Estimating Identifiable Causal Effects on Markov Equivalence Class through Double Machine Learning

Yonghan Jung [1] **Jin Tian** [2] **Elias Bareinboim** [3]

Abstract

General methods have been developed for estimating causal effects from observational data under causal assumptions encoded in the form of a causal graph. Most of this literature assumes that the underlying causal graph is completely specified. However, only observational data is available in most practical settings, which means that one can learn at most a Markov equivalence class (MEC) of the underlying causal graph. In this paper, we study the problem of causal estimation from a MEC represented by a partial ancestral graph (PAG), which is learnable from observational data. We develop a general estimator for any identifiable causal effects in a PAG. The result fills a gap for an end-to-end solution to causal inference from observational data to effects estimation. Specifically, we develop a complete identification algorithm that derives an influence function for any identifiable causal effects from PAGs. We then construct a double/debiased machine learning (DML) estimator that is robust to model misspecification and biases in nuisance function estimation, permitting the use of modern machine learning techniques. Simulation results corroborate with the theory.

1. Introduction

Inferring causal effects from observational data is a fundamental task in machine learning and various empirical sciences. There exists a growing literature studying the conditions under which causal conclusions can be drawn from non-experimental data (Pearl, 2000; Bareinboim & Pearl, 2016; Pearl & Mackenzie, 2018). In particular, the literature of *causal effect identification* (Pearl, 2000, Def. 3.2.4) investigates whether, given a causal graph G encoding qualitative knowledge about the domain, an interventional distribution $P(Y = y|do(X = x))$ (for short, $P_x(y)$), representing the causal effect of the treatment X on the outcome Y, can be uniquely inferred from the observational distribution $P(V)$ (Pearl, 1995; Tian & Pearl, 2003; Huang & Valtorta, 2006; Shpitser & Pearl, 2006; Lee & Bareinboim, 2020). There is also a large literature on estimating causal effects from finite samples drawn from $P(V)$ when the corresponding causal estimand is in the form of covariate adjustment (or its sequential variants) (Rosenbaum & Rubin, 1983; Pearl & Robins, 1995; Robins et al., 2000; Bang & Robins, 2005; Van Der Laan & Rubin, 2006; Hill, 2011), including doubly robust estimators for addressing model misspecification (Robins et al., 1994; Bang & Robins, 2005; Van Der Laan & Rubin, 2006; Rotnitzky & Smucler, 2020; Smucler et al., 2020; Fulcher et al., 2020). Recently, machine learning (ML) based methods have been developed for estimating any causal effects from finite samples whenever they are identifiable given a causal graph (Jung et al., 2020a;b; 2021).

Despite the power of these results, their applicability is contingent upon one having a causal graph, which may be hard to manually specify. In practical settings, one may attempt to learn the causal graph using structural learning algorithms from the available observational data (Pearl, 2000; Spirtes et al., 2000; Peters et al., 2017). Still, in principle, only a *Markov equivalence class (MEC)* of the underlying causal graph can be inferred from non-experimental data (Spirtes et al., 2000; Zhang, 2008b) without assumptions about the underlying causal mechanisms (Peters et al., 2017). There is a growing interest in causal identification in MECs (Zhang, 2008a; Perkovic et al., 2017; Jaber et al., 2018a;b). In particular, an algorithm called IDP has recently been developed for identifying causal effects in a MEC represented by a *partial ancestral graph (PAG)* (Jaber et al., 2019), which is both sufficient and necessary (i.e., complete). PAGs are learnable from observational data using causal structural learning algorithms (e.g. FCI (Zhang, 2008b)).

