct causal effect is mediated through one or multiple latent variables, but that is, again, not part of a complete model.

To illustrate the definition of complete models, consider the graphs in Figure 2. The graph $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$ for $V = \{A, B, C, D\}$ in Figure 2a is not unconfounded, but after the "clique" (Figure 2b) or "noise" expansion (Figure 2c), it becomes an unconfounded graph. The remark after Theorem 1 suggests that the nonparametric system of equations is a natural statistical model associated with such graphs. In other words, it is reasonable to require $\mathbb{P}(\mathrm{G}_1') = \mathbb{P}_{\mathrm{E}}(\mathrm{G}_1')$ and $\mathbb{P}(\mathrm{G}_2') = \mathbb{P}_{\mathrm{E}}(\mathrm{G}_2')$. Figure 2d shows another possible expansion of G that involves a latent variable F, but the expanded graph is confounded because of $A \longleftrightarrow C \longleftarrow F$ and $B \longrightarrow F \longleftrightarrow D$ (one can further expand the bidirected edges to make the graph unconfounded). By requiring the model $\mathbb{P}(\mathrm{G})$ to be complete with respect to unconfounded graphs, it should contain the V-marginals of $\mathbb{P}(\mathrm{G}_1')$, $\mathbb{P}(\mathrm{G}_2')$.

Figure 3 visualizes the following Theorem on complete ADMG models.

Theorem 2. *Among the ADMG models in Figure 1a, only the CE, E, NE, and UM models are complete with respect to unconfounded graph expansions (taking $\mathcal{G}_0(V) = \mathbb{G}^*_{\mathrm{UA}}(V)$ in Definition 2).*

Figure 1a shows that the E and NE models are equivalent, so Theorem 2 really identifies three different complete models associated with ADMGs. Between them, the choice rests on how one interprets unconfounded ADMGs. By applying the Equivalence argument to Figure 1b, the E/NE model is the most natural interpretation for unconfounded ADMGs. For this reason, we believe that the E/NE model should be used as the default interpretation of general ADMGs.

The rest of this article is organized as follows. In Section 2, we introduce some basic notation and terminology for graphical statistical models. In Section 3, we formally define the statistical models that appear in Theorem 1. In Section 4, we outline a proof of the assertion that the nonparametric equation system is nested Markov by building a theory for causality based on ADMGs. In Section 5, we give some further remarks. Technical proofs can be found in the Appendix.

2 Basic notation and terminology

2.1 Conditional independence

Intuitively, a graphical statistical model imposes algebraic (and semi-algebraic) constraints on probability distributions according to certain structures in the graph. Perhaps the simplest form of algebraic constraints on probability distributions is conditional independence: for disjoint subsets $V_{\mathcal{J}}, V_{\mathcal{K}}, V_{\mathcal{L}}$ of V, define

$$V_{\mathcal{J}} \perp\!\!\!\perp V_{\mathcal{K}} \mid V_{\mathcal{L}} \textbf{ under } \mathsf{P} \iff \mathsf{p}(v_{\mathcal{J}}, v_{\mathcal{K}} \mid v_{\mathcal{L}}) = \mathsf{p}(v_{\mathcal{J}} \mid v_{\mathcal{L}})\, \mathsf{p}(v_{\mathcal{K}} \mid v_{\mathcal{L}}) \text{ for all } v_{\mathcal{L}} \text{ such that } \mathsf{p}(v_{\mathcal{L}}) > 0,$$

where $\mathsf{p}(v_{\mathcal{J}}, v_{\mathcal{K}} \mid v_{\mathcal{L}})$ is the conditional density function of $V_{\mathcal{J}}$ and $V_{\mathcal{K}}$ given $V_{\mathcal{L}}$ evaluated at $(v_{\mathcal{J}}, v_{\mathcal{K}}, v_{\mathcal{L}})$ (under law P) and other conditional densities are defined similarly.

Conditional independence satisfies a number of "graphoid axioms" that bear a close relation to graph separation; see Pearl (1988) and Lauritzen (1996).

2.2 Walk algebra

We adopt the notation and terminology in Zhao (2024) to describe the walk algebra generated by directed mixed graphs. For $V_j, V_k \in V$, we say w is a *walk* from V_j to V_k if it is a sequence of connecting edges (edge directions are ignored when deciding connection), its first edge starts at V_j, and its last edge ends at V_k. We say a walk is *blocked* by $L \subseteq V$ if

1. w contains a collider V_l (so part of w looks like $\hookrightarrow V_l \hookleftarrow$) such that $V_l \notin L$; or
2. w contains a non-collider such that $V_l \in L$.

This is slightly different from (but in ADMGs equivalent to) the notion of blocking for paths usually used in the literature which requires that no desendants of any collider V_l is in L. See Zhao (2024) for further discussion.

If $\{V_j\}$, $\{V_k\}$, and L are disjoint and there exists an unblocked walk from V_j to V_k given L, we say V_j is *m-connected* to V_k given L and write $V_j \rightsquigarrow * \leftsquigarrow V_k \mid L$ **in** G; the half arrowheads mean the walk can end with or without arrowheads on both sides, and the asterisk is a wildcard character to indicate that the walk may have any number of colliders. If no such walk exists, we say V_j and V_k are *m-separated* given L in G and write **not** $V_j \rightsquigarrow * \leftsquigarrow V_k \mid L$ **in** G. This definition naturally extends to sets of subvertices: for disjoint $J, K, L \subset V$, we write

$$J \rightsquigarrow * \leftsquigarrow K \mid L \textbf{ in } \mathrm{G} \iff V_j \rightsquigarrow * \leftsquigarrow V_k \mid L \textbf{ in } \mathrm{G} \text{ for some } V_j \in J, V_k \in K.$$

We now introduce some special types of walks and associated concepts that play important roles in the theory of ADMG models:

1. $\longrightarrow$ and $\longleftrightarrow$: these are the basic edges that generate the walk algebra. We write $\mathrm{pa}_{\mathrm{G}}(V_j) = \{V_k \in V : V_k \longrightarrow V_j\}$ as the *parents* of V_j and $\mathrm{ch}_{\mathrm{G}}(V_j) = \{V_k \in V : V_j \longrightarrow V_k\}$ as the *children* of V_j. When it is more convenient to work with indicies of the variables, we often use the notation $\mathrm{pa}_{\mathrm{G}}(j) = \{k \in [d] : V_k \longrightarrow V_j\}$ and likewise for $\mathrm{ch}_{\mathrm{G}}(j)$. We sometimes omit the graph G in the subscript when it is clear from the context.
2. $\rightsquigarrow$: this means a *(right-)directed walk* that consists of one or more $\longrightarrow$. We write $\mathrm{an}(V_j) = \{V_k \in V : V_k \longrightarrow V_j\}$ as the *ancestors* of V_j and $\mathrm{de}(V_j) = \{V_k \in V : V_j \longrightarrow V_k\}$ as the *descendants* of V_j. The corresponding indices are denoted as $\mathrm{an}(j)$ and $\mathrm{de}(j)$, respectively. We say a subset of vertices $K \subseteq V$ is *ancestral* in G if it contains all its ancestors, that is,

$$\mathrm{an}(K) = \{V_k \in V : V_k \rightsquigarrow K \textbf{ in } \mathrm{G}\} \subseteq K.$$

   This concept is useful because the ancestral marginal of an ADMG is simply its induced subgraph, that is, if K is ancestral, then $\mathrm{margin}_K(\mathrm{G}) = (K, \mathcal{D} \cap (K \times K), \mathcal{B} \cap (K \times K))$ for $\mathrm{G} = (V, \mathcal{D}, \mathcal{B})$.
3. $\leftharpoonup\!\sim\!\rightharpoondown$: this means an *arc*, a walk with no colliders.
4. $\leftrightsquigarrow$: these are all arcs that are not $\rightsquigarrow$ or $\leftsquigarrow$ (consisting of one or more $\longleftarrow$). When a walk like $\leftrightsquigarrow$ is a path, we call it a *confounding arc*.
5. $\longleftrightarrow * \longleftrightarrow$: this means a walk consisting one or more $\longleftrightarrow$. Vertices connected by such walks are said to be in the same *district*.[3]
6. $\hookrightarrow * \hookleftarrow$: this means a *collider-connected walk* in which all non-endpoints are colliders. The set $\mathrm{mb}_{\mathrm{G}}(V_j) = \{V_k \in V : V_k \hookrightarrow * \hookleftarrow V_j \textbf{ in } \mathrm{G}\}$ is called the *Markov boundary* of V_j in G.[4] The corresponding set of indicies is denoted as $\mathrm{mb}_{\mathrm{G}}(j)$.
7. $\hookrightarrow * \longleftrightarrow$: this is a collider-connected walk that ends with an arrowhead. The set $\mathrm{mbg}_{\mathrm{G}}(V_j) = \{V_k \in V : V_k \hookrightarrow * \longleftrightarrow V_j \textbf{ in } \mathrm{G}\}$ is called the *Markov background* of V_j in G.[5] The corresponding set of indicies is denoted as $\mathrm{mbg}_{\mathrm{G}}(j)$.
8. $\rightsquigarrow * \leftsquigarrow$: this is a walk consisting of one or more arcs, which is simply any walk in the graph.

A formal definition of these and some other important types of walks can be found in Zhao (2024).

2.3 Marginalization

As our argument rests on considering latent variable "expansions" or "explanations" of graphical models, let us take some care to define the related concepts. Consider an ADMG $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$. We restrict ourselves to the case where each vertex $V_j \in V, j = 1, \ldots, d$, of the graph is a finite-dimensional real random variable, so $\mathbb{V}_j \subseteq \mathbb{R}^{n_j}$ for some $n_j \in \mathbb{Z}^+$. We assume that $\mathbb{V}_j$ is a measure space and the choice of measure will be implicit in the definitions below; in practice, this is usually the Lebesgue measure if the random variable is continuous or the counting measure if the random variable is discrete. Let $\mathbb{V} = \mathbb{V}_1 \times \cdots \times \mathbb{V}_d$ and $\mathbb{P}(\mathbb{V})$ denote the set of all probability measures on $\mathbb{V}$ with a density function, so $\mathbb{P}(\mathbb{V})$ is isomorphic to the set of non-negative functions on $\mathbb{V}$ with integral 1.

[3] This terminology is due to Richardson (2003). The same concept is called *c-component* in Tian and Pearl (2002).

[4] This is closely related to the concept of Markov blanket and Markov boundary (minimal Markov blanket) in conditional independence models; see Pearl (1988).

[5] When V_j has no children in G, it is obvious that $\mathrm{mb}_{\mathrm{G}}(V_j) = \mathrm{mbg}_{\mathrm{G}}(V_j)$. For this reason, Richardson, Evans, et al. (2023) also refers to $\mathrm{mbg}_{\mathrm{G}}(V_j)$ as the Markov blanket/boundary of V_j. However, this terminology is confusing when V_j is not childless and is avoided here.

The marginalization operator can act on product spaces, probability distributions, and graphs. For any subset $\mathcal{J} \subseteq [d]$ and $J = V_{\mathcal{J}} \subseteq V$, denote the J-marginal of $\mathbb{V}$ as

$$\mathrm{margin}_J(\mathbb{V}) = \mathbb{V}_{\mathcal{J}} = \prod_{j \in \mathcal{J}} \mathbb{V}_j.$$

Further, let $\mathrm{margin}_J(\mathsf{P})$ denote the marginal distribution of J when the joint distribution of V is P, so the density function of $\mathrm{margin}_J(\mathsf{P})$ is simply the marginal density function $\mathsf{p}(v_{\mathcal{J}})$ of $V_{\mathcal{J}}$. Finally, for an ADMG $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$, its J-marginal is defined as its image under the map

$$\begin{aligned} \mathrm{margin}_J : \mathbb{G}^*_{\mathrm{A}}(V) &\to \mathbb{G}^*_{\mathrm{A}}(J), \\ \mathrm{G} &\mapsto \mathrm{G}', \end{aligned}$$

where G' is defined by the following equivalences for all $V_j, V_k \in J$ such that $V_j \neq V_k$:

$$\begin{aligned} V_j \longrightarrow V_k \textbf{ in } \mathrm{G}' &\iff P[V_j \rightsquigarrow V_k \mid J \textbf{ in } \mathrm{G}] \neq \emptyset, \\ V_j \longleftrightarrow V_k \textbf{ in } \mathrm{G}' &\iff P[V_j \leftrightsquigarrow V_k \mid J \textbf{ in } \mathrm{G}] \neq \emptyset. \end{aligned}$$

Here, $P[V_j \rightsquigarrow V_k \mid J \textbf{ in } \mathrm{G}]$ is the set of all directed paths (P means paths) from V_j to V_k in G with no non-endpoints in J, and $P[V_j \leftrightsquigarrow V_k \mid J \textbf{ in } \mathrm{G}]$ is the set of all confounding arcs (paths with no collider and two end-point arrowheads) from V_j to V_k in G with no non-endpoints in J. Marginalization of directed mixed graphs is often referred to as "latent projection" in the literature and is first considered by Verma and Pearl (1990). The reader is invited to check that the graphs in Figures 2b to 2d all marginalize to the graph in Figure 2a.

3 Statistical models associated with directed mixed graphs

3.1 Global Markov (GM) property

The global Markov model assumes that every m-separation in the graph implies a conditional independence in the probability distribution.

Definition 3. The *global Markov model* with respect to $\mathrm{G} \in \mathbb{G}^*(V)$ is defined as

$$\begin{aligned} &\mathbb{P}_{\mathrm{GM}}(\mathrm{G}, \mathbb{V}) \\ =&\{\mathsf{P} \in \mathbb{P}(\mathbb{V}) : \textbf{not } J \leftrightsquigarrow * \leftrightsquigarrow K \mid L \textbf{ in } \mathrm{G} \Longrightarrow J \perp\!\!\!\perp K \mid L \textbf{ under } \mathsf{P} \text{ for all disjoint } J, K, L \subset V\}. \end{aligned}$$

The global Markov model takes simpler forms in some subclasses of $\mathbb{G}^*(V)$. When the graph is directed (so $\mathrm{G} \in \mathbb{G}^*_{\mathrm{D}}(V)$), walks must consist of directed edges (if we ignore bidirected loops) so $\leftrightsquigarrow * \leftrightsquigarrow$ can be written as $\overset{\mathrm{d}}{\leftrightsquigarrow} * \overset{\mathrm{d}}{\leftrightsquigarrow}$ (where d means the walk consist of directed edges only). When the graph is bidirected (so $\mathrm{G} \in \mathbb{G}^*_{\mathrm{B}}(V)$), walks must consist of bidirected edges and $\leftrightsquigarrow * \leftrightsquigarrow$ can be written as $\longleftrightarrow * \longleftrightarrow$ (meaning one or more bidirected edges). See Zhao (2024) for further discussion.

3.2 Unconditional Markov (UM) model

The next model only requires unconditional independences in the global Markov model.

Definition 4. The *unconditional Markov model* with respect to $\mathrm{G} \in \mathbb{G}^*(V)$ is defined as

$$\mathbb{P}_{\mathrm{UM}}(\mathrm{G}, \mathbb{V}) = \{\mathsf{P} \in \mathbb{P}(\mathbb{V}) : \textbf{not } J \rightsquigarrow K \textbf{ in } \mathrm{G} \Longrightarrow J \perp\!\!\!\perp K \textbf{ under } \mathsf{P} \text{ for all disjoint } J, K \subset V\}.$$

When the graph $\mathrm{G} \in \mathbb{G}_{\mathrm{B}}^*(V)$ is bidirected, this reduces to the *connected set Markov property* in Richardson (2003) which says every connected set (via bidirected edges) is independent of its non-neighbours.

3.3 Ordered local Markov (LM) property

Given an ADMG $\mathrm{G} \in \mathbb{G}_{\mathrm{A}}^*(V)$, we say a strict order $\prec$ on the vertex set V is a *topological order* of G if

$$V_k \longrightarrow V_j \textbf{ in } \mathrm{G} \Longrightarrow V_k \prec V_j \text{ for all } V_j, V_k \in V.$$

An ADMG may have multiple topological orders. Let $\mathrm{pre}_{\prec}(V_j) = \{V_k \in V : V_k \prec V_j\}$ collect all vertices before V_j in the order $\prec$.