Even though these are quite general results, it remains an open challenge to estimate the resulting causal expressions from finite samples. For concreteness, consider the PAG in Fig. 1 as an exam-

[1]Department of Computer Science, Purdue University, USA [2]Department of Computer Science, Iowa State University, USA [3]Department of Computer Science, Columbia University, USA. Correspondence to: Yonghan Jung <jung222@purdue.edu>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

ple. The IDP algorithm identifies $P_x(y_1, y_2, y_3, y_4) = P(y_4|y_3, y_2, y_1, x, r)P(y_1)\sum_r P(y_2, y_3|x, r)P(r)$. The only viable general-purpose method currently available for estimating arbitrary causal estimands like this is the "plug-in" estimators (Casella & Berger, 2002), which estimate each conditional probability in the estimand (e.g., $P(y_4|y_3, y_2, y_1, x, r)$), called *nuisance functions* or *nuisances* in short, often by assuming a parametric model, and plug them into the equation. However, plug-in estimators are vulnerable to model misspecification in that all nuisance models need to be correctly specified for the estimator to be consistent. They also often suffer from biases in estimating the nuisances. In recent years, it is common to learn nuisance functions using highly flexible ML models, particularly in high-dimensional settings, including methods such as random forests (Breiman, 2001), boosted regression trees (Freund et al., 1996), and deep neural networks (Bengio, 2009). In practice, these ML methods inherently trade off regularization bias with overfitting often causing acute bias in the plug-in estimators of the target estimand such that these estimators will not achieve desirable $\sqrt{N}$-consistency (Chernozhukov et al., 2018), where N is the sample size.

We will exploit in this paper the *double/debiased machine learning* (DML) framework proposed in (Chernozhukov et al., 2018). This framework provides estimators that achieve $\sqrt{N}$-consistency with respect to the target estimand while admitting the use of highly flexible ML methods for estimating the nuisances at a slower $N^{-1/4}$ rate convergence ('*debiasedness*'). DML has been applied in causal inference including in the context of the backdoor/ignorability and instrumental variables (Robins et al., 1994; Bang & Robins, 2005; Van Der Laan & Rubin, 2006; Díaz & van der Laan, 2013; Benkeser et al., 2017; Kennedy et al., 2017; Rotnitzky & Smucler, 2020; Smucler et al., 2020; Colangelo & Lee, 2020) and in some specific settings (Toth & van der Laan, 2016; Rudolph & van der Laan, 2017; Fulcher et al., 2020; Kennedy, 2020a; Bhattacharya et al., 2020). Recently, DML has been used for estimating causal effects when the causal graph is fully specified (Jung et al., 2021).

Our goal will be to develop a general estimator for any identifiable causal effects in PAGs (when the causal graph is unknown). In particular, we will develop a DML estimator for identifiable causal effects in PAGs, named *DML-IDP*, by deriving their *influence functions (IF)* based on the semiparametric theory (Van der Vaart, 2000). Our results fill in a gap for a purely data-driven, end-to-end solution to causal effects estimation, i.e., from observational data $\mathcal{D} \rightarrow$ PAG G by structure learning algorithm $\rightarrow$ identifiability of target effect $P_x(y)$ by IDP $\rightarrow$ **estimating $P_x(y)$ from $\mathcal{D}$ by DML-IDP**. Specifically, our contributions are as follows:

1. We develop a complete systematic procedure that derives an IF for any identifiable causal effects in a PAG over discrete endogenous variables.

2. We develop a DML estimator (DML-IDP) for any identifiable causal effects in a PAG with discrete variables, which enjoy debiasedness and doubly robustness against model misspecification and biases in nuisances estimation. Experimental studies corroborate with the theory.

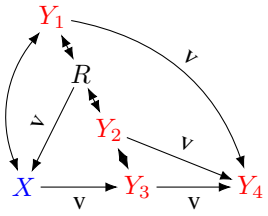


Figure 1: An example PAG. Nodes representing the treatment (X) and outcome ($\mathbf{Y}$) are marked in blue and red respectively. Causal effect $P_x(\mathbf{y})$ is identifiable. 'v' on edges stands for '*visible*' edges.

The proofs are provided in Appendix B in suppl. material.

2. Preliminaries

Each variable is represented with a capital letter (V) and its realized value with the small letter (v). We use bold letters ($\mathbf{V}$) to denote sets of variables. We use $I_{\mathbf{v}'}(\mathbf{V})$ to represent the indicator function such that $I_{\mathbf{v}'}(\mathbf{V}) = 1$ if and only if $\mathbf{V} = \mathbf{v}'$; $I_{\mathbf{v}'}(\mathbf{V}) = 0$ otherwise. For function $f(\mathbf{v})$ and a distribution $P(\mathbf{v})$, $\mathbb{E}_P[f(\mathbf{V})] \equiv \sum_{\mathbf{v}} f(\mathbf{v})P(\mathbf{v})$, and $\|f(\mathbf{V})\|_2 \equiv \sqrt{\mathbb{E}_P[(f(\mathbf{V}))^2]}$. $\widehat{f}$ is said to converge to f at rate r_N if $\|\widehat{f}(\mathbf{V}) - f(\mathbf{V})\|_2 = O_P(1/r_N)$.

Structural Causal Models. We use the language of structural causal models (SCMs) as our basic semantical framework (Pearl, 2000). Each SCM M over a set of variables $\mathbf{V}$ induces a distribution $P(\mathbf{v})$ and a causal graph G that is a directed acyclic graph (DAG) with bidirected arrows (edges). Solid-directed arrows encode functional relationships between observed variables, and bidirected arrows encode unobserved latent confounders. Within the structural semantics, performing an intervention and setting $\mathbf{X} = \mathbf{x}$ is represented through the do-operator, $do(\mathbf{X} = \mathbf{x})$, which encodes the operation of replacing the original equations of $\mathbf{X}$ by the constant $\mathbf{x}$ and induces a submodel $M_{\mathbf{x}}$ and an interventional distribution $P(\mathbf{v}|do(\mathbf{X} = \mathbf{x})) \equiv P_{\mathbf{x}}(\mathbf{v})$.

Partial Ancestral Graphs (PAGs). Given non-experimental data, only a *Markov equivalence class (MEC)* of the underlying causal graph can be inferred which includes a set of graphs with the same conditional independences (Zhang, 2007). A PAG provides a graphical representation of a MEC. PAGs may contain directed ($\rightarrow$) or bidirected ($\leftrightarrow$) edges, representing ancestral relations, and

edges with circles (e.g., $\{\circ\!\!\rightarrow, \circ\!\!-\!\!\circ\}$) indicating structural uncertainty (see Figs. 1 and 2 for example PAGs).

Given a PAG, a path between X and Y is *potentially directed* from X to Y if there is no arrowhead $\{<, >\}$ on the path pointing towards X. Y is called a *possible descendant* of X and X a *possible ancestor* of Y and denoted $X \in An(Y)$ if there is a potentially directed path from X to Y. Y is called a *possible child* of X and denoted $Y \in Ch(X)$, and X a *possible parent* of Y and denoted $X \in Pa(Y)$, if they are adjacent and the edge is not into X. By stipulation, $X \in An(X)$, $X \in Pa(X)$, and $X \in Ch(X)$. For a set of nodes $\mathbf{X}$, we have $Pa(\mathbf{X}) = \bigcup_{X\in\mathbf{X}} Pa(X)$ and $Ch(\mathbf{X}) = \bigcup_{X\in\mathbf{X}} Ch(X)$. If the edge marks on a path between X and Y are all circles, we call the path a circle path. We refer to the closure of nodes connected with circle paths as a *bucket*. Nodes $\mathbf{V}$ in a PAG G are partitioned into a unique set of buckets $\mathbf{V} = \bigcup_{i=1}^{n} \mathbf{B}_i$. There exists a topological order over buckets $\mathbf{B}_1 \prec \cdots \prec \mathbf{B}_n$ that defines a partial order over $\mathbf{V}$, which is valid in all the causal graphs in the MEC. This is named a *partial topological order (PTO)* and could be assigned by (Jaber et al., 2018a, Algo. 2). Given a PTO $\prec$ and a set $\mathbf{C} \subseteq \mathbf{V}$, we denote $\text{pre}_{\mathbf{C}}(\mathbf{B}_i) \equiv (\bigcup_{j<i} \mathbf{B}_j) \cap \mathbf{C}$ and use $\text{pre}(\mathbf{B}_i) \equiv \text{pre}_{\mathbf{V}}(\mathbf{B}_i)$. An *inducing path* is a path on which every node V_i (except for the endpoints) is a *collider* on the path and every collider is an ancestor of an endpoint. A directed edge $X \rightarrow Y$ in a PAG is *visible* and denoted $X \overset{\text{v}}{\rightarrow} Y$ if there exists no causal graph in the corresponding MEC where there is an inducing path between X and Y that is into X. Given a PAG G and a set $\mathbf{C} \subseteq \mathbf{V}$, $G(\mathbf{C})$ denotes the subgraph composed of nodes $\mathbf{C}$ and edges therein.