Recall that the *Markov boundary* of $V_j \in V$ in $\mathrm{G} \in \mathbb{G}_{\mathrm{A}}^*(V)$ is defined as all vertices that can be connected to V_j via colliders:

$$\mathrm{mb}_{\mathrm{G}}(V_j) = \{V_k \in V : V_k \hookrightarrow * \hookleftarrow V_j \textbf{ in } \mathrm{G}\}.$$

If an ancestral set $K \subseteq V$ contains V_j ($V_j \in K$) but not any children of V_j ($V_j \not\twoheadrightarrow K$ **in** G), the Markov boundary of V_j in K is defined as

$$\mathrm{mb}_{\mathrm{G}}(V_j, K) = \{V_k \in K : V_k \hookleftarrow\!\!\!\!\rightarrow * \hookleftarrow\!\!\!\!\rightarrow V_j \textbf{ in } \mathrm{G}\} = \mathrm{mb}_{\mathrm{G}_K}(V_j) = \mathrm{mbg}_{\mathrm{G}}(V_j) \cap K,$$

where G_K is the subgraph of G restricted to K. The reader is invited to verify the last two equalities.

Definition 5. The *ordered local Markov model* with respect to $\mathrm{G} \in \mathbb{G}_{\mathrm{A}}^*(V)$ and a topological order $\prec$ of G is defined as

$$\mathbb{P}_{\mathrm{LM}}(\mathrm{G}, \prec, \mathbb{V}) = \Big\{ \mathsf{P} \in \mathbb{P}(\mathbb{V}) : V_j \perp\!\!\!\perp K \setminus \mathrm{mb}_{\mathrm{G}}(V_j, K) \setminus V_j \mid \mathrm{mb}_{\mathrm{G}}(V_j, K) \textbf{ under } \mathsf{P} \\ \text{for all } V_j \text{ and ancestral } K \text{ such that } V_j \in K \subseteq \mathrm{pre}_{\prec}(V_j) \Big\}.$$

This definition is due to Richardson (2003, p. 151). It can be shown that the model $\mathbb{P}_{\mathrm{LM}}(\mathrm{G}, \prec, \mathbb{V})$ actually does not depend on which topological order $\prec$ is used. For this reason, we will write it as $\mathbb{P}_{\mathrm{LM}}(\mathrm{G}, \mathbb{V})$.

When G is a DAG (i.e. $\mathrm{G} \in \mathbb{G}_{\mathrm{DA}}^*(V)$), the Markov boundary of $V_j \in V$ reduces to

$$\mathrm{mb}_{\mathrm{G}}(V_j) = \{V_k \in V : V_k \longrightarrow V_j, V_k \longleftarrow V_j, \text{ or } V_k \longrightarrow V_l \longleftarrow V_j \text{ for some } V_l \in V\}.$$

If K is an ancestral set that contains V_j but none of its children, it is easy to see that the Markov boundary of V_j in K is precisely its parents (and thus does not depend on K):

$$\mathrm{mb}_{\mathrm{G}}(V_j, K) = \mathrm{pa}_{\mathrm{G}}(V_j) = \{V_k \in V : V_k \longrightarrow V_j\}.$$

Therefore, the definition of ordered local Markov model for DAGs is consistent with that in Lauritzen (1996, p. 50).

3.4 Factorization (F) and exogenous factorization (EF) properties

Definition 6. For a DAG $\mathrm{G} \in \mathbb{G}^*_{\mathrm{DA}}(V)$, the *factorization model* is defined as

$$\mathbb{P}_{\mathrm{F}}(\mathrm{G}, \mathbb{V}) = \Big\{ \mathsf{P} \in \mathbb{P}(\mathbb{V}) : \mathsf{p}(v) = \prod_{j=1}^{p} \mathsf{p}(v_j \mid v_{\mathrm{pa}_{\mathrm{G}}(j)}) \text{ whenever the right hand side is well defined} \Big\},$$

where $\mathsf{p}(v)$ is the density function of V and $\mathsf{p}(v_j \mid v_{\mathrm{pa}_{\mathrm{G}}(j)})$ is the conditional density function V_j given its parents in G.

Some authors refer to a probability distribution in the above model as a *Bayesian network*, a terminology due to Pearl (1985). Next, we given an extension of this definition to unconfounded ADMGs.

Definition 7. Consider an unconfounded ADMG $\mathrm{G} \in \mathbb{G}^*_{\mathrm{UA}}(V)$ with exogenous vertices $E \subseteq V$. The *exogenous factorization model* with respect to G and E is defined as

$$\begin{aligned}\mathbb{P}_{\mathrm{EF}}(\mathrm{G}, \mathbb{V}) = \Big\{ \mathsf{P} \in \mathbb{P}(\mathbb{V}) : \mathsf{p}(v) = \mathsf{p}(e) \prod_{V_j \in V \setminus E} \mathsf{p}(v_j \mid v_{\mathrm{pa}_{\mathrm{G}}(j)}) &\text{ whenever well defined,} \\ \mathrm{margin}_E(\mathsf{P}) \in \mathbb{P}_{\mathrm{GM}}(\mathrm{margin}_E(\mathrm{G}), \mathrm{margin}_E(\mathbb{V})) &\Big\},\end{aligned}$$

where $\mathsf{p}(e) = \mathrm{margin}_E(\mathsf{p})$ is the marginal density function of E.

It is easy to see that $\mathbb{P}_{\mathrm{EF}}(\mathrm{G}, \mathbb{V}) = \mathbb{P}_{\mathrm{F}}(\mathrm{G}, \mathbb{V})$ if $\mathrm{G} \in \mathbb{G}^*_{\mathrm{DA}}(V)$ and $\mathbb{P}_{\mathrm{EF}}(\mathrm{G}, \mathbb{V}) = \mathbb{P}_{\mathrm{GM}}(\mathrm{G}, \mathbb{V})$ if $\mathrm{G} \in \mathbb{G}^*_{\mathrm{B}}(V)$. So exogenous factorization is a concept that genearlizes factorization with respec to DAGs and global Markov property with respect to bidirected graphs. The requirement that the marginal distribution of E is global Markov is not essential and can be replaced by other equivalent definitions (see Figure 1d).

3.5 Nested Markov (NM) property

The nested Markov property is defined through the *fixing* (or *do*) operator that applies to product spaces, probability distributions and graphs (Richardson, Evans, et al. 2023). First, for any $J \subseteq V$, $\mathrm{do}_J(\mathbb{V}) = \mathrm{margin}_J(\mathbb{V})$ is simply the subspace corresponding to J. Given $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$, we say $V_j \in V$ is *fixable* in G if there does not exist $V_k \in V$ such that $V_j \rightsquigarrow V_k$ and $V_j \longleftrightarrow * \longleftrightarrow V_k$ **in** G. When fixing operator acts on a graph $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$, it means that we remove a fixable vertex V_j in G and all its edges, that is,

$$\begin{aligned}\mathrm{do}_{V_j} : \mathbb{G}^*_{\mathrm{A}}(V) &\to \mathbb{G}^*_{\mathrm{A}}(V_{-j}), \\ \mathrm{G} &\mapsto \mathrm{G}_{V_{-j}},\end{aligned}$$

where $\mathrm{G}_{V_{-j}}$ is the subgraph on $V_{-j} = V \setminus \{V_j\}$.

Finally, we define the fixing operator on probability distribiutions. For any fixable $V_j \in V$ and $v_j \in \mathbb{V}_j$, the fixing operator $\mathrm{do}_{V_j = v_j} : \mathbb{P}(\mathbb{V}) \to \mathbb{P}(\mathbb{V}_{-j})$ is defined as the following transformation of the density function:

$$(\mathrm{do}_{V_j = v_j}(\mathsf{p}))(v_{-j}) = \frac{\mathsf{p}(v)}{\mathsf{p}(v_j \mid v_{\mathrm{mbg}_{\mathrm{G}}(j)})}.$$

It is easy to verify that the image is indeed a density function for V_{-j} (non-negative and integrates to 1) that is indexed by $v_j \in \mathbb{V}_j$.[6] We deliberately denoted the fixing operator as $\mathrm{do}_{V_j=v_j}$ because it corresponds to identifying the interventional distribution of V_{-j} when V_j is set to v_j; see Proposition 3 below.

Now consider a sequence of distinct vertices $J = V_{\mathcal{J}} = (V_{j_1}, \dots, V_{j_n})$. Denote

$$\mathrm{do}_J = \mathrm{do}_{V_{j_1}} \circ \cdots \circ \mathrm{do}_{V_{j_n}},$$

which can be applied to product spaces, graphs, and probability distributions (after further specifying the value that $V_{\mathcal{J}}$ is fixed at). We say J is fixable if V_{j_m} is fixable in $\mathrm{do}_{V_{j_1}} \circ \cdots \circ \mathrm{do}_{V_{j_{m-1}}}(\mathrm{G})$ for all $m = 1, \dots, n$. Not all permutations of J are fixable, but all fixable permutations of J define the same fixing operator as shown by Richardson, Evans, et al. (2023); see also the remark at the end of Section 4.3. So with a slight abuse of notation, do_J can also be defined for any (unordered) subset $J \subseteq V$ that has at least one fixable permutation. We use the convention that $\mathrm{do}_\emptyset(\mathsf{P}) = \mathsf{P}$. Finally, the fixing operator on state spaces is "dual" to marginalization in the sense that $\mathrm{do}_{V_{\mathcal{J}}}(\mathbb{V}) = \mathrm{margin}_{V \setminus V_{\mathcal{J}}}(\mathbb{V})$.

The nested Markov model in Richardson, Evans, et al. (2023) requires that the probability distribution, after fixing, is global Markov with respect to the fixed graph.

Definition 8. The *nested Markov model* with respect to $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$ is defined as

$$\begin{aligned}&\mathbb{P}_{\mathrm{NM}}(\mathrm{G}, \mathbb{V})\\ =&\{\mathsf{P} \in \mathbb{P}(\mathbb{V}) : \mathrm{do}_{V_{\mathcal{J}}=v_{\mathcal{J}}}(\mathsf{P}) \in \mathbb{P}_{\mathrm{GM}}(\mathrm{do}_J(\mathrm{G}), \mathrm{do}_J(\mathbb{V})) \text{ for all fixable } J = V_{\mathcal{J}} \subseteq V \text{ and } v_{\mathcal{J}} \in \mathbb{V}_{\mathcal{J}}\}.\end{aligned}$$

3.6 Augmentation (A) criterion

The augmentation criterion links statistical models associated with directed graphs with those associated with undirected graphs. To this end, let us introduce some additional notation. Let $\mathbb{UG}(V)$ denote the collection of all simple undirected graphs with vertex set V; specifically, $\mathbb{UG}(V)$ contains all graphs $\mathrm{G}' = (V, \mathcal{E})$ such that $\mathcal{E} \subseteq V \times V$, $(V_j, V_j) \notin \mathcal{E}$, and $(V_j, V_k) \in \mathcal{E}$ implies that $(V_k, V_j) \in \mathcal{E}$ for all $V_j, V_k \in V$. This definition is not different from a bidirected graph besides the requirement of no self-loops, but the semantics of undirected and bidirected graphs (in terms of graph separation) are different. Specifically, for $\mathrm{G}' \in \mathbb{UG}(V)$ and disjoint subsets $J, K, L \subset V$, we say L *separate* J and K in G' and write

$$\textbf{not } J \mathrel{—\!\!*\!\!—} K \mid L \textbf{ in } \mathrm{G}',$$

if every path from a vertex in J to a vertex in K contains an non-endpoint in L. The global Markov model associated with an undirected graph $\mathrm{G}' \in \mathbb{UG}(V)$ is defined as

$$\begin{aligned}\mathbb{P}_{\mathrm{GM}}(\mathrm{G}', \mathbb{V}) = \{\mathsf{P} \in \mathbb{P}(\mathbb{V}) : \textbf{not } J \mathrel{—\!\!*\!\!—} K \mid L \textbf{ in } \mathrm{G}' \Longrightarrow J \perp\!\!\!\perp K \mid L \textbf{ under } \mathsf{P}\\ \text{for all disjoint } J, K, L \subset V\}.\end{aligned}$$

Consider the following *augmentation* map from directed mixed graphs to undirected graphs:

$$\begin{aligned}\mathrm{augment} : \mathbb{G}^*(V) &\to \mathbb{UG}(V),\\ \mathrm{G} &\mapsto \mathrm{G}',\end{aligned}$$

[6] Fixing is well defined whenever $\mathsf{p}(v_j \mid v_{\mathrm{mbg}_{\mathrm{G}}(j)})$ is not 0 or ∞. An argument similar to that in Pollard (2001, Theorem 5.12) shows that such event has probability 0 and thus is inconsequential in defining the density function of the probability distribution after fixing.

where $\mathrm{G}' = \mathrm{augment}(\mathrm{G})$ is an undirected graph with the same vertex set V such that

$$V_j \,\text{—}\, V_k \textbf{ in } \mathrm{G}' \iff V_j \longleftrightarrow * \longleftrightarrow V_k \textbf{ in } \mathrm{G} \text{ for all } V_j, V_k \in V, V_j \neq V_k.$$

That is, V_j is connected to all vertices in its Markov boundary. When this map is restricted to DAGs, this is often known as *moralization* in the literature because it connects any two parents with the same child (Lauritzen and Wermuth 1989; Frydenberg 1990). For ADMGs, the augmentation criterion below is introduced in Richardson (2003).

Definition 9. The augmentation model for $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$ is defined as

$$\begin{aligned} &\mathbb{P}_{\mathrm{A}}(\mathrm{G}, \mathbb{V}) \\ =&\{\mathsf{P} \in \mathbb{P}(\mathbb{V}) : \mathrm{margin}_J(\mathsf{P}) \in \mathbb{P}_{\mathrm{GM}}(\mathrm{augment} \circ \mathrm{margin}_J(\mathrm{G}), \mathrm{margin}_J(\mathbb{V})) \text{ for all ancestral } J \subseteq V\}. \end{aligned}$$

3.7 Pairwise (PE), clique (CE), and noise (NE) expansions

One way to define statistical models associated with a general ADMG is through expanding the graph to “simpler graphs”. First, let us define graph expansion, which is simply the pre-image of graph marginalization. Specifically, given $\mathrm{G} \in \mathbb{G}^*(V)$, define

$$\mathrm{expand}(\mathrm{G}) = \{\mathrm{G}' \in \mathbb{G}^*(V') : V' \supseteq V, \mathrm{margin}_V(\mathrm{G}') = \mathrm{G}\}.$$

Obviously, graph marginalization is not injective, so $\mathrm{expand}(\mathrm{G})$ is an infinite set of graphs that can marginalize to G.

There are several possible “natural” definitions that pick a specific element of $\mathrm{expand}(\mathrm{G})$ as “the” expanded graph. Consider $V = \{V_1, \ldots, V_d\}$ and $\mathrm{G} \in \mathbb{G}^*(V)$. The *pairwise expansion* replaces every bidirected edge by a latent common parent. Formally, the pairwise expansion graph $\mathrm{expand}_{\mathrm{P}}(\mathrm{G})$ has vertex set $V \cup E$ with $E = \{E_{jk} : V_j \longleftrightarrow V_k \textbf{ in } \mathrm{G}, j < k\}$ and the following edges:

$$\begin{aligned} E_{jk} \longrightarrow V_j &\textbf{ in } \mathrm{expand}_{\mathrm{P}}(\mathrm{G}), \text{ for all } j, k \in [d] \text{ such that } V_j \longleftrightarrow V_k \textbf{ in } \mathrm{G}, \\ V_j \longrightarrow V_k &\textbf{ in } \mathrm{expand}_{\mathrm{P}}(\mathrm{G}), \text{ for all } j, k \in [d] \text{ such that } V_j \longrightarrow V_k \textbf{ in } \mathrm{G}\,. \end{aligned}$$

The *clique expansion* replaces every bidirected clique (in which every two vertices are connected by a bidirected edge) by a latent common parent. Formally, if we let $\mathcal{C}(\mathrm{G})$ denote (the vertex indicies of) all bidirected cliques in G, that is,[7]

$$\mathcal{C}(\mathrm{G}) = \{\mathcal{J} \subseteq 2^{[d]} : V_j \longleftrightarrow V_k \text{ for all } j, k \in \mathcal{J}\},$$

then the clique expansion graph $\mathrm{expand}_{\mathrm{C}}(\mathrm{G})$ has vertex set $V \cup E$ with $E = \{E_{\mathcal{J}} : \mathcal{J} \in \mathcal{C}(\mathrm{G})\}$ and the following edges:

$$\begin{aligned} E_{\mathcal{J}} \longrightarrow V_j &\textbf{ in } \mathrm{expand}_{\mathrm{C}}(\mathrm{G}), \text{ for all } j \in \mathcal{J} \in \mathcal{C}(\mathrm{G}), \\ V_j \longrightarrow V_k &\textbf{ in } \mathrm{expand}_{\mathrm{C}}(\mathrm{G}), \text{ for all } j, k \in [d] \text{ such that } V_j \longrightarrow V_k \textbf{ in } \mathrm{G}\,. \end{aligned}$$

It is easy to see that pairwise and clique expansion graphs are DAGs.