Causal Effect Identification. Given a DAG G over $\mathbf{V}$, an effect $P_{\mathbf{x}}(\mathbf{y})$ where $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$ is *identifiable* if $P_{\mathbf{x}}(\mathbf{y})$ is computable from the distribution $P(\mathbf{v})$ in any SCM that induces G (Pearl, 2000, p. 77). One key notion is called *confounded components (for short, C-component)* : closures of nodes connected with a path composed solely of bi-directed edges $V_i \leftrightarrow V_j$ (Tian & Pearl, 2002).

Given a PAG G over $\mathbf{V}$, a query $P_{\mathbf{x}}(\mathbf{y})$ is *identifiable* if and only if $P_{\mathbf{x}}(\mathbf{y})$ is identifiable with the same expression in every DAG in the MEC represented by the PAG G. A complete identification algorithm in PAGs called IDP has been developed (Jaber et al., 2019) (also presented in Appendix A for convenience) based on *possible C-component* (PC-component) and *definite C-component* (DC-component):

Definition 1 (PC **&** DC**-component** (Jaber et al., 2018a))**.** In a PAG (or its subgraph), two nodes are in the same PC-*component* if there is a path between them s.t. (1) all non-endpoint nodes along the path are colliders, and (2) none of the edges is visible. Two nodes are in the same DC-*component* if they are connected with a bi-directed path.

For a set of variables $\mathbf{X}$, we will use $\mathcal{C}(\mathbf{X})$ to denote the union of the PC-components that contain variables in $\mathbf{X}$. For any $\mathbf{C} \subseteq \mathbf{V}$, the quantity $Q[\mathbf{C}] \equiv P_{\mathbf{v}\setminus\mathbf{c}}(\mathbf{c})$, called a C-*factor*, is defined as the distribution of $\mathbf{C}$ under an intervention on $\mathbf{V}\setminus\mathbf{C}$. IDP algorithm is based on the following results for identification and decomposition of C-factors.

Proposition 1 (Jaber et al., 2018b)**.** *Let G be a PAG over $\mathbf{V}$, $\mathbf{T} = \cup_{i=1}^{m} \mathbf{B}_i$ be the union of a set of buckets, and $\mathbf{X} \subseteq \mathbf{T}$ be a bucket. Given $P_{\mathbf{v}\setminus\mathbf{t}}$ (i.e., $Q[\mathbf{T}]$) and a PTO $\mathbf{B}_1 \prec \cdots \prec \mathbf{B}_m$ with respect to $G(\mathbf{T})$, $Q[\mathbf{T}\setminus\mathbf{X}]$ is identifiable if and only if $\mathcal{C}(\mathbf{X}) \cap Ch(\mathbf{X}) \subseteq \mathbf{X}$ in $G(\mathbf{T})$. If identifiable,*

$$Q[\mathbf{T}\setminus\mathbf{X}] = \frac{P_{\mathbf{v}\setminus\mathbf{t}}}{\mathcal{Q}_{\mathbf{S}_\mathbf{X}}} \sum_{\mathbf{x}} \mathcal{Q}_{\mathbf{S}_\mathbf{X}},$$

where $\mathcal{Q}_{\mathbf{S}_\mathbf{X}} \equiv \prod_{i|\mathbf{B}_i\subseteq\mathbf{S}_\mathbf{X}} P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_i|pre_{\mathbf{T}}(\mathbf{b}_i))$ and $\mathbf{S}_\mathbf{X} = \bigcup_{X\in\mathbf{X}} \mathbf{S}_X$ with $\mathbf{S}_X$ being the DC-component of X in $G(\mathbf{T})$.