[7] One can also define bidirected cliques as the *maximal* sets connected by bidirected edges in the graph (i.e. “districts” in the terminology of Richardson (2003)), but that does not change the clique expansion model. The definition employed here simplifies our proof in the Appendix that the clique expansion model is complete (particularly Lemma 8).

The *noise expansion*, on the other hand, results in an unconfounded graph where the bidirected and directed edges are "separated". Formally, the noise expansion graph $\mathrm{expand_N}(\mathrm{G})$ has vertex set $V \cup E$ with $E = \{E_1, \dots, V_d\}$ and the following edges:

$$
\begin{aligned}
E_j &\longrightarrow V_j \textbf{ in } \mathrm{expand_N}(\mathrm{G}), \text{ for all } j \in [d],\\
E_j &\longleftrightarrow E_k \textbf{ in } \mathrm{expand_N}(\mathrm{G}), \text{ for all } j,k \in [d] \text{ such that } V_j \longleftrightarrow V_k \textbf{ in } \mathrm{G},\\
V_j &\longrightarrow V_k \textbf{ in } \mathrm{expand_N}(\mathrm{G}), \text{ for all } j,k \in [d] \text{ such that } V_j \longrightarrow V_k \textbf{ in } \mathrm{G}\,.
\end{aligned}
$$

Definition 10. For $\mathrm{G} = (V, \mathcal{B}, \mathcal{D}) \in \mathbb{G}_{\mathrm{A}}^*(V)$, the *pairwise expansion model*, *clique expansion model*, and *clique expansion model* are defined as the V-marginal of the global Markov model for the corresponding expanded graphs:

$$
\begin{aligned}
\mathbb{P}_{\mathrm{PE}}(\mathrm{G}, \mathbb{V}) &= \mathrm{margin}_V \Big(\mathbb{P}_{\mathrm{GM}}\big(\mathrm{expand_P}(\mathrm{G}), \mathbb{V} \times [0,1]^{|\mathcal{B}|} \big) \Big),\\
\mathbb{P}_{\mathrm{CE}}(\mathrm{G}, \mathbb{V}) &= \mathrm{margin}_V \Big(\mathbb{P}_{\mathrm{GM}}\big(\mathrm{expand_C}(\mathrm{G}), \mathbb{V} \times [0,1]^{|\mathcal{C}(\mathrm{G})|} \big) \Big),\\
\mathbb{P}_{\mathrm{NE}}(\mathrm{G}, \mathbb{V}) &= \mathrm{margin}_V \Big(\mathbb{P}_{\mathrm{GM}}\big(\mathrm{expand_N}(\mathrm{G}), \mathbb{V} \times [0,1]^{|V|} \big) \Big).
\end{aligned}
$$

This definition assumes that the latent variables are supported in the unit interval, which is large enough for most purposes.

3.8 Nonparametric equation (E) systems

Definition 11 (Nonparametric system)**.** Consider $\mathrm{G} \in \mathbb{G}_{\mathrm{A}}^*(V)$. The *nonparametric equation system* $\mathbb{P}_{\mathrm{E}}(\mathrm{G}, \mathbb{V})$ collects all probability distribution $\mathsf{P} \in \mathbb{P}(\mathbb{V})$ on a random vector $V = (V_1, \dots, V_d)$ such that the following event has probability 1 under P: V solves the equations

$$
V_j = f_j(V_{\mathrm{pa_G}(j)}, E_j),\ j = 1, \dots, d \tag{3}
$$

for some (measurable) functions $f_j : \mathbb{V}_{\mathrm{pa_G}(j)} \times [0,1] \to \mathbb{V}_j$, $j = 1, \dots, d$ and random vector $E = (E_1, \dots, E_d) \in [0,1]^d$ whose joint distribution Q is unconditionally Markov with respect to the bidirected component of G, that is, for all disjoint $\mathcal{J}, \mathcal{K} \subset [d]$, we have

$$
V_{\mathcal{J}} \not\leftrightarrow V_{\mathcal{K}} \textbf{ in } \mathrm{G} \Longrightarrow E_{\mathcal{J}} \perp\!\!\!\perp E_{\mathcal{K}} \textbf{ under } \mathsf{Q}\,. \tag{4}
$$

This definition is closely related to the "semi-Markovian" causal model in Pearl (2009, p. 30), but there are some subtle distinctions. First, Pearl does not explicitly state (4) as the Markov condition on the distribution of the noise variables and just calls the model semi-Markovian if the noises are correlated. In another definition of semi-Markovian models, Bareinboim et al. (2022, p. 542-543) define its causal diagram by adding a bidirected edge between V_j and V_k if the corresponding noise variables are correlated. However, equation (4) is stronger: it further requries the pairwise independence relationships can be combined (so the conditional independences form a compositional semi-graphoid). Second, Pearl intends to interpret (3) as a causal model. However, to formalize a causal models, it is not enough to consider the variables V themselves. One needs to further consider the potential outcome of V under interventions; see Section 4 below.

4 Causal Markov model and the nested Markov property

The nonparametric equation system gives a natural definition of potential outcomes using recursive substitution (Pearl 2009; Richardson and Robins 2013). In this Section, we will introduce this causal model and use it to prove that the nonparametric equation system is nested Markov as formally stated below.

Theorem 3. *For* $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$ *and any product space* $\mathbb{V}$*, we have* $\mathbb{P}_{\mathrm{E}}(\mathrm{G}, \mathbb{V}) \subseteq \mathbb{P}_{\mathrm{NM}}(\mathrm{G}, \mathbb{V})$*. In other words, the implication* $E \Rightarrow NM$ *in Figure 1a holds.*

4.1 The causal Markov model

Consider an ADMG $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$ and a nonparametric equation system as given in Definition 11. Roughly speaking, we can interpret the equations in (3) causally by requiring that those equations still hold in an intervention that sets some of the variables to a fixed value. To formalize this heuristic, let us first define what we mean by the causal model associated with an ADMG.

Let the random variable $V_j(v_{\mathcal{I}})$ denote the *potential outcome* of V_j under an intervention that sets $V_{\mathcal{I}}$ to $v_{\mathcal{I}}$, $j \in [d]$, $\mathcal{I} \subseteq [d]$. The *potential outcome schedule* is the collection of all potential outcomes:

$$V(\cdot) = (V_j(v_{\mathcal{I}}) : j \in [d], \mathcal{I} \subseteq [d], v_{\mathcal{I}} \in \mathbb{V}_{\mathcal{I}}).$$

Let $\mathbb{CP}(\mathbb{V})$ denote the collection of all probability distributions P on $V(\cdot)$ such that $V_j(v_{\mathcal{I}})$ takes value in $\mathbb{V}_j$, that is, $\mathrm{margin}_{V_j(v_{\mathcal{I}})}(\mathsf{P}) \in \mathbb{P}(\mathbb{V}_j)$. To avoid cluttering, we will denote $\mathrm{pa}_{\mathrm{G}}(j)$ as $\mathrm{pa}(j)$ below.

Definition 12 (Causal Markov model)**.** We say $\mathsf{P} \in \mathbb{CP}(\mathbb{V})$ is *causal Markov* with respect to $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$ if the following are true:

1. The potential outcomes are consistent with each other in the sense that the next event has probability 1 under P:
$$V_j(v_{\mathcal{I}}) = V_j(v_{\mathrm{pa}(j)\cap\mathcal{I}}, V_{\mathrm{pa}(j)\setminus\mathcal{I}}(v_{\mathcal{I}})), \text{ for all } j \in [d], \mathcal{I} \subseteq [d], v \in \mathbb{V}. \tag{5}$$

2. The distribution of the basic potential outcomes is unconditionally Markov with respect to the bidirected component of G, that is, for all disjoint $\mathcal{J}, \mathcal{K} \subset [d]$, we have
$$V_{\mathcal{J}} \not\leftrightarrow V_{\mathcal{K}} \textbf{ in } \mathrm{G} \Longrightarrow V_{\mathcal{J}}(v) \perp\!\!\!\perp V_{\mathcal{K}}(v) \textbf{ under } \mathsf{P} \text{ for all } v \in \mathbb{V}. \tag{6}$$

The *causal Markov model* associated with G is then defined as

$$\mathbb{CP}(\mathrm{G}, \mathbb{V}) = \{\mathsf{P} \in \mathbb{CP}(\mathbb{V}) : \mathsf{P} \text{ is causal Markov with respect to G}\}.$$

This definition generalizes the single-world causal model introduced by Richardson and Robins (2013) in two ways: first, the causal diagram can be an ADMG instead of just a DAG; second, this definition does not use structural equations to define potential outcomes.

Note that the directed and bidirected edges play different epistemic roles in this definition. The directed edges represent direct causal effects, and the bidirected edges represent exogenous correlation. Importantly, this model does not assume that the exogenous correlations arise from latent common causes. In the author's opinion, this is more transparent than the approach in Richardson, Evans, et al. (2023, Section 4.3) and implicitly taken in Pearl's work that assumes a *causal* model with respect to some unspecified DAG expansion of the ADMG. It is difficult to conceptualize potential outcomes of the latent variables without knowing what they are. In contrast, the causal Markov model above only requires potential outcomes of the variables in the ADMG.

The equations in the E model (see Definition 11) give a natural definition of potential outcomes via the following recursion:

$$V_j(v_{\mathcal{I}}) = f_j(v_{\mathrm{pa}(j)\cap\mathcal{I}}, V_{\mathrm{pa}(j)\setminus\mathcal{I}}(v_{\mathcal{I}}), E_j),\ j = 1,\dots,d. \tag{7}$$

The distribution of the potential outcome schedule is then entirely determined by the functions $f_1,\dots,f_d$ and the distribution of the noise variables $E_1,\dots,E_d$. This is often referred to as the *structural equation model* (Pearl 2009) or *structural causal model* (Peters, Janzing, and Schölkopf 2017; Bareinboim et al. 2022), although the assumption on the distribution of E is not always clearly stated. The distribution of potential outcomes defined via (7) is causal Markov with respect to G: (5) immediately follows from (7), and (6) immediately follows from (4).[8] We summarize this observation as a Lemma.[9]

Lemma 1. *For any* $\mathrm{G} \in \mathbb{G}_{\mathrm{A}}^*(V)$ *and product space* $\mathbb{V}$*, we have* $\mathbb{P}_{\mathrm{E}}(\mathrm{G},\mathbb{V}) \subseteq \mathrm{margin}_V(\mathbb{CP}(\mathrm{G},\mathbb{V}))$.

4.2 Properties of the causal Markov model

We will next introduce three key properties of the causal Markov model and use them to prove Theorem 3. The first property, as summarized by the next Proposition, is often referred to as "consistency of potential outcomes" in the causal inference literature.

Proposition 1. *Suppose* $\mathsf{P} \in \mathbb{CP}(\mathrm{G},\mathbb{V})$ *for some* $\mathrm{G} \in \mathbb{G}_{\mathrm{A}}^*(V)$*. Then for all disjoint* $V_{\mathcal{I}}, V_{\mathcal{I}'} \subset V$*, we have*

$$\mathsf{P}(V(v_{\mathcal{I}}, v_{\mathcal{I}'}) = V(v_{\mathcal{I}}) \mid V_{\mathcal{I}'}(v_{\mathcal{I}}) = v_{\mathcal{I}'}) = 1,\ \textit{for all}\ v \in \mathbb{V}.$$

The second property is that the causal Markov property implies the global Markov property at different "levels" of the potential outcomes. To formally describe this, let us generalize the definition of single world intervention graphs (SWIGs) in Richardson and Robins (2013) from DAGs to ADMGs. Given $\mathrm{G} \in \mathbb{G}_{\mathrm{A}}^*(V)$, let $\mathrm{G}(v_{\mathcal{I}})$ denote the graph obtained by removing all outgoing edges from $V_{\mathcal{I}}$ (i.e. edges like $V_{\mathcal{I}} \longrightarrow *$) and relabelling V_j as $V_j(v_{\mathcal{I}})$ for all $V_j \in V$.[10] Let V_{-j} denote the complement of V_j in V and $V_{-\mathcal{J}}$ denote the complement of $V_{\mathcal{J}}$.

The next Proposition generalizes similar results for DAGs in the literature, for example, Theorem 1.4.1 in Pearl (2009) and Proposition 11 in Richardson and Robins (2013).

[8] Note that (4) implies more than (6): the conditional independence $V_{\mathcal{J}}(v) \perp\!\!\!\perp V_{\mathcal{K}}(v') \mid V_{\mathcal{L}}(v'')$ is also true for all $v, v', v'' \in \mathbb{V}$ that are not the same. These "cross-world" assumptions, however, are not possible to verify by any experiment (Richardson and Robins 2013).

[9] Note that our definition of causal Markov model is a collection of probability distributions on the potential outcomes schedule and does not require defining potential outcomes via structural equations. It is natural to ask if this is indeed more general, that is, whether the reverse of Lemma 1 is true. It is observed in Richardson and Robins (2013, p. 22) that one can use the potential outcomes to define structural equations as

$$f_j(v_{\mathrm{pa}(j)}, E_j) = V_j(v_{\mathrm{pa}(j)}),\ j = 1,\dots,d,$$

where $E_j = (V_j(v_{\mathrm{pa}(j)}) : v_{\mathrm{pa}(j)} \in \mathbb{V}_{\mathrm{pa}(j)})$ collects all basic potential outcomes for V_j. However, the range of E_j is $\mathbb{V}_j^{\mathbb{V}_{\mathrm{pa}(j)}}$, whose cardinality is not always the same as that of $[0,1]$ (i.e. the continuum). Furthermore, independence of the "noise" in (4) does not directly follow from single-world independence of the potential outcomes in (6).

[10] We do not consider the "fixed vertex" v_i for $i \in \mathcal{I}$ as in Richardson and Robins (2013), because we are only interested in the distribution of $V(v_{\mathcal{I}})$ here.

Proposition 2. *Suppose* $\mathsf{P} \in \mathbb{CP}(\mathrm{G}, \mathbb{V})$ *for some* $\mathrm{G} \in \mathbb{G}_{\mathrm{A}}^*(V)$*. Then* $\mathrm{margin}_{V(v_{\mathcal{I}})}(\mathsf{P}) \in \mathbb{P}_{\mathrm{GM}}(\mathrm{G}(v_{\mathcal{I}}), \mathbb{V})$ *for all* $V_{\mathcal{I}} \subseteq V$ *and* $v \in \mathbb{V}$*.*

The third property establishes the connection between fixability and causal identification. Mathematically speaking, causal identification refers to injectivity of the map $\mathrm{margin}_V : \mathbb{CP}(\mathrm{G}, \mathbb{V}) \to \mathbb{P}(\mathbb{V})$, that is, it asks whether we can determine the distribution of the potential outcomes schedule from the distribution of the observed outcomes. The next Proposition shows that if a vertex V_j is fixable in G, then the distribution of $V(v_j)$ can be identified.

Proposition 3. *Suppose* $\mathsf{P} \in \mathbb{CP}(\mathrm{G}, \mathbb{V})$ *for some* $\mathrm{G} \in \mathbb{G}_{\mathrm{A}}^*(V)$*. If* $V_j \in V$ *is fixable in* G*, then*

$$\frac{\mathsf{p}(V_j(v_j) = \tilde{v}_j, V_{-j}(v_j) = v_{-j})}{\mathsf{p}(V_j = v_j, V_{-j} = v_{-j})} = \frac{\mathsf{p}(V_j = \tilde{v}_j \mid V_{\mathrm{mbg}(j)} = v_{\mathrm{mbg}(j)})}{\mathsf{p}(V_j = v_j \mid V_{\mathrm{mbg}(j)} = v_{\mathrm{mbg}(j)})}, \text{ for all } v \in \mathbb{V} \text{ and } v_j^* \in \mathbb{V}_j. \quad (8)$$

The proof of Propositions 1 to 3 can be found in the Appendix.

4.3 Proof of Theorem 3

Next, we use the above properties of the causal Markov model to prove Theorem 3.

Consider any $\mathsf{P}_V \in \mathbb{P}_{\mathrm{E}}(\mathrm{G}, \mathbb{V})$. By Lemma 1, there exists $\mathsf{P} \in \mathbb{CP}(\mathrm{G}, \mathbb{V})$ such that $\mathrm{margin}_V(\mathsf{P}) = \mathsf{P}_V$. Consider any fixable vertex V_j in G. By rewritting equation (8) in Proposition 3 as

$$\mathsf{p}(V_j(v_j) = \tilde{v}_j, V_{-j}(v_j) = v_{-j}) = (\mathrm{do}_{V_j = v_j}(\mathsf{p}_V))(v_{-j}) \cdot \mathsf{p}(V_j = \tilde{v}_j \mid V_{\mathrm{mbg}(j)} = v_{\mathrm{mbg}(j)}),$$

we find, after marginalizing out $\tilde{v}_j$, that

$$\mathsf{p}(V_{-j}(v_j) = v_{-j}) = (\mathrm{do}_{V_j = v_j}(\mathsf{p}_V))(v_{-j}).$$

By using Proposition 2 and the fact that the subgraph of $\mathrm{G}(v_{\mathcal{I}})$ on $V_{-j}(v_{\mathcal{I}})$ is isomorphic to the subgraph of G on V_{-j}, this shows that

$$\mathrm{do}_{V_j = v_j}(\mathsf{P}_V) = \mathrm{margin}_{V_{-j}(v_j)}(\mathsf{P}) \in \mathbb{P}_{\mathrm{GM}}(\mathrm{margin}_{V_{-j}(v_j)}(\mathrm{G}(v_j)), \mathbb{V}_{-j}) = \mathbb{P}_{\mathrm{GM}}(\mathrm{do}_{V_j}(\mathrm{G}), \mathrm{do}_{V_j}(\mathbb{V})).$$

By applying this argument repeatedly, we see that, for any fixable sequence $V_{\mathcal{J}}$ and $v_{\mathcal{J}} \in \mathbb{V}_{\mathcal{J}}$,

$$\mathrm{do}_{V_{\mathcal{J}} = v_{\mathcal{J}}}(\mathsf{P}_V) = \mathrm{margin}_{V_{-\mathcal{J}}(v_{\mathcal{J}})}(\mathsf{P}) \in \mathbb{P}_{\mathrm{GM}}(\mathrm{do}_{V_{\mathcal{J}}}(\mathrm{G}), \mathrm{do}_{V_{\mathcal{J}}}(\mathbb{V})). \quad (9)$$

This shows that $\mathsf{P}_V \in \mathbb{P}_{\mathrm{NM}}(\mathrm{G}, \mathbb{V})$ and completes our proof of Theorem 3.

Note that (9) immediately implies that the order of fixing does not matter, that is, when fixing is applied sequentially for two different fixable permutations of the same subset of variables, the results are the same.

5 Conclusion and discussion

The conclusion of this article can be summarized in one simple sentence:

Use ADMGs, not DAGs, for agnostic causal inference.

Of course, it is not new to use ADMGs for causal inference. After all, Wright (1934) have used them nearly a century ago because two types of edges are needed to describe two different types of dependence (causal and statistical correlation) in a linear structural equation model, and this tradition is kept in social science; see e.g. Bollen (1989) and the popular LISREL software (Jöreskog and Sörbom 2018). Moreover, ADMGs are used in the groundbreaking do-calculus (Pearl 1995, 2009) and the ID algorithm for causal identification (Tian and Pearl 2002; Richardson, Evans, et al. 2023). But the point here is different: we believe that causal inference can and should be entirely based on ADMGs.

We next break down the supporting arguments for this conclusion into theoretical and practical considerations.

Theoretically, as is argued in this article, the ADMG-based E/NE model is more "natural" than the DAG-based CE model. Both of them are complete (with respect to unconfounded graph expansions) in the sense given in Definition 2, but the E/NE model is more natural in unconfounded graphs as shown by applying the Equivalence argument to Figure 1b. In particular, although most statisticians will agree that the unconditional Markov (UM) model is too large to be interesting, the CE model does not show superiority over the UM model through the lens of Theorems 1 and 2:

1. Both of the CE and UM models are complete with respect to unconfounded graph expansions;
2. The CE model is equivalent to GM and many other models in DAGs, while the UM model is equivalent to GM and many other models in bidirected graphs.

Thus, by commiting to the CE model and a DAG-based theory for causality, one dismisses the class of bidirected graphs as an interesting primitive object to study. This commitment results in many complicated inequality constraints on the probability distribution that are difficult to characterize (Fritz 2012; Evans 2016).

Practically, using ADMGs instead of DAGs (with possible latent variables) allow applied statisticians to focus on the variables being investigated and the variables that can potentially be measured in their study. Moreover, practitioners do not need to justify why any bidirected edge is assumed in the graph, because exactly why two variables are exogenously correlated is not crucial for causal identification (through the do-calculus or ID algorithm). Rather, practitioners should focus on defending the lack of bidirected or directed edges between some variables, which is why causal identification is possible. By using ADMGs and drawing bidirected edges, practitioners are instinctively encouraged to think about the missing bidirected edges. For example, this approach is taken in Guo and Zhao (2023) who develop a new procedure for confounder selection by iteratively expanding possible bidirected edges in the graph.

Of course, when there are good reasons to believe two variables have a common cause, practitioners are still encouraged to include it in the graph even if the common cause cannot be measured. Latent mixture models can still be used if they are deemed reasonable for the specific problem, and alternative identification strategies such as those using proxies of the unmeasured common causes remain useful (see e.g. Tchetgen et al. 2024). Moreover, the E/NE model is really just a small modification of the CE model: it is not hard to show that they are equivalent when the bidirected edges can be partitioned into multiple cliques.

So what is all the fuss about? What the article really criticizes is the following interpretation of ADMGs that is commonly found in verbal communications about causal graphs:

> ADMG is just a convenient shortcut to represent some unspecified large causal DAG that generate the data.

This interpretation is mathematically unnecessary to define a structural causal model, makes obscure ontological assumptions, and tends to discourage practitioners from deliberating over the real assumptions in the graph (i.e. missing edges).

There are some important open problems to consider in future work. First, it would be interesting to understand the inequality constraints implied by the E/NE model, in addition to the equality constraints in the nested Markov model. Second, ADMG can also be used to describe quantum mechanics models, which are also submodels of the nested Markov model (Navascués and Wolfe 2020). A quick investigation shows that the E/NE model does not contain nor is constainted by the quantum mechanics model: the E/NE model has a more relaxed interpretation of bidirected graphs but a local interpretation of directed edges. It would be interesting to study their relations further and consider super-models that contain both of them. Third, many modern causal inference methods use graphical diagrams to identify the causal estimands of interest and then estimate those parameters using influence-function based methods. These methods typically require pathwise differentiability of the estimands within the model, and it would be interesting to study that for the E/NE model defined here.

Acknowledgement

This work is in part supported by the Engineering and Physical Sciences Research Council (grant number EP/V049968/1). The author thanks Wenjie Hu for discussion on the causal Markov model for ADMGs and Thomas Richardson for constructive feedback on an earlier draft.

References

Balke, Alexander and Judea Pearl (Sept. 1997). "Bounds on Treatment Effects from Studies with Imperfect Compliance". In: *Journal of the American Statistical Association* 92.439, pp. 1171–1176. ISSN: 0162-1459. DOI: 10.1080/01621459.1997.10474074.

Bareinboim, Elias et al. (Mar. 2022). "On Pearl's Hierarchy and the Foundations of Causal Inference". In: *Probabilistic and Causal Inference: The Works of Judea Pearl*. 1st ed. Vol. 36. New York, NY, USA: Association for Computing Machinery, pp. 507–556. ISBN: 978-1-4503-9586-1.

Bollen, Kenneth A. (Apr. 1989). *Structural Equations with Latent Variables*. 1st ed. Wiley. ISBN: 978-0-471-01171-2 978-1-118-61917-9. DOI: 10.1002/9781118619179.

Evans, Robin J. (2016). "Graphs for Margins of Bayesian Networks". In: *Scandinavian Journal of Statistics* 43.3, pp. 625–648. ISSN: 1467-9469. DOI: 10.1111/sjos.12194.

Fritz, Tobias (Oct. 2012). "Beyond Bell's Theorem: Correlation Scenarios". In: *New Journal of Physics* 14.10, p. 103001. ISSN: 1367-2630. DOI: 10.1088/1367-2630/14/10/103001.

Frydenberg, Morten (1990). "The Chain Graph Markov Property". In: *Scandinavian Journal of Statistics* 17.4, pp. 333–353. ISSN: 0303-6898. JSTOR: 4616181.

Guo, F. Richard and Qingyuan Zhao (Oct. 2023). *Confounder Selection via Iterative Graph Expansion*. DOI: 10.48550/arXiv.2309.06053. arXiv: 2309.06053 [math, stat].

Jöreskog, K. G. and D. Sörbom (2018). *LISREL 10 for Windows*. Scientific Software International, Inc. Skokie, IL.

Kiiveri, Harri, T. P. Speed, and J. B. Carlin (Feb. 1984). "Recursive Causal Models". In: *Journal of the Australian Mathematical Society* 36.1, pp. 30–52. ISSN: 0263-6115. DOI: 10.1017/S1446788700027312.

Lauritzen, S. L. and N. Wermuth (Mar. 1989). “Graphical Models for Associations between Variables, Some of Which Are Qualitative and Some Quantitative”. In: *The Annals of Statistics* 17.1, pp. 31–57. ISSN: 0090-5364, 2168-8966. DOI: `10.1214/aos/1176347003`.

Lauritzen, Steffen L. (1996). *Graphical Models*. Oxford Statistical Science Series. Oxford: Clarendon Press.

Navascués, Miguel and Elie Wolfe (Jan. 2020). “The Inflation Technique Completely Solves the Causal Compatibility Problem”. In: *Journal of Causal Inference* 8.1, pp. 70–91. ISSN: 2193-3685. DOI: `10.1515/jci-2018-0008`.

Pearl, Judea (1985). “Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning”. In: *Proceedings of the 7th Conference of the Cognitive Science Society*, pp. 329–334.

— (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. ISBN: 1-55860-479-0.

— (Dec. 1995). “Causal Diagrams for Empirical Research”. In: *Biometrika* 82.4, pp. 669–688. ISSN: 0006-3444. DOI: `10.1093/biomet/82.4.669`.

— (2009). *Causality*. 2nd ed. Cambridge: Cambridge University Press. ISBN: 978-0-521-89560-6. DOI: `10.1017/CBO9780511803161`.

Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf (Oct. 2017). *Elements of Causal Inference: Foundations and Learning Algorithms*. The MIT Press. ISBN: 978-0-262-03731-0.

Pollard, David (2001). *A User’s Guide to Measure Theoretic Probability*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press. ISBN: 978-0-521-80242-0. DOI: `10.1017/CBO9780511811555`.

Richardson, Thomas (2003). “Markov Properties for Acyclic Directed Mixed Graphs”. In: *Scandinavian Journal of Statistics* 30.1, pp. 145–157. DOI: `10.1111/1467-9469.00323`.

Richardson, Thomas S., Robin J. Evans, et al. (Feb. 2023). “Nested Markov Properties for Acyclic Directed Mixed Graphs”. In: *The Annals of Statistics* 51.1, pp. 334–361. ISSN: 0090-5364, 2168-8966. DOI: `10.1214/22-AOS2253`.

Richardson, Thomas S and James M Robins (2013). *Single World Intervention Graphs (SWIGs): A Unification of the Counterfactual and Graphical Approaches to Causality*. Tech. rep. 128. Center for the Statistics and the Social Sciences, University of Washington Series.

Tchetgen, Eric J. Tchetgen et al. (Aug. 2024). “An Introduction to Proximal Causal Inference”. In: *Statistical Science* 39.3, pp. 375–390. ISSN: 0883-4237, 2168-8745. DOI: `10.1214/23-STS911`.

Tian, Jin and Judea Pearl (Aug. 2002). “On the Testable Implications of Causal Models with Hidden Variables”. In: *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*. UAI’02. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 519–527. ISBN: 978-1-55860-897-9.

Verma, Thomas S and Judea Pearl (1990). “Equivalence and Synthesis of Causal Models”. In: *Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence (UAI-1990)*. Cambridge, MA, USA, pp. 220–227.

Wright, Sewall (1934). “The Method of Path Coefficients”. In: *The Annals of Mathematical Statistics* 5.3, pp. 161–215. DOI: `10.1214/aoms/1177732676`.

Zhao, Qingyuan (July 2024). *A Matrix Algebra for Graphical Statistical Models*. arXiv: `2407.15744 [math, stat]`.

A Technical proofs

A.1 Proof of Theorem 1

As mentioned previously, many implications and equivalences in Figure 1 are already proved the literature. We will identify the new claims and then prove them in a sequence of Lemmas.

Relations in Figure 1a: it follows from the definition that PE $\Rightarrow$ CE, E $\Rightarrow$ NE, NM $\Rightarrow$ GM $\Rightarrow$ UM. It is shown in Richardson (2003, Theorem 2) that LM $\Leftrightarrow$ GM $\Leftrightarrow$ A and in Richardson, Evans, et al. (2023, Theorem 46) that CE $\Rightarrow$ NM. In Lemma 6 in the main text, it is shown that CE $\Rightarrow$ NE. It follows from Lemma 5 below that NE $\Rightarrow$ E and from Theorem 3 in the main text that E $\Rightarrow$ NM.

Relations in Figure 1b: it follows from Lemmas 2 to 4 below that E $\Leftrightarrow$ EF $\Leftrightarrow$ GM and GM $\Leftrightarrow$ NM. The rest of the relations follow from Figure 1a.

Relations in Figure 1c: it is shown in Lauritzen (1996, Theorem 3.27) that GM $\Leftrightarrow$ F (although there is a gap in the proof of Lauritzen (1996, Proposition 3.25); see the remark after Richardson (2003, Corollary 2)). By definition, $\mathrm{expand}_{\mathrm{P}}(\mathrm{G}) = \mathrm{G}$ because a DAG has no bidirected edges (recall that we do not consider bidirected loops). So, by definition, PE $\Leftrightarrow$ GM. The rest of the equivalences and implications follow from Figure 1b (because DAGs are unconfounded).

Relations in Figure 1d: it is shown in Richardson (2003, Theorem 3) that GM $\Leftrightarrow$ UM. The rest of the equivalences and implications follow from Figure 1b (because bidirected graphs are unconfounded).

It remains to show that the relations in Figure 1 are "tight" in the sense that when the two models are not connected by $\Leftrightarrow$ in Figure 1, there exists some graph in the corresponding class such that the models are not equal. It suffices to consider the following cases:

1. When G is a DAG, UM $\Rightarrow$ GM is not always true. For example, consider the graph $A \longleftarrow B \longrightarrow C$, for which the GM model contains the additional conditional independence $A \perp\!\!\!\perp C \mid B$.
2. When G is bidirected, GM $\Rightarrow$ CE and CE $\Rightarrow$ PE are not always true. This is closely related to Bell's inequalities in quantum mechanics; see Fritz (2012) for some examples.
3. When G is an ADMG, GM $\Rightarrow$ NM is generally not true. A well known example is the "Verma constraint" (Verma and Pearl 1990; Richardson, Evans, et al. 2023).
4. When G is an ADMG, NM $\Rightarrow$ NE is generally not true. This is because NM only contains equality constraints and latent variable models such as the E model may contain inequality constraints. A well known example is the Balke-Pearl bound for the instrumental variable graph (Balke and Pearl 1997).

Proof of new claims

Lemma 2. *For* $\mathrm{G} \in \mathbb{G}^*_{\mathrm{UA}}(V)$ *and any product space* $\mathbb{V}$*, we have* $\mathbb{P}_{\mathrm{E}}(\mathrm{G}, \mathbb{V}) = \mathbb{P}_{\mathrm{EF}}(\mathrm{G}, \mathbb{V})$.