Definition 2 (**Region** $\mathcal{R}_{\mathbf{A}}^{\mathbf{C}}$ (Jaber et al., 2019))**.** Given a PAG G over $\mathbf{V}$ and $\mathbf{A} \subseteq \mathbf{C} \subseteq \mathbf{V}$, the *region* of $\mathbf{A}$ w.r.t. $\mathbf{C}$, denoted $\mathcal{R}_{\mathbf{A}}^{\mathbf{C}}$, is the union of the buckets in $G(\mathbf{C})$ that contain nodes in the PC-component $\mathcal{C}(\mathbf{A})$ of $\mathbf{A}$ in $G(\mathbf{C})$.

Proposition 2 (Jaber et al., 2019)**.** *Given a PAG G over $\mathbf{V}$ and a set $\mathbf{C} \subseteq \mathbf{V}$, $Q[\mathbf{C}]$ can be decomposed as $Q[\mathbf{C}] = \frac{Q[\mathcal{R}_\mathbf{A}]\cdot Q[\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_\mathbf{A}}]}{Q[\mathcal{R}_\mathbf{A}\cap\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_\mathbf{A}}]}$ for any $\mathbf{A} \subseteq \mathbf{C}$, where $\mathcal{R}_{(\cdot)} = \mathcal{R}_{(\cdot)}^{\mathbf{C}}$.*

Semiparametric Theory. We aim to estimate a target estimand $\psi \equiv \Psi(P)$ that is a functional of $P(\mathbf{V})$ (e.g., $\Psi(P) = \sum_z P(y|x,z)P(z)$) from finite samples $\mathcal{D} = \{\mathbf{V}_{(i)}\}_{i=1}^{N}$ drawn from P. Let a *parametric submodel* $P_t \equiv P(\mathbf{v})(1 + tg(\mathbf{v}))$ for any $t \in \mathbb{R}$ and bounded mean-zero function $g(\cdot)$ over random variables $\mathbf{V}$. If a functional $\Psi(P_t)$ is pathwise (formally, Gâteaux) differentiable at $t = 0$, then there exists a function $\phi(\mathbf{V}; \psi, \eta)$ (shortly ϕ), called an *influence function (IF)* for ψ, where $\eta = \eta(P)$ stands for the set of nuisance functions comprising ϕ, satisfying $\mathbb{E}_P[\phi] = 0$, $\mathbb{E}_P[\phi^2] < \infty$, and $\frac{\partial}{\partial t}\Psi(P_t)|_{t=0} = \mathbb{E}_P[\phi(\mathbf{V}; \psi, \eta)S_t(\mathbf{V}; t = 0)]$ where $S_t(\mathbf{v}; t = 0) \equiv \frac{\partial}{\partial t} \log P_t(\mathbf{v})|_{t=0}$ is the score function (Van der Vaart, 2000, Chap. 25). Given an IF ϕ, a Regular and Asymptotic Linear (RAL) estimator T_N can be constructed satisfying $T_N - \psi = \frac{1}{N}\sum_{i=1}^{N} \phi(\mathbf{V}_{(i)}; \psi, \eta) + o_P(N^{-1/2})$. When the IF can be decomposed as $\phi(\mathbf{V}; \psi, \eta) = \mathcal{V}(\mathbf{V}; \eta) - \psi$ for some function $\mathcal{V}(\mathbf{V}; \eta)$, called the *uncentered influence function (UIF)*, the corresponding RAL estimator is $T_N = \frac{1}{N}\sum_{i=1}^{N} \mathcal{V}(\mathbf{V}_{(i)}, \widehat{\eta})$ where $\widehat{\eta}$ denotes nuisances estimated from sample $\mathcal{D}$ (Kennedy, 2020a). We will focus on deriving UIFs in this paper. Once we have a UIF the corresponding IF could be expressed as $\phi(\mathbf{V}; \psi, \eta) = \mathcal{V}(\mathbf{V}; \eta) - \mathbb{E}_P[\mathcal{V}(\mathbf{V}; \eta)]$.

We make the following assumptions throughout the paper, which ascertain that the estimands will be pathwise differentiable.