Proof. Let $E \subseteq V$ denote a set of exogenous vertices in G. It follows from the definition that $\mathbb{P}_{\mathrm{E}}(\mathrm{G}, \mathbb{V}) \subseteq \mathbb{P}_{\mathrm{EF}}(\mathrm{G}, \mathbb{V})$. For the reverse, consider $\mathsf{P} \in \mathbb{P}_{\mathrm{EF}}(\mathrm{G}, \mathbb{V})$, so

$$\mathsf{p}(V = v) = \mathsf{p}(E = e) \prod_{V_j \notin E} \mathsf{p}(V_j = v_j \mid V_{\mathrm{pa}(j)} = v_{\mathrm{pa}(j)}),$$

where p is the density function of P and $\mathrm{pa}(j)$ is the parent set of V_j in G. For any $j = 1, \dots, d$ such that $V_j \notin E$, define $E'_j = \mathsf{P}(V_j \mid V_{\mathrm{pa}(j)})$, where $\mathsf{P}(v_j \mid v_{\mathrm{pa}(j))})$ is the conditional cumulative distribution function of V_j at v_j given $V_{\mathrm{pa}(j)} = v_{\mathrm{pa}(j)}$. Thus

$$V_j = \begin{cases} \mathsf{Q}_j(E'_j \mid V_{\mathrm{pa}(j)}), & \text{if } V_j \notin E, \\ E_j, & \text{otherwise,} \end{cases}$$

where $\mathsf{Q}_j(\cdot \mid v_{\mathrm{pa}(j)})$ is the conditional quantile function of V_j given $V_{\mathrm{pa}(j)} = v_{\mathrm{pa}(j)}$. Thus, V satisfies a system of equations with respect to G. Using the equivalence of GM and UM for bidirected graphs, it is easy to verify that the distribution of the noise variables in the system is global Markov with respect to the bidirected component of G because it factorizes as

$$\mathsf{p}(E = e) \prod_{V_j \notin E} \mathsf{p}(E'_j = e'_j).$$

This shows that $\mathsf{P} \in \mathbb{P}_{\mathrm{E}}(\mathrm{G}, \mathbb{V})$ and hence $\mathbb{P}_{\mathrm{EF}}(\mathrm{G}, \mathbb{V}) \subseteq \mathbb{P}_{\mathrm{E}}(\mathrm{G}, \mathbb{V})$. □

Lemma 3. *For $\mathrm{G} \in \mathbb{G}^*_{\mathrm{UA}}(V)$ and any product space $\mathbb{V}$, we have $\mathbb{P}_{\mathrm{EF}}(\mathrm{G}, \mathbb{V}) = \mathbb{P}_{\mathrm{GM}}(\mathrm{G}, \mathbb{V})$.*

Proof. By considering a topological order $\prec$ for G with the exogenous vertices being the smallest, it is straightforward to show that the ordered local Markov property implies the exogenous factorization property. Hence $\mathbb{P}_{\mathrm{GM}}(\mathrm{G}, \mathbb{V}) = \mathbb{P}_{\mathrm{LM}}(\mathrm{G}, \mathbb{V}) \subseteq \mathbb{P}_{\mathrm{EF}}(\mathrm{G}, \mathbb{V})$.

We next prove the reverse direction by using the augmentation criterion. Let $E \subseteq V$ denote a set of exogenous vertices in G; suppose $E = V_{\mathcal{E}}$ where $\mathcal{E} \subseteq [d]$. It is easy to see that if $\mathsf{P} \in \mathbb{P}_{\mathrm{EF}}(\mathrm{G}, \mathbb{V})$, then P factorizes according to augment(G) (the factorization property with respect to the augmentation graph, which is undirected, means that the density function p can be written as a product of terms that depend on the undirected cliques of the graph). By the Hammersley-Clifford theorem (Lauritzen 1996, p. 36), we have $\mathsf{P} \in \mathbb{P}_{\mathrm{GM}}(\mathrm{augment}(\mathrm{G}), \mathbb{V})$. Now consider any $\mathcal{J} \subseteq [d]$ such that $J = V_{\mathcal{J}}$ is ancestral. For $\mathsf{P} \in \mathbb{P}_{\mathrm{EF}}(\mathrm{G}, \mathbb{V})$, the joint density function can be factorized as

$$\mathsf{p}(v) = \mathsf{p}(v_{\mathcal{E}\cap\mathcal{J}})\, \mathsf{p}(v_{\mathcal{E}\setminus\mathcal{J}} \mid v_{\mathcal{E}\cap\mathcal{J}}) \prod_{j \in \mathcal{J}\setminus\mathcal{E}} \mathsf{p}(v_j \mid v_{\mathrm{pa}(j)}) \prod_{j \notin \mathcal{J}\cup\mathcal{E}} \mathsf{p}(v_j \mid v_{\mathrm{pa}(j)}).$$

By noting that all variables in the third term must belong to the ancestral set $V_{\mathcal{J}}$, it is easy to see that

$$\mathsf{p}(v_{\mathcal{J}}) = \mathsf{p}(v_{\mathcal{E}\cap\mathcal{J}}) \prod_{j \in \mathcal{J}\setminus\mathcal{E}} \mathsf{p}(v_j \mid v_{\mathrm{pa}(j)}).$$

Recall that the ancestral margin of an ADMG is simply its corresponding subgraph. This shows that $\mathrm{margin}_J(\mathsf{P}) \in \mathbb{P}_{\mathrm{EF}}(\mathrm{margin}_J(\mathrm{G}), \mathrm{margin}_J(\mathbb{V}))$, and by the same argument above,

$$\mathrm{margin}_J(\mathsf{P}) \in \mathbb{P}_{\mathrm{GM}}(\mathrm{augment} \circ \mathrm{margin}_J(\mathrm{G}), \mathrm{margin}_J(\mathbb{V})).$$

Therefore, $\mathsf{P} \in \mathbb{P}_{\mathrm{A}}(\mathrm{G}, \mathbb{V})$ and hence $\mathbb{P}_{\mathrm{EF}}(\mathrm{G}, \mathbf{V}) \subseteq \mathbb{P}_{\mathrm{A}}(\mathrm{G}, \mathbb{V}) = \mathbb{P}_{\mathrm{GM}}(\mathrm{G}, \mathbb{V})$. □

Lemma 4. *For $\mathrm{G} \in \mathbb{G}^*_{\mathrm{UA}}(V)$ and any product space $\mathbb{V}$, we have $\mathbb{P}_{\mathrm{NM}}(\mathrm{G}, \mathbb{V}) = \mathbb{P}_{\mathrm{GM}}(\mathrm{G}, \mathbb{V})$.*

Proof. The fact that $\mathbb{P}_{\mathrm{NM}}(\mathrm{G}, \mathbb{V}) \subseteq \mathbb{P}_{\mathrm{GM}}(\mathrm{G}, \mathbb{V})$ follows from the definition. The reverse direction is implied by Lemma 3 and $\mathbb{P}_{\mathrm{EF}}(\mathrm{G}, \mathbb{V}) \subseteq \mathbb{P}_{\mathrm{NM}}(\mathrm{G}, \mathbb{V})$, which we prove next. Consider any $\mathsf{P} \in \mathbb{P}_{\mathrm{EF}}(\mathrm{G}, \mathbb{V})$. Because G is unconfounded, it is easy to see that every vertex $V_j \in V$ is fixable. Thus, it suffices to show that

$$\mathrm{do}_{V_j = v_j}(\mathsf{P}) \in \mathbb{P}_{\mathrm{EF}}(\mathrm{do}_{V_j}(\mathrm{G}), \mathrm{do}_{V_j}(\mathbb{V})), \text{ for all } V_j \in V, \tag{10}$$

which implies that $\mathrm{do}_{V_j=v_j}(\mathsf{P})$ is global Markov with respect to $\mathrm{do}_{V_j}(\mathrm{G})$ by Lemma 3. By applying this result recursively, we find that $\mathrm{do}_J(\mathsf{P})$ is global Markov with respect to $\mathrm{do}_J(\mathrm{G})$ for all $J \subseteq V$.

Next we prove the claim in (10). When V_j is exogenous (i.e. $V_j \in E = V_{\mathcal{E}}$), we can factorize the density function of P as

$$\mathsf{p}(v) = \mathsf{p}(v_{\mathcal{E}\setminus\{j\}})\,\mathsf{p}(v_j \mid v_{\mathcal{E}\setminus\{j\}}) \prod_{V_k \in V\setminus E} \mathsf{p}(v_k \mid v_{\mathrm{pa}(k)}).$$

It is easy to see that the Markov background of V_j is the district containing V_j:

$$\mathrm{mbg}_{\mathrm{G}}(V_j) = \{V_k \in V : V_k \longleftrightarrow * \longleftrightarrow V_j\} = \mathrm{mb}_{\mathrm{G}}(V_j, E).$$

Therefore, $\mathsf{p}(v_j \mid v_{\mathrm{mbg}_{\mathrm{G}}(j)}) = \mathsf{p}(v_j \mid v_{\mathcal{E}\setminus\{j\}})$ and the operator $\mathrm{do}_{V_j=v_j}$ maps p to

$$\mathsf{p}(v_{\mathcal{E}\setminus\{j\}}) \prod_{V_k \in V\setminus E} \mathsf{p}(v_k \mid v_{\mathrm{pa}(k)}),$$

which satisfies the exogenous factorization property with respect to $\mathrm{do}_{V_j}(\mathrm{G})$, the subgraph that removes V_j. (Note that although v_j may appear in the second product, this is still a density function of V_{-j} because v_j is fixed.) We are left with the case that V_j is endogenous (i.e. $V_j \notin E$). In this case, the Markov background of V_j is simply the parent set of V_j, so the operator $\mathrm{do}_{V_j=v_j}$ maps p to

$$\mathsf{p}(v_{\mathcal{E}}) \prod_{V_k \in V\setminus E\setminus V_j} \mathsf{p}(v_k \mid v_{\mathrm{pa}(k)}),$$

which, again, satisfies the exogenous factorization property with respect to $\mathrm{do}_{V_j}(\mathrm{G})$. □

Lemma 5. *For* $\mathrm{G} \in \mathbb{G}_{\mathrm{A}}^*(V)$ *and any product space* $\mathbb{V}$*, we have* $\mathbb{P}_{\mathrm{NE}}(\mathrm{G}, \mathbb{V}) \subseteq \mathbb{P}_{\mathrm{E}}(\mathrm{G}, \mathbb{V})$.

Proof. We first consider the case that all random variables are real-valued (so $\mathbb{V}_j \subseteq \mathbb{R}$) and any distribution $\mathsf{P} \in \mathbb{P}_{\mathrm{NE}}(\mathrm{G}, \mathbb{V})$ on V. By definition, there exists a distribution P' on (V, E) such that $\mathsf{P}' \in \mathbb{P}_{\mathrm{GM}}\big(\mathrm{G}', \mathbb{V}\times[0,1]^{|V|}\big)$ for $\mathrm{G}' = \mathrm{expand}_{\mathrm{N}}(\mathrm{G})$ and $\mathrm{margin}_V(\mathsf{P}') = \mathsf{P}$. Because G' is unconfounded, P' must satisfy the exogenous factorization property (Lemma 3):

$$\mathsf{p}'(V = v \mid E = e) = \prod_{j=1}^{d} \mathsf{p}'(V_j = v_j \mid V_{\mathrm{pa}(j)} = v_{\mathrm{pa}(j)}, E_j = e_j),$$

where $\mathrm{pa}(j) = \mathrm{pa}_{\mathrm{G}}(j)$ contains indicies for the parents of V_j in G and the marginal distribution of E is global Markov with respect to the bidirected component of G. Let $\mathsf{P}'(v_j \mid v_{\mathrm{pa}(j)}, e_j)$ denote the conditional cumulative distribution function of V_j given $V_{\mathrm{pa}(j)} = v_{\mathrm{pa}(j)}$ and $E_j = e_j$, and let $\mathsf{Q}'(\cdot \mid v_{\mathrm{pa}(j)}, e_j)$ denote the associated conditional quantile function. Let $E_1', \dots, E_d'$ be independent uniform random variables over $[0,1]$ and let $V' = (V_1', \dots, V_d')$ be defined recursively by

$$V_j' = \mathsf{Q}'(E_j' \mid V_{\mathrm{pa}(j)}', E_j),\ j = 1, \dots, d.$$

Using the Galois connections for the distribution and quantile functions (i.e. $\mathsf{Q}(e) \leq v$ if and only if $e \leq \mathsf{P}(v)$ for any pair of distribution and quantile functions (P, Q) and $e \in [0,1]$, $v \in \mathbb{R}$), it is easy to show that V' has the same distribution P as V. Let $h : [0,1] \times [0,1] \to [0,1]$ be any (measurable) bijection.[11] It is obvious that V_j' is a function of $V_{\mathrm{pa}(j)}'$ and $h(E_j, E_j')$, and the distribution of

[11] One simple construction is to alternate between the digits in the binary expansion of the two arguments.

$h(E_1, E_1'), \dots, h(E_d, E_d')$ is global Markov with respect to the bidirected component of G. Thus $\mathsf{P} \in \mathbb{P}_{\mathrm{E}}(\mathrm{G}, \mathbb{V})$.

For general $\mathbb{V}_1, \dots, \mathbb{V}_d$, the above argument can be easily extended by introducing an order on the entries of $V_j \in V$ (if V_j is indeed multivariate) and applying the conditional quantile transform recursively according to that order. □

Lemma 6. *For* $\mathrm{G} \in \mathbb{G}_{\mathrm{A}}^*(V)$ *and any product space* $\mathbb{V}$*, we have* $\mathbb{P}_{\mathrm{CE}}(\mathrm{G}, \mathbb{V}) \subseteq \mathbb{P}_{\mathrm{NE}}(\mathrm{G}, \mathbb{V})$.

Proof. Consider $\mathsf{P} \in \mathbb{P}_{\mathrm{CE}}(\mathrm{G}, \mathbb{V})$ and let $\mathrm{G}' = \mathrm{expand}_{\mathrm{C}}(\mathrm{G})$. By definition, there exists a distribution $\mathsf{P}' \in \mathbb{P}_{\mathrm{GM}}(\mathrm{G}', \mathbb{V} \times [0,1]^{|\mathcal{C}(\mathrm{G})|})$ on V and $E_{\mathcal{J}}, \mathcal{J} \in \mathcal{C}(\mathrm{G})$ such that $\mathsf{P} = \mathrm{margin}_V(\mathsf{P}')$. Because G' is unconfounded, P' must satisfy the exogenous factorization property (Lemma 3):

$$\mathsf{p}'(V = v \mid E = e) = \prod_{j=1}^{d} \mathsf{p}'(V_j = v_j \mid V_{\mathrm{pa}(j)} = v_{\mathrm{pa}(j)}, \tilde{E}_j = \tilde{e}_j),$$

where $\mathrm{pa}(j)$ is the parent set of V_j in G and $\tilde{E}_j = (E_{\mathcal{J}} : \mathcal{J} \in \mathcal{C}(\mathrm{G}), j \in \mathcal{J})$ collects latent variables in G' with a directed edge to V_j. It is easy to see that the distribution of $(\tilde{E}_1, \dots, \tilde{E}_d)$ is global Markov with respect to the bidirected component of G. Let h_j be a (measurable) bijection that maps $[0,1]^{|\tilde{E}_j|}$ to $[0,1]$. Thus, the distribution of $(V_1, \dots, V_d, h_1(\tilde{E}_1), \dots, h_d(\tilde{E}_d))$ satisfies the exogenous factorization property (and thus the global Markov property by Lemma 3) with respect to the noise expansion graph $\mathrm{expand}_{\mathrm{N}}(\mathrm{G})$. This shows that $\mathsf{P} \in \mathbb{P}_{\mathrm{NE}}(\mathrm{G}, \mathbb{V})$. □

A.2 Proof of Theorem 2

For $\mathrm{G} \in \mathbb{G}_{\mathrm{A}}^*(V)$, let the collection of all unconfounded expansions of G be denoted as

$$\mathrm{expand}_{\mathrm{U}}(\mathrm{G}) = \bigcup_{V' \supset V} \{\mathrm{G}' \in \mathbb{G}_{\mathrm{UA}}^*(V') : \mathrm{margin}_V(\mathrm{G}') = \mathrm{G}\}.$$

Equation (2) can then be rewritten as

$$\mathbb{P}(\mathrm{G}) = \bigcup_{\mathrm{G}' \in \mathrm{expand}_{\mathrm{U}}(\mathrm{G})} \mathrm{margin}_V \left(\mathbb{P}(\mathrm{G}', \mathbb{V} \times [0,1]^{|V(\mathrm{G}')| - |V|})\right),$$

where $V(\mathrm{G}')$ is the vertex set of G'.

E/NE is complete

We first show that the E model (which is equivalent to NE by Figure 1a) is complete by proving the next result.