Assumption 1 (Discreteness of variables). *The set of variables* $\mathbf{V}$ *in the PAG are discrete.*

Assumption 2 (Positivity of conditional probabilities). *There exists a fixed* $\epsilon \in (0, 0.5)$ *s.t.* $P(\mathbf{a}|\mathbf{b}) > \epsilon$ *for any* $\mathbf{A}, \mathbf{B} \subseteq \mathbf{V}$.

The results can be extended to continuous cases with additional conditions such that the corresponding influence functions are well-defined (Robins, 2000; Neugebauer & van der Laan, 2007; Díaz & van der Laan, 2013; Kennedy et al., 2017; Chernozhukov et al., 2019).

Double/Debiased Machine Learning (DML). DML methods (Chernozhukov et al., 2018) are based on two ideas: (1) use a *Neyman orthogonal score*[1] to estimate the target ψ, and (2) use *cross-fitting*[2] to construct the estimator. DML estimators guarantee $\sqrt{N}$-consistency even when the estimates $\widehat{\eta}$ of (possibly high-dimensional) nuisance functions converge at a much slower $N^{-1/4}$ rate ('*debiasedness*'), allowing the use of a broad array of modern ML methods that do not meet certain smoothness/complexity restrictions (i.e., *Donsker* class). Neyman-orthogonal scores may coincide with IFs - a fact we exploit in this paper.

3. IFs for Canonical Expressions

Before deriving IFs for any identifiable causal effects in PAGs, in this section, we derive IFs for two typical functionals that often appear in the expressions of causal effects, called here *canonical expressions*.

3.1. Canonical expression 1

Definition 3 (Canonical expression 1 (CE-1)). Let $\mathbf{T} = \{\mathbf{B}_1 < \cdots < \mathbf{B}_n\}$ be a set of ordered sets[3]. Let $\mathbf{C} \subseteq \mathbf{T}$ be a subset composed of $\mathbf{B}_i \in \mathbf{T}$ and $\mathbf{A}$ be a subset of variables contained in $\mathbf{C}$. A quantity $\mathcal{Q}$ is said to be (in the form of) a *canonical expression 1 (CE-1)* if it is in the following form:

$$\mathcal{Q} = \sum_{\mathbf{a}} \prod_{\mathbf{B}_i \in \mathbf{C}} P(\mathbf{b}_i | \text{pre}_{\mathbf{T}}(\mathbf{b}_i)). \tag{1}$$

For concreteness, we show the causal effect $P_{\mathbf{x}}(y)$ (for $\mathbf{X} = \{X_1, X_2\}$) in the PAG in Fig. 2a can be expressed as a CE-1 as follows:

Given a PTO $\mathbf{V} = \{C \prec B \prec A \prec X_1 \prec Z \prec X_2 \prec Y\}$, we have $Q[\mathbf{V} \setminus X_2]$ is identifiable from $Q[\mathbf{V}] = P(\mathbf{V})$ by Prop. 1 as X_2 is a bucket satisfying $\mathcal{C}(X_2) \cap Ch(X_2) = \{X_2\}$ and $\mathbf{S}_{X_2} = \{X_2\}$, and we obtain $Q[\mathbf{V} \setminus X_2] = P_{x_2}(\mathbf{v} \setminus x_2)$ as follow:

$$Q[\mathbf{V} \setminus X_2] = \frac{P(\mathbf{v})}{P(x_2|\text{pre}(x_2))} = P(y|\text{pre}(y))P(\text{pre}(x_2)).$$

For $\mathbf{T} \equiv \mathbf{V} \setminus \{X_2\}$, $Q[\mathbf{T} \setminus X_1]$ is identifiable from $Q[\mathbf{T}]$ by Prop. 1 as X_1 is a bucket satisfying $\mathcal{C}(X_1) \cap Ch(X_1) = \{X_1\}$ and $\mathbf{S}_{X_1} = \{X_1\}$, and we obtain $Q[\mathbf{T} \setminus X_1] = P_{x_1,x_2}(\mathbf{t} \setminus \{x_1\})$ as follow:

$$\begin{aligned} Q[\mathbf{T} \setminus X_1] &= \frac{P_{x_2}(\mathbf{t})}{P_{x_2}(x_1|\text{pre}_{\mathbf{T}}(x_1))} \\ &= P(y|\text{pre}(y))P(z|\text{pre}(z))P(a,b,c) \end{aligned}$$