Proposition 4. *For any* $\mathrm{G} \in \mathbb{G}_{\mathrm{A}}^*(V)$*,* $|V| = d$*, and product space* $\mathbb{V} = \mathbb{V}_1 \times \cdots \times \mathbb{V}_d$*, we have*

$$\begin{aligned}
\mathbb{P}_{\mathrm{E}}(\mathrm{G}, \mathbb{V}) &= \mathrm{margin}_V \left(\mathbb{P}_{\mathrm{E}}\left(\mathrm{expand}_{\mathrm{N}}(\mathrm{G}), \mathbb{V} \times [0,1]^{|V|}\right)\right) \\
&= \bigcup_{\mathrm{G}' \in \mathrm{expand}_{\mathrm{U}}(\mathrm{G})} \mathrm{margin}_V \left(\mathbb{P}_{\mathrm{E}}\left(\mathrm{G}', \mathbb{V} \times [0,1]^{|V(\mathrm{G}')| - |V|}\right)\right) \\
&= \bigcup_{\mathrm{G}' \in \mathrm{expand}(\mathrm{G})} \mathrm{margin}_V \left(\mathbb{P}_{\mathrm{E}}\left(\mathrm{G}', \mathbb{V} \times [0,1]^{|V(\mathrm{G}')| - |V|}\right)\right).
\end{aligned}$$

Proof. We have

$$\begin{aligned}
\mathbb{P}_{\mathrm{E}}(\mathrm{G},\mathbb{V}) &\subseteq \mathrm{margin}_V(\mathbb{P}_{\mathrm{EF}}(\mathrm{expand}_{\mathrm{N}}(\mathrm{G}),\mathbb{V}\times[0,1]^{|V|})) && \text{(By definition)}\\
&= \mathrm{margin}_V(\mathbb{P}_{\mathrm{E}}(\mathrm{expand}_{\mathrm{N}}(\mathrm{G}),\mathbb{V}\times[0,1]^{|V|})) && \text{(By Theorem 1)}\\
&\subseteq \bigcup_{\mathrm{G}'\in\mathrm{expand}_{\mathrm{U}}(\mathrm{G})} \mathrm{margin}_V(\mathbb{P}_{\mathrm{E}}(\mathrm{G}',\mathbb{V}\times[0,1]^{|V(\mathrm{G}')|-|V|})) && \text{(By definition)}\\
&\subseteq \bigcup_{\mathrm{G}'\in\mathrm{expand}(\mathrm{G})} \mathrm{margin}_V(\mathbb{P}_{\mathrm{E}}(\mathrm{G}',\mathbb{V}\times[0,1]^{|V(\mathrm{G}')|-|V|})). && \text{(By definition)}
\end{aligned}$$

It remains to prove that $\mathbb{P}_{\mathrm{E}}(\mathrm{G},\mathbb{V}) \supseteq \mathrm{margin}_V(\mathbb{P}_{\mathrm{E}}(\mathrm{G}',\mathbb{V}\times[0,1]^{|V(\mathrm{G}')|-|V|}))$ for all $\mathrm{G}'\in\mathrm{expand}(\mathrm{G})$. This follows from Lemma 7 below. $\square$

Lemma 7. *For all $\mathrm{G}\in\mathbb{G}^*_{\mathrm{A}}(V)$ and $\tilde{V}\subseteq V$ that takes value in the subspace $\tilde{\mathbb{V}}\subseteq\mathbb{V}$, we have*

$$\mathrm{margin}_{\tilde{V}}(\mathbb{P}_{\mathrm{E}}(\mathrm{G},\mathbb{V})) \subseteq \mathbb{P}_{\mathrm{E}}(\mathrm{margin}_{\tilde{V}}(\mathrm{G}),\tilde{\mathbb{V}}). \tag{11}$$

Proof. Because marginalization is associative, it suffices to prove this for $\tilde{V} = V\setminus\{V_j\}$ for all $V_j\in V$. Consider $\mathsf{P}\in\mathbb{P}_{\mathrm{E}}(\mathrm{G},\mathbb{V})$, so V satisfy the equations in (3) and E satisfies (4). We need to show that $\mathrm{margin}_{\tilde{V}}(\mathsf{P})\in\mathbb{P}_{\mathrm{E}}(\mathrm{margin}_{\tilde{V}}(\mathrm{G},\tilde{\mathbb{V}}))$.

Consider the following modifications of the equations:

$$V_k = \begin{cases} f_k(V_{\mathrm{pa}(k)},E_k), & \text{if } k\notin\mathrm{ch}(j) \text{ and } k\neq j,\\ f_k(V_{\mathrm{pa}(k)\setminus\{j\}}, f_j(V_{\mathrm{pa}(j)},E_j),E_k), & \text{if } k\in\mathrm{ch}(j),\end{cases} \tag{12}$$

where $V_{\mathrm{pa}(k)}$ is the set of parents of V_k and $V_{\mathrm{ch}(j)}$ is the set of children of V_j in G. In words, we eliminate V_j by plugging $V_j = f_j(V_{\mathrm{pa}(j)},E_j)$ in all the equations for the children of V_j in G. We claim that this results in a nonparametric system with respect to $\tilde{\mathrm{G}} = \mathrm{margin}_{\tilde{V}}(\mathrm{G})$:

$$V_k = \tilde{f}_k(V_{\mathrm{pa}_{\tilde{\mathrm{G}}}(k)},\tilde{E}_k),\ k\neq j, \tag{13}$$

where $\mathrm{pa}_{\tilde{\mathrm{G}}}(k)$ is the parent of k in $\tilde{\mathrm{G}}$,

$$\tilde{E}_k = \begin{cases} E_k, & \text{if } k\notin\mathrm{ch}(j) \text{ and } k\neq j,\\ g(E_k,E_j), & \text{if } k\in\mathrm{ch}(j),\end{cases}$$

and g is any bi-measurable[12] bijective map from $[0,1]^2$ to $[0,1]$ (for example, g can be defined by interlacing the decimal expansions of its two arguments). To see this, marginalizing out V_j in G introduces the directed edges $V_{\mathrm{pa}(j)}\longrightarrow V_{\mathrm{ch}(j)}$, which are respected in the modified equations. Thus, the right hand side of (13) collects all the variables on the right hand side of (12). It remains to prove that $\tilde{E}$ obeys the global Markov property with respect to the bidirected component of $\tilde{\mathrm{G}}$.

Consider disjoint $J,K,L\subset\tilde{V}$ such that

$$\textbf{not } J \longleftrightarrow * \longleftrightarrow K \mid L \textbf{ in } \tilde{\mathrm{G}}. \tag{14}$$

Because all bidirected edges in G between vertices in $\tilde{V}$ are contained in $\tilde{\mathrm{G}}$, it follows that

$$\textbf{not } J \longleftrightarrow * \longleftrightarrow K \mid L \textbf{ in } \mathrm{G}\,. \tag{15}$$

[12]Meaning both g and its inverse are measurable.

Let $J = V_{\mathcal{J}}, K = V_{\mathcal{K}}, L = V_{\mathcal{L}}$. It follows from the Markov property of E that

$$E_{\mathcal{J}} \perp\!\!\!\perp E_{\mathcal{K}} \mid E_{\mathcal{L}}. \tag{16}$$

By construction,

$$\tilde{E}_{\mathcal{J}} = \begin{cases} E_{\mathcal{J}}, & \text{if } \mathcal{J} \cap \mathrm{ch}(j) = \emptyset, \\ h(E_{\mathcal{J} \cup \{j\}}), & \text{if } \mathcal{J} \cap \mathrm{ch}(j) \neq \emptyset, \end{cases}$$

where h is some bijective map and similarly for $\tilde{E}_{\mathcal{K}}$ and $\tilde{E}_{\mathcal{L}}$. We prove $\tilde{E}_{\mathcal{J}} \perp\!\!\!\perp \tilde{E}_{\mathcal{K}} \mid \tilde{E}_{\mathcal{L}}$ by considering the following cases:

1. $\mathcal{J} \cap \mathrm{ch}(j) = \mathcal{K} \cap \mathrm{ch}(j) = \mathcal{L} \cap \mathrm{ch}(j) = \emptyset$. The desired conclusion immediately follows from (16).
2. $\mathcal{J} \cap \mathrm{ch}(j) \neq \emptyset$, $\mathcal{K} \cap \mathrm{ch}(j) = \mathcal{L} \cap \mathrm{ch}(j) = \emptyset$. We claim that

$$\textbf{not } V_j \longleftrightarrow * \longleftrightarrow K \mid L, J \textbf{ in } \mathrm{G},$$

   otherwise there exists a walk like $J \longleftarrow V_j \longleftrightarrow * \longleftrightarrow K \mid L, J$ **in** G that marginalizes to $J \longleftrightarrow * \longleftrightarrow K \mid L, J$ **in** $\tilde{\mathrm{G}}$, which contradicts (14). By the Markov property of E, we have

$$E_j \perp\!\!\!\perp E_{\mathcal{K}} \mid E_{\mathcal{L}}, E_{\mathcal{J}}.$$

   By (16) and the chain rule for conditional independence, we obtain $E_{\mathcal{J} \cup \{j\}} \perp\!\!\!\perp E_{\mathcal{K}} \mid E_{\mathcal{L}}$.
3. $\mathcal{J} \cap \mathrm{ch}(j) \neq \emptyset$, $\mathcal{K} \cap \mathrm{ch}(j) = \emptyset$, $\mathcal{L} \cap \mathrm{ch}(j) \neq \emptyset$. We claim that

$$\textbf{not } J \longleftrightarrow * \longleftrightarrow K \mid L, V_j \textbf{ in } \mathrm{G}\,.$$

   If this not true, there exists a walk like $J \longleftrightarrow * \longleftrightarrow V_j \longleftrightarrow * \longleftrightarrow K \mid L$ **in** G because of (15). Thus, we have $J \longleftarrow V_j \longleftrightarrow * \longleftrightarrow K \mid L$ **in** G, which, after marginalization, contradicts (14). It follows from the above claim that $E_{\mathcal{J}} \perp\!\!\!\perp E_{\mathcal{K}} \mid E_{\mathcal{L} \cup \{j\}}$ and hence $E_{\mathcal{J} \cup \{j\}} \perp\!\!\!\perp E_{\mathcal{K}} \mid E_{\mathcal{L} \cup \{j\}}$.
4. $\mathcal{J} \cap \mathrm{ch}(j) = \emptyset$, $\mathcal{K} \cap \mathrm{ch}(j) \neq \emptyset$. This is symmetric to the last two cases.
5. $\mathcal{J} \cap \mathrm{ch}(j) \neq \emptyset$, $\mathcal{K} \cap \mathrm{ch}(j) \neq \emptyset$. This is not possible, because the confounding arc $J \longleftarrow V_j \longleftrightarrow V_j \longrightarrow K$ **in** G implies $J \longleftrightarrow K$ **in** $\tilde{\mathrm{G}}$, which contradicts (14).

This completes our proof of (17). □

Clique expansion is complete

Next, we prove that the CE model for ADMGs is the completion of the CE model for unconfounded graphs.

Proposition 5. *For any* $\mathrm{G} \in \mathbb{G}_{\mathrm{A}}^*(V)$, $|V| = d$, *and product space* $\mathbb{V} = \mathbb{V}_1 \times \cdots \times \mathbb{V}_d$, *we have*

$$\begin{aligned} \mathbb{P}_{\mathrm{CE}}(\mathrm{G}, \mathbb{V}) &= \mathrm{margin}_V \left(\mathbb{P}_{\mathrm{CE}} \left(\mathrm{expand}_{\mathrm{C}}(\mathrm{G}), \mathbb{V} \times [0,1]^{|V|} \right) \right) \\ &= \bigcup_{\mathrm{G}' \in \mathrm{expand}_{\mathrm{U}}(\mathrm{G})} \mathrm{margin}_V \left(\mathbb{P}_{\mathrm{CE}} \left(\mathrm{G}', \mathbb{V} \times [0,1]^{|V(\mathrm{G}')| - |V|} \right) \right) \\ &= \bigcup_{\mathrm{G}' \in \mathrm{expand}(\mathrm{G})} \mathrm{margin}_V \left(\mathbb{P}_{\mathrm{CE}} \left(\mathrm{G}', \mathbb{V} \times [0,1]^{|V(\mathrm{G}')| - |V|} \right) \right). \end{aligned}$$

Proof. The proof is similar to that of Proposition 4. Recall that $\mathrm{expand}_{\mathrm{C}}(\mathrm{G})$ is always a DAG. We have

$$\begin{aligned}
\mathbb{P}_{\mathrm{CE}}(\mathrm{G},\mathbb{V}) &= \mathrm{margin}_V(\mathbb{P}_{\mathrm{GM}}(\mathrm{expand}_{\mathrm{C}}(\mathrm{G}),\mathbb{V}\times[0,1]^{|V|})) && \text{(By definition)}\\
&= \mathrm{margin}_V(\mathbb{P}_{\mathrm{CE}}(\mathrm{expand}_{\mathrm{C}}(\mathrm{G}),\mathbb{V}\times[0,1]^{|V|})) && \text{(By Theorem 1)}\\
&\subseteq \bigcup_{\mathrm{G}'\in\mathrm{expand}_{\mathrm{U}}(\mathrm{G})} \mathrm{margin}_V(\mathbb{P}_{\mathrm{CE}}(\mathrm{G}',\mathbb{V}\times[0,1]^{|V(\mathrm{G}')|-|V|})) && \text{(By definition)}\\
&\subseteq \bigcup_{\mathrm{G}'\in\mathrm{expand}(\mathrm{G})} \mathrm{margin}_V(\mathbb{P}_{\mathrm{CE}}(\mathrm{G}',\mathbb{V}\times[0,1]^{|V(\mathrm{G}')|-|V|})). && \text{(By definition)}
\end{aligned}$$

The reverse direction follows from Lemma 8 below. □

Lemma 8. *For all* $\mathrm{G}\in\mathbb{G}_{\mathrm{A}}^{*}(V)$ *and* $\tilde{V}\subseteq V$ *that takes value in the subspace* $\tilde{\mathbb{V}}\subseteq\mathbb{V}$*, we have*

$$\mathrm{margin}_{\tilde{V}}(\mathbb{P}_{\mathrm{CE}}(\mathrm{G},\mathbb{V}))\subseteq\mathbb{P}_{\mathrm{CE}}(\mathrm{margin}_{\tilde{V}}(\mathrm{G}),\tilde{\mathbb{V}}). \tag{17}$$

Proof. Similar to the proof of Lemma 7, it suffices to prove this for $\tilde{V}=V\setminus\{V_j\}$ for all $V_j\in V$. Let $\tilde{\mathrm{G}}=\mathrm{margin}_{\tilde{V}}(\mathrm{G})$.

Let $\mathsf{P}\in\mathbb{P}_{\mathrm{CE}}(\mathrm{G},\mathbb{V})$, so by definition, there exists $\mathsf{P}'\in\mathbb{P}_{\mathrm{GM}}(\mathrm{expand}_{\mathrm{C}}(\mathrm{G}),\mathbb{V}\times[0,1]^{|\mathcal{C}(\mathrm{G})|})$ such that $\mathsf{P}=\mathrm{margin}_V(\mathsf{P}')$. Because $\mathrm{expand}_{\mathrm{C}}(\mathrm{G})$ is a DAG, this means that P' is also a nonparametric system of equations (by Theorem 1), that is

$$V_k=f_k(V_{\mathrm{pa}_{\mathrm{G}}(k)},C_k),\ k=1,\dots,d$$

for some functions $f_1,\dots,f_d$, $C_k=(E_{\mathcal{J}}:k\in\mathcal{J}\in\mathcal{C}(\mathrm{G}))$, and $E_{\mathcal{J}},\mathcal{J}\in\mathcal{C}(\mathrm{G})$ are independent random variables over $[0,1]$ under P'. We would like to show that $\mathrm{margin}_{\tilde{V}}(\mathsf{P}')\in\mathbb{P}_{\mathrm{CE}}(\tilde{\mathrm{G}},\tilde{\mathbb{V}})$, which requires us to rewrite the equations as

$$V_k=\tilde{f}_k(V_{\mathrm{pa}_{\tilde{\mathrm{G}}}(k)},\tilde{C}_k),\ k\neq j, \tag{18}$$

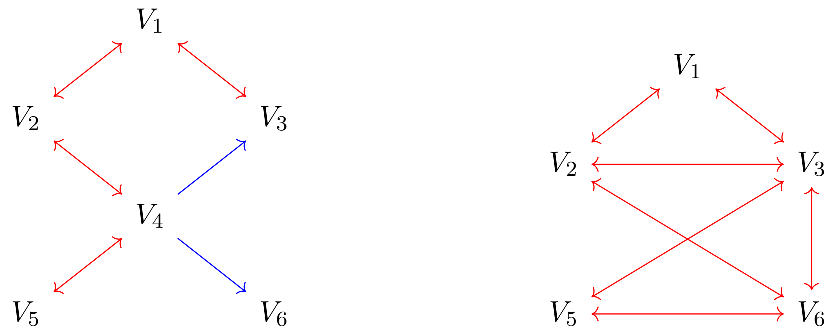
where $\mathrm{pa}_{\tilde{\mathrm{G}}}(k)$ is the parent of k in $\tilde{\mathrm{G}}$, $\tilde{C}_k=(\tilde{E}_{\tilde{\mathcal{J}}}:k\in\tilde{\mathcal{J}}\in\mathcal{C}(\tilde{\mathrm{G}}))$, and $\tilde{E}_{\tilde{\mathcal{J}}}$, $\tilde{\mathcal{J}}\in\mathcal{C}(\tilde{\mathrm{G}})$ are independent.

It is not difficult to see that

1. Any bidirected clique in G that does not contain V_j remains a bidirected clique in $\tilde{\mathrm{G}}$. That is, for any $\mathcal{J}\in\mathcal{C}(\mathrm{G})$ such that $j\notin\mathcal{J}$, we have $\mathcal{J}\in\mathcal{C}(\tilde{\mathrm{G}})$. In this case, define $\tilde{E}_{\mathcal{J}}=E_{\mathcal{J}}$ (unless it is redefined below).
2. Any bidirected clique in G that contains V_j, after removing V_j and adding $V_{\mathrm{ch}_{\mathrm{G}}(j)}$, is a bidirected clique in $\tilde{\mathrm{G}}$. That is, for any $\mathcal{J}\in\mathcal{C}(\mathrm{G})$ such that $j\in\mathcal{J}$, we have $\tilde{\mathcal{J}}=\mathcal{J}\setminus\{j\}\cup\mathrm{ch}_{\mathrm{G}}(j)\in\mathcal{C}(\tilde{\mathrm{G}})$. In this case, define

$$\tilde{E}_{\tilde{\mathcal{J}}}=\begin{cases}E_{\mathcal{J}}, & \text{if } \tilde{\mathcal{J}}\notin\mathcal{C}(\mathrm{G}),\\ g_{\mathcal{J}}(E_{\mathcal{J}},E_{\tilde{\mathcal{J}}}), & \text{if } \tilde{\mathcal{J}}\in\mathcal{C}(\mathrm{G}) \text{ (this redefines the variable)},\end{cases}$$

where $g_{\mathcal{J}}$ is an appropriate bijection from its domain to $[0,1]$.



(a) Cliques: 1, 2, 3, 4, 5, 6, 12, 13, 24, 45.

(b) Cliques: 1, 2, 3, 5, 6, 12, 13, 23, 26, 35, 36, 56, 123, 236, 356.

Figure 4: Marginalization can create many new cliques.

There may be other cliques in $\tilde{\mathrm{G}}$, but we do not need to consider them and will set the corresponding $\tilde{E}$ variable to be 0. See Example 1 below.

It is easy to see that $\tilde{E}_{\tilde{\mathcal{J}}}$, $\mathcal{J} \in \mathcal{C}(\tilde{\mathrm{G}})$ are independent because each variable $E_{\mathcal{J}}$ appears in exactly one $\tilde{E}_{\tilde{\mathcal{J}}}$. Now we prove that (18) is true. Similar to the proof of Lemma 7, we eliminate V_j by plugging its equation in all the equations for the children of V_j, so

$$V_k = \begin{cases} f_k(V_{\mathrm{pa}_{\mathrm{G}}(k)}, C_k), & \text{if } k \notin \mathrm{ch}_{\mathrm{G}}(j) \text{ and } k \neq j, \\ f_k(V_{\mathrm{pa}_{\mathrm{G}}(k)\setminus\{j\}}, f_j(V_{\mathrm{pa}(j)}, C_j), C_k), & \text{if } k \in \mathrm{ch}_{\mathrm{G}}(j), \end{cases}$$

Let us first prove (18) for $k \notin \mathrm{ch}_{\mathrm{G}}(j)$, so $\mathrm{pa}_{\tilde{\mathrm{G}}}(k) = \mathrm{pa}_{\mathrm{G}}(k)$. It suffices to show that every term in $E_{\mathcal{J}} \in C_k$ (so $k \in \mathcal{J} \in \mathcal{C}(\mathrm{G})$) shows up in $\tilde{C}_k$. This is true because

1. If $j \notin \mathcal{J}$, then $E_{\mathcal{J}}$ is contained in $\tilde{E}_{\mathcal{J}} \in \tilde{C}_k$ by construction;
2. If $j \in \mathcal{J}$, then $E_{\mathcal{J}}$ is contained in $\tilde{E}_{\tilde{\mathcal{J}}}$ for $\tilde{\mathcal{J}} = \mathcal{J} \setminus \{j\} \cup \mathrm{ch}_{\mathrm{G}}(j)$ (it is easy to check that $k \in \tilde{\mathcal{J}}$ and $\tilde{\mathcal{J}} \in \mathcal{C}(\tilde{\mathrm{G}})$ so $\tilde{E}_{\tilde{\mathcal{J}}} \in \tilde{C}_k$).

Next us first prove (18) for $k \in \mathrm{ch}_{\mathrm{G}}(j)$, so $\mathrm{pa}_{\tilde{\mathrm{G}}}(k) = \mathrm{pa}_{\mathrm{G}}(k) \setminus \{j\} \cup \mathrm{pa}_{\mathrm{G}}(j)$. It suffices to show that every term $E_{\mathcal{J}} \in C_j \cup C_k$ appears on the right hand side of (18). If $E_{\mathcal{J}} \in C_k$ the same argument as above (for $k \notin \mathrm{ch}_{\mathrm{G}}(j)$) applies. If $E_{\mathcal{J}} \in C_j$ (so $j \in \mathcal{J}$), we can use the second argument as above (note that $k \in \tilde{\mathcal{J}}$ is still true because $k \in \mathrm{ch}_{\mathrm{G}}(j)$). □

Example 1. As an example to illustrate this proof, let G be the graph in Figure 4a, so the nonparametric equation system for the clique expansion graph is given by

$$\begin{aligned}
V_1 &= f_1(E_1, E_{12}, E_{13}) &&= \tilde{f}_1(\tilde{E}_1, \tilde{E}_{12}, \tilde{E}_{13}, \tilde{E}_{123}), \\
V_2 &= f_2(E_2, E_{12}, E_{24}) &&= \tilde{f}_2(\tilde{E}_2, \tilde{E}_{12}, \tilde{E}_{23}, \tilde{E}_{24}, \tilde{E}_{123}, \tilde{E}_{236}), \\
V_3 &= f_3(f_4(E_4, E_{24}, E_{45}), E_3, E_{13}) &&= \tilde{f}_3(\tilde{E}_3, \tilde{E}_{13}, \tilde{E}_{23}, \tilde{E}_{35}, \tilde{E}_{36}, \tilde{E}_{123}, \tilde{E}_{236}, \tilde{E}_{356}), \\
V_5 &= f_5(E_5, E_{45}) &&= \tilde{f}_5(\tilde{E}_5, \tilde{E}_{35}, \tilde{E}_{56}, \tilde{E}_{356}), \\
V_6 &= f_6(f_4(E_4, E_{24}, E_{45}), E_6) &&= \tilde{f}_6(\tilde{E}_6, \tilde{E}_{26}, \tilde{E}_{36}, \tilde{E}_{56}, \tilde{E}_{236}, \tilde{E}_{356}),
\end{aligned}$$

where $\tilde{E}_{\cdot} = E_{\cdot}$ for $\cdot \in \{1, 2, 3, 5, 6, 12, 13\}$, $\tilde{E}_{35} = E_4$, $\tilde{E}_{236} = E_{24}$, $\tilde{E}_{356} = E_{45}$, and $\tilde{E}_{\cdot} = 0$ for $\cdot \in \{23, 24, 26, 35, 36, 123\}$.

UM is complete

Let us first consider an example.

Example 2. Consider the instrumental variable graph

$$\begin{tikzcd} Z \arrow[r] & X \arrow[r] \arrow[r, bend left, leftrightarrow] & Y \end{tikzcd}$$

and its clique expansion

$$\begin{tikzcd} & & E_{XY} \arrow[dl] \arrow[dr] & \\ Z \arrow[r] & X \arrow[rr] & & Y \end{tikzcd}$$

The UM model for the instrumental variable graph contains all probability distributions of (Z, X, Y) because they are all connected by arcs. On the other hand, it is well known that a latent variable interpretation of this graph imposes inequality constraints (Balke and Pearl 1997). However, this implicitly assumes the usual interpretation of DAGs (e.g. factorization, global Markov, or any of their equivalences in Figure 1c), which contains all distributions of (Z, X, Y, E_{XY}) such that $Z \perp\!\!\!\perp E_{XY}$ and $Y \perp\!\!\!\perp Z \mid X, E_{XY}$. In contrast, the UM model contains all distributions of (Z, X, Y, E_{XY}) such that $E_{XY} \perp\!\!\!\perp Z$, which impose no constraint on the marginal distribution of (Z, X, Y).

We now prove that the UM model is complete with respect to unconfounded graph expansions. First, we show

$$\mathbb{P}_{\mathrm{UM}}(\mathrm{G}) \subseteq \bigcup_{\mathrm{G}' \in \mathrm{expand}_{\mathrm{U}}(\mathrm{G})} \mathrm{margin}_V(\mathbb{P}_{\mathrm{UM}}(\mathrm{G}'))$$

with an almost trivial construction. Consider any $\mathsf{P} \in \mathbb{P}_{\mathrm{UM}}(\mathrm{G})$ with density function $\mathsf{p}(v)$. Consider the clique expansion of G and the density function

$$\mathsf{p}'(v, e) = \mathsf{p}(v)\, \mathsf{q}(e),$$

where q is density function of the uniform distribution over $[0, 1]^{|\mathcal{C}(\mathrm{G})|}$ (so $\mathsf{q}(e) = 1$ for all e). It is obvious that p' marginalizes to p, and p' satisfies the unconditional Markov property with respect to the clique expansion graph.

The reverse direction follows from the fact that marginalization preserves m-connection. That is, for disjoint $J, K \subseteq V \subseteq V'$ and graphs $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V), \mathrm{G}' \in \mathbb{G}^*_{\mathrm{A}}(V')$, if $\mathrm{margin}_V(\mathrm{G}') = \mathrm{G}$, then $J \rightsquigarrow K$ **in** G if and only if $J \rightsquigarrow K$ **in** G′ (see, for example, Guo and Zhao 2023, Theorem 2).

Other ADMG models are not complete

Because the E model is equivalent to the NM, LM, GM, and A models when the graph is unconfounded, by Proposition 4, the corresponding model for general ADMGs as defined by (2) is also the E model. By Theorem 1, the E model is different from the NM, LM, GM, and A models for general ADMGs. Thus, the NM, LM, GM, and A models are not complete with respect to unconfounded graph expansions.

It remains to show that PE is not complete. Consider the "bidirected 3-loop" with edges $A \longleftrightarrow B$, $B \longleftrightarrow C$, and $C \longleftrightarrow A$. If the PE model is complete, it should contain the (A, B, C)-marginal of the DAG with edges $U \longrightarrow A$, $U \longrightarrow B$, $U \longrightarrow C$, which places no restrictions on the distribution of (A, B, C). However, the PE model has some inequality constraints; see Fritz (2012, Example 2.11). So the PE model is "too small".

A.3 Proof of Proposition 1

Suppose $V_{\mathcal{I}'}(v_{\mathcal{I}}) = v_{\mathcal{I}'}$. Consider any $V_j \in V$. It follows from (5) that

$$V_j(v_{\mathcal{I}}, v_{\mathcal{I}'}) = V_j(v_{\mathrm{pa}(j)\cap\mathcal{I}}, v_{\mathrm{pa}(j)\cap\mathcal{I}'}, V_{\mathrm{pa}(j)\setminus(\mathcal{I}\cup\mathcal{I}')}(v_{\mathcal{I}}, v_{\mathcal{I}'}))$$

and

$$\begin{aligned} V_j(v_{\mathcal{I}}) &= V_j(v_{\mathrm{pa}(j)\cap\mathcal{I}}, V_{\mathrm{pa}(j)\cap\mathcal{I}'}(v_{\mathcal{I}}), V_{\mathrm{pa}(j)\setminus(\mathcal{I}\cup\mathcal{I}')}(v_{\mathcal{I}})) \\ &= V_j(v_{\mathrm{pa}(j)\cap\mathcal{I}}, v_{\mathrm{pa}(j)\cap\mathcal{I}'}, V_{\mathrm{pa}(j)\setminus(\mathcal{I}\cup\mathcal{I}')}(v_{\mathcal{I}})). \end{aligned}$$

Thus, it suffices to show that, if $\mathrm{pa}(j) \setminus (\mathcal{I} \cup \mathcal{I}')$ is not empty,

$$V_{\mathrm{pa}(j)\setminus(\mathcal{I}\cup\mathcal{I}')}(v_{\mathcal{I}}, v_{\mathcal{I}'}) = V_{\mathrm{pa}(j)\setminus(\mathcal{I}\cup\mathcal{I}')}(v_{\mathcal{I}}).$$

The proof can then be completed by an induction argument.

A.4 Proof of Proposition 2

Consider $\mathrm{G} \in \mathbb{G}^*_{\mathrm{A}}(V)$ and $\mathsf{P} \in \mathbb{CP}(\mathrm{G}, \mathbb{V})$. Because G is acyclic, for any $V_{\mathcal{I}} \subset V$, there always exists $V_j \notin V_{\mathcal{I}}$ such that $\mathrm{de}_{\mathrm{G}}(V_j) \subseteq V_{\mathcal{I}}$. By the definition of causal Markov model and in particular (6), $\mathrm{margin}_{V(v)}(\mathsf{P}) \in \mathbb{P}_{\mathrm{GM}}(\mathrm{G}(V(v)), \mathbb{V})$. Proposition 2 then follows from repeatedly applying the following result.

Lemma 9 (Recursive substitution preserves global Markov property)**.** *Consider any* $V_{\mathcal{I}} \subset V$ *and* $V_j \notin V_{\mathcal{I}}$ *such that* $\mathrm{de}_{\mathrm{G}}(V_j) \subseteq V_{\mathcal{I}}$. *Let* $\mathcal{I}' = \mathcal{I} \cup \{j\}$. *If* $\mathrm{margin}_{V(v_{\mathcal{I}'})}(\mathsf{P}) \in \mathbb{P}_{\mathrm{GM}}(\mathrm{G}(v_{\mathcal{I}'}), \mathbb{V})$*, then* $\mathrm{margin}_{V(v_{\mathcal{I}})}(\mathsf{P}) \in \mathbb{P}_{\mathrm{GM}}(\mathrm{G}(v_{\mathcal{I}}), \mathbb{V})$

We will abbreviate $\mathrm{ch}_{\mathrm{G}}(V_j)$ as $\mathrm{ch}(V_j)$ below. The following observations will be useful in our proof of Lemma 9:

(i) We have $V_k(v_{\mathcal{I}}) = V_k(v_{\mathcal{I}'})$ for any $V_k \notin \mathrm{ch}(V_j)$.
(ii) $\mathrm{G}(v_{\mathcal{I}})$ has all the edges in $\mathrm{G}(v_{\mathcal{I}'})$ (after relabeling the vertices using $V_k(v_{\mathcal{I}'}) \mapsto V_k(v_{\mathcal{I}})$) and additionally the edges $V_j(v_{\mathcal{I}}) \longrightarrow V_{\mathrm{ch}(j)}(v_{\mathcal{I}})$.
(iii) It follows from the previous observation that any m-separation in $\mathrm{G}(v_{\mathcal{I}})$ also holds $\mathrm{G}(v_{\mathcal{I}'})$ (after relabeling the vertices using $V_k(v_{\mathcal{I}'}) \mapsto V_k(v_{\mathcal{I}})$).
(iv) There are no edges like $V_{\mathrm{ch}(j)}(v_{\mathcal{I}}) \longrightarrow *$ (as $V_{\mathrm{ch}(j)} \subseteq V_{\mathcal{I}}$ by assumption).

To prove Lemma 9, it suffices to show that for all disjoint $V_{\mathcal{K}}, V_{\mathcal{L}}, V_{\mathcal{M}} \subset V$,

$$\textbf{not } V_{\mathcal{K}}(v_{\mathcal{I}}) \leftsquigarrow\rightsquigarrow * \leftsquigarrow\rightsquigarrow V_{\mathcal{L}}(v_{\mathcal{I}}) \mid V_{\mathcal{M}}(v_{\mathcal{I}}) \textbf{ in } \mathrm{G}(v_{\mathcal{I}}) \Longrightarrow V_{\mathcal{K}}(v_{\mathcal{I}}) \perp\!\!\!\perp V_{\mathcal{L}}(v_{\mathcal{I}}) \mid V_{\mathcal{M}}(v_{\mathcal{I}}) \textbf{ under } \mathsf{P}. \quad (19)$$

We will prove (19) by considering two separate cases.

Lemma 10. *Under the assumptions in Lemma 9, the implication in* (19) *is true if* $V_j \in V_{\mathcal{K}} \cup V_{\mathcal{L}}$.

Proof. By symmetry, it suffices to prove (19) when $V_j \in V_{\mathcal{L}}$. First, we claim that the m-separation in (19) implies

$$\textbf{not } V_{\mathcal{K}}(v_{\mathcal{I}}) \leftsquigarrow\rightsquigarrow * \leftsquigarrow\rightsquigarrow V_{\mathcal{L}}(v_{\mathcal{I}}), V_{\mathcal{M}\,\cap\,\mathrm{ch}(j)}(v_{\mathcal{I}}) \mid V_{\mathcal{M}\,\setminus\,\mathrm{ch}(j)}(v_{\mathcal{I}}) \textbf{ in } \mathrm{G}(v_{\mathcal{I}}). \quad (20)$$

We prove this claim by contradiction. Suppose (20) is not true, so there exists $V_m \in V_{\mathcal{L}} \cup V_{\mathcal{M} \cap \mathrm{ch}(j)}$ such that

$$V_{\mathcal{K}}(v_{\mathcal{I}}) \rightsquigarrow * \leftsquigarrow V_m(v_{\mathcal{I}}) \mid V_{\mathcal{M} \setminus \mathrm{ch}(j)}(v_{\mathcal{I}}) \textbf{ in } \mathrm{G}(v_{\mathcal{I}}).$$

First, note that by observation (iv), if a vertex in $V_{\mathrm{ch}(j)}(v_{\mathcal{I}})$ is a non-endpoint in a walk, it is a collider. Thus, $V_m \in V_{\mathcal{M} \cap \mathrm{ch}(j)}$ (the $V_m \in V_{\mathcal{L}}$ case gives an immediate contradiction with the m-separation in (19)). Again, by using observation (iv) and the fact that $V_{\mathrm{ch}(j)}(v_{\mathcal{I}})$ can only be colliders, we know

$$V_{\mathcal{K}}(v_{\mathcal{I}}) \rightsquigarrow * \leftrightsquigarrow V_m(v_{\mathcal{I}}) \mid V_{\mathcal{M} \setminus \{m\}}(v_{\mathcal{I}}) \textbf{ in } \mathrm{G}(v_{\mathcal{I}}).$$

Because $m \in \mathrm{ch}(j)$, this shows

$$V_{\mathcal{K}}(v_{\mathcal{I}}) \rightsquigarrow * \leftrightsquigarrow V_m(v_{\mathcal{I}}) \longleftarrow V_j(v_{\mathcal{I}}) \mid V_{\mathcal{M}}(v_{\mathcal{I}}) \textbf{ in } \mathrm{G}(v_{\mathcal{I}}).$$

This again contradicts the m-separation in (19).

Using observation (iii), (20) implies that

$$\textbf{not } V_{\mathcal{K}}(v_{\mathcal{I}'}) \rightsquigarrow * \leftsquigarrow V_{\mathcal{L}}(v_{\mathcal{I}'}), V_{\mathcal{M} \cap \mathrm{ch}(j)}(v_{\mathcal{I}'}) \mid V_{\mathcal{M} \setminus \mathrm{ch}(j)}(v_{\mathcal{I}'}) \textbf{ in } \mathrm{G}(v_{\mathcal{I}'})$$

So by the global Markov property of $\mathrm{margin}_{V(v_{\mathcal{I}'})}(\mathsf{P})$, we have

$$V_{\mathcal{K}}(v_{\mathcal{I}'}) \perp\!\!\!\perp V_{\mathcal{L}}(v_{\mathcal{I}'}), V_{\mathcal{M} \cap \mathrm{ch}(j)}(v_{\mathcal{I}'}) \mid V_{\mathcal{M} \setminus \mathrm{ch}(j)}(v_{\mathcal{I}'}) \textbf{ under } \mathsf{P}. \tag{21}$$

Next we show that the same conditional independence for potential outcomes under $v_{\mathcal{I}}$ is also true. We have, for any $\tilde{v} \in \mathbb{V}$,

$$\begin{aligned}
&\mathsf{p}(V_{\mathcal{K}}(v_{\mathcal{I}}) = \tilde{v}_{\mathcal{K}} \mid V_{\mathcal{L}}(v_{\mathcal{I}}) = \tilde{v}_{\mathcal{L}}, V_{\mathcal{M} \cap \mathrm{ch}(j)}(v_{\mathcal{I}}) = \tilde{v}_{\mathcal{M} \cap \mathrm{ch}(j)}, V_{\mathcal{M} \setminus \mathrm{ch}(j)}(v_{\mathcal{I}}) = \tilde{v}_{\mathcal{M} \setminus \mathrm{ch}(j)}) \\
&= \mathsf{p}(V_{\mathcal{K}}(v_{\mathcal{I}}, \tilde{v}_j) = \tilde{v}_{\mathcal{K}} \mid V_{\mathcal{L}}(v_{\mathcal{I}}, \tilde{v}_j) = \tilde{v}_{\mathcal{L}}, V_{\mathcal{M} \cap \mathrm{ch}(j)}(v_{\mathcal{I}}, \tilde{v}_j) = \tilde{v}_{\mathcal{M} \cap \mathrm{ch}(j)}, V_{\mathcal{M} \setminus \mathrm{ch}(j)}(v_{\mathcal{I}}, \tilde{v}_j) = \tilde{v}_{\mathcal{M} \setminus \mathrm{ch}(j)}) \\
&= \mathsf{p}(V_{\mathcal{K}}(v_{\mathcal{I}}, \tilde{v}_j) = \tilde{v}_{\mathcal{K}} \mid V_{\mathcal{M} \setminus \mathrm{ch}(j)}(v_{\mathcal{I}}, \tilde{v}_j) = \tilde{v}_{\mathcal{M} \setminus \mathrm{ch}(j)}) \\
&= \mathsf{p}(V_{\mathcal{K}}(v_{\mathcal{I}}) = \tilde{v}_{\mathcal{K}} \mid V_{\mathcal{M} \setminus \mathrm{ch}(j)}(v_{\mathcal{I}}) = \tilde{v}_{\mathcal{M} \setminus \mathrm{ch}(j)}),
\end{aligned}$$

the first equality follows from consistency of potential outcomes (Proposition 1) and the assumption that $V_j \in V_{\mathcal{L}}$, the secone equality follows from (21), and the last equality follows from observation (i) (the m-separation in (19) implies that $V_j(v_{\mathcal{I}}) \not\rightsquigarrow V_{\mathcal{K}}(v_{\mathcal{I}})$). This shows that

$$V_{\mathcal{K}}(v_{\mathcal{I}}) \perp\!\!\!\perp V_{\mathcal{L}}(v_{\mathcal{I}}), V_{\mathcal{M} \cap \mathrm{ch}(j)}(v_{\mathcal{I}}) \mid V_{\mathcal{M} \setminus \mathrm{ch}(j)}(v_{\mathcal{I}}) \textbf{ under } \mathsf{P},$$

which immediately implies the conditional independence in (19) by the weak union property of conditional independence. $\square$

Lemma 11. *Under the assumptions in Lemma 9, the implication in* (19) *is true if* $V_j \notin V_{\mathcal{K}} \cup V_{\mathcal{L}}$.

Proof. If $V_j \not\rightsquigarrow V_{\mathcal{K}} \cup V_{\mathcal{L}} \cup V_{\mathcal{M}}$, then the implication in (19) immediately follows from the global Markov property of $\mathrm{margin}_{V(v_{\mathcal{I}'})}(\mathsf{P})$ and observation (i). We now assume $V_j \longrightarrow V_{\mathcal{K}} \cup V_{\mathcal{L}} \cup V_{\mathcal{M}}$.

We claim that

$$\textbf{not } V_{\mathcal{K}}(v_{\mathcal{I}}) \rightsquigarrow * \leftsquigarrow V_{\mathcal{L}}(v_{\mathcal{I}}) \mid V_{\mathcal{M}}(v_{\mathcal{I}}), V_j(v_{\mathcal{I}}), \tag{22}$$

Otherwise, we have

$$V_{\mathcal{K}}(v_{\mathcal{I}}) \rightsquigarrow * \leftrightsquigarrow V_j(v_{\mathcal{I}}) \leftrightsquigarrow * \leftsquigarrow V_{\mathcal{L}}(v_{\mathcal{I}}) \mid V_{\mathcal{M}}(v_{\mathcal{I}}),$$

By appending the edge $V_j \longrightarrow V_{\mathcal{K}} \cup V_{\mathcal{L}} \cup V_{\mathcal{M}}$, this leads to a contradiction with the m-separation in (19).

Further, we claim that

$$\textbf{not } V_{\mathcal{K}}(v_{\mathcal{I}}) \leftrightsquigarrow * \leftrightsquigarrow V_j(v_{\mathcal{I}}) \mid V_{\mathcal{M}}(v_{\mathcal{I}}) \quad \text{or} \quad \textbf{not } V_{\mathcal{L}}(v_{\mathcal{I}}) \leftrightsquigarrow * \leftrightsquigarrow V_j(v_{\mathcal{I}}) \mid V_{\mathcal{M}}(v_{\mathcal{I}}).$$

Otherwise, we have

$$V_{\mathcal{K}}(v_{\mathcal{I}}) \leftrightsquigarrow * \leftrightsquigarrow V_j(v_{\mathcal{I}}) \leftrightsquigarrow * \leftrightsquigarrow V_{\mathcal{L}}(v_{\mathcal{I}}) \mid V_{\mathcal{M}}(v_{\mathcal{I}}).$$

The case where $V_j(v_{\mathcal{I}})$ is a collider already shown to be impossible above. In the other case, all $V_j(v_{\mathcal{I}})$ in this walk are not colliders and it contradicts (22).

Without loss of generality, let us assume

$$\textbf{not } V_{\mathcal{K}}(v_{\mathcal{I}}) \leftrightsquigarrow * \leftrightsquigarrow V_j(v_{\mathcal{I}}) \mid V_{\mathcal{M}}(v_{\mathcal{I}}).$$

By composing this with the m-separation in (19), we obtain

$$\textbf{not } V_{\mathcal{K}}(v_{\mathcal{I}}) \leftrightsquigarrow * \leftrightsquigarrow V_{\mathcal{L}\cup\{j\}}(v_{\mathcal{I}}) \mid V_{\mathcal{M}}(v_{\mathcal{I}}).$$

It follows from Lemma 10 that

$$V_{\mathcal{K}}(v_{\mathcal{I}}) \perp\!\!\!\perp V_{\mathcal{L}\cup\{j\}}(v_{\mathcal{I}}) \mid V_{\mathcal{M}}(v_{\mathcal{I}}),$$

which implies the conditional independence in (19). □

A.5 Proof of Proposition 3

Let us first prove the following graphical result.

Lemma 12. *A vertex $V_j \in V$ is fixable in $\mathrm{G} \in \mathbb{G}_{\mathrm{A}}^*(V)$ if and only if*

$$\textbf{not } V_j(v_j) \leftrightsquigarrow * \leftrightsquigarrow V_{\mathrm{de}_{\mathrm{G}}(j)}(v_j) \mid V_{\mathrm{nd}_{\mathrm{G}}(j)}(v_j) \textbf{ in } \mathrm{G}(v_j), \tag{23}$$

where $\mathrm{nd}_{\mathrm{G}}(j) = [d] \setminus \{j\} \setminus \mathrm{de}_{\mathrm{G}}(j)$ *collects the indicies of the non-descendants of* V_j *in* G.

Proof. Because $V_j(v_j)$, $V_{\mathrm{de}(j)}(v_j)$, and $V_{\mathrm{nd}(j)}(v_j)$ gives a partition of the vertex set of $\mathrm{G}(v_j)$, the m-separation in (23) is equivalent to

$$\textbf{not } V_j(v_j) \longrightarrow * \longleftarrow V_{\mathrm{de}(j)}(v_j) \mid V_{\mathrm{nd}(j)}(v_j) \textbf{ in } \mathrm{G}(v_j),$$

which is further equivalent to

$$\textbf{not } V_j(v_j) \longleftrightarrow * \longleftrightarrow V_{\mathrm{de}(j)}(v_j) \mid V_{\mathrm{nd}(j)}(v_j) \textbf{ in } \mathrm{G}(v_j)$$

because $V_j(v_j)$ has no children and acyclicity of G (so $V_{\mathrm{de}(j)}(v_j) \not\to V_j(v_j), V_{\mathrm{nd}(j)}(v_j)$). By the definition of $\mathrm{G}(v_j)$, the last condition is equivalent to

$$\textbf{not } V_j \longleftrightarrow * \longleftrightarrow V_{\mathrm{de}(j)} \mid V_{\mathrm{nd}(j)} \textbf{ in } \mathrm{G},$$

Again, because V_j, $V_{\mathrm{de}(j)}$, and $V_{\mathrm{nd}(j)}$ partition the vertex set of G, this is equivalent to

$$\textbf{not } V_j \longleftrightarrow * \longleftrightarrow V_{\mathrm{de}(j)} \textbf{ in } \mathrm{G},$$

which is exactly what fixability of V_j means. □

We now turn to prove Proposition 3. The consistency property (5) implies that $V_{\text{nd}(j)}(v_j) = V_{\text{nd}(j)}$ and $V_j(v_j) = V_j$. So by factorizing the joint density of $V(v_j)$, we have

$$\begin{aligned}
&\mathsf{p}(V_j(v_j) = \tilde{v}_j, V_{-j}(v_j) = v_{-j}) \\
&= \mathsf{p}(V_{\text{nd}(j)} = v_{\text{nd}(j)})\, \mathsf{p}(V_j = \tilde{v}_j \mid V_{\text{nd}(j)} = v_{\text{nd}(j)})\, \mathsf{p}(V_{\text{de}(j)}(v_j) = v_j \mid V_{\text{nd}(j)} = v_{\text{nd}(j)}, V_j = \tilde{v}_j) \\
&= \mathsf{p}(V_{\text{nd}(j)} = v_{\text{nd}(j)})\, \mathsf{p}(V_j = \tilde{v}_j \mid V_{\text{nd}(j)} = v_{\text{nd}(j)})\, \mathsf{p}(V_{\text{de}(j)}(v_j) = v_j \mid V_{\text{nd}(j)} = v_{\text{nd}(j)}, V_j = v_j) \\
&= \mathsf{p}(V_{\text{nd}(j)} = v_{\text{nd}(j)})\, \mathsf{p}(V_j = \tilde{v}_j \mid V_{\text{nd}(j)} = v_{\text{nd}(j)})\, \mathsf{p}(V_{\text{de}(j)} = v_j \mid V_{\text{nd}(j)} = v_{\text{nd}(j)}, V_j = v_j),
\end{aligned}$$

where the second equality follows from fixability of V_j and Lemma 12, and the last equality follows from the consistency of potential outcomes (Proposition 1). By factorizing $\mathsf{p}(V = v)$ in a similar way and rearranging the terms, we obtain

$$\frac{\mathsf{p}(V_j(v_j) = \tilde{v}_j, V_{-j}(v_j) = v_{-j})}{\mathsf{p}(V_j = v_j, V_{-j} = v_{-j})} = \frac{\mathsf{p}(V_j = \tilde{v}_j \mid V_{\text{nd}(j)} = v_{\text{nd}(j)})}{\mathsf{p}(V_j = v_j \mid V_{\text{nd}(j)} = v_{\text{nd}(j)})}. \tag{24}$$

It is easy to see that

$$\textbf{not } V_j \leftsquigarrow\!\!\rightsquigarrow * \leftsquigarrow\!\!\rightsquigarrow V_{\text{nd}(j)\setminus\text{mbg}(j)} \mid V_{\text{mbg}(j)} \textbf{ in } \mathrm{G}\,.$$

By Proposition 2, we have

$$V_j \perp\!\!\!\perp V_{\text{nd}(j)\setminus\text{mbg}(j)} \mid V_{\text{mbg}(j)} \textbf{ under } \mathsf{P}\,.$$

Equation (8) then immediately follows from (24).