by the equality $P_{x_2}(x_1|\text{pre}_{\mathbf{T}}(x_1)) = P(x_1|\text{pre}_{\mathbf{T}}(x_1))$. Finally, the causal effect $P_{\mathbf{x}}(y)$ is given as a CE-1 as:

$$P_{\mathbf{x}}(y) = \sum_{z,a,b,c} Q[\mathbf{T} \setminus X_1]. \tag{2}$$

We derive an IF for functionals in the form of CE-1 as follows:

Lemma 1 (UIF for CE-1). *Let the target estimand* $\psi = \mathcal{Q}$ *be a CE-1 given by Eq. (1) in Def. 3. Let* $\mathbf{Y} \equiv \mathbf{C} \setminus \mathbf{A}$, *and* $\mathbf{X} \equiv \mathbf{T} \setminus \mathbf{C} \equiv \{\mathbf{B}_{j_1} < \cdots < \mathbf{B}_{j_m}\}$ *where* $\mathbf{B}_{j_s} \in \mathbf{T}$. *Let* $\mathbf{C}$ *be partitioned with respect to* $\mathbf{X}$ *as* $\mathbf{C} = \bigcup_{k=0}^{m} \mathbf{C}_k$, *where* $\mathbf{C}_k \equiv \{\mathbf{B}_r \in \mathbf{C} : j_k < r < j_{k+1}\} \equiv \{\mathbf{B}_{k_{\min}} < \cdots < \mathbf{B}_{k_{\max}}\}$ *with* $j_0 \equiv 0$ *and* $j_{m+1} \equiv n+1$. *Let* P_π *be a distribution over* $\mathbf{T}$ *given by* $P_\pi \equiv I_{\mathbf{x}}(\mathbf{X}) \prod_{\mathbf{B}_i \in \mathbf{C}} P(\mathbf{B}_i | pre_{\mathbf{T}}(\mathbf{B}_i))$. *Then,* $\mathcal{V}(\mathbf{T}; \eta = (\boldsymbol{\omega}, \boldsymbol{\theta}))$ *in the following is a UIF for* ψ*:*

$$\mathcal{V}(\mathbf{T}; \eta = (\boldsymbol{\omega}, \boldsymbol{\theta})) = \theta_{0,1} + \sum_{\substack{k=1 \\ \mathbf{C}_k \neq \emptyset}}^{m} \omega_k \left(\theta_{k,1} - \theta_{k,2}\right), \tag{3}$$

where $\boldsymbol{\omega} \equiv \{\omega_k |\ \mathbf{C}_k \neq \emptyset,\ k \in \{1, \cdots, m\}\}$ *and* $\boldsymbol{\theta} \equiv \{\theta_{0,1}\} \cup \{(\theta_{k,1}, \theta_{k,2}) |\ \mathbf{C}_k \neq \emptyset,\ k \in \{1, \cdots, m\}\}$ *are nuisances given by*

$$\begin{aligned} \omega_k &\equiv \prod_{r=1}^{k} \frac{I_{\mathbf{b}_{j_r}}(\mathbf{B}_{j_r})}{P(\mathbf{B}_{j_r} | pre_{\mathbf{T}}(\mathbf{B}_{j_r}))}, \\ \theta_{k,1} &\equiv \mathbb{E}_{P_\pi}\left[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{B}_{k_{\max}}, pre_{\mathbf{T}}(\mathbf{B}_{k_{\max}})\right], \\ \theta_{k,2} &\equiv \mathbb{E}_{P_\pi}\left[I_{\mathbf{y}}(\mathbf{Y}) | pre_{\mathbf{T}}(\mathbf{B}_{k_{\min}})\right], \end{aligned}$$

where $\theta_{0,1} = \mathbb{E}_{P_\pi}[I_{\mathbf{y}}(\mathbf{Y})]$ *if* $\mathbf{C}_0 = \emptyset$.

For concreteness, we apply Lemma 1 to derive a UIF for $\psi \equiv P_{x_1,x_2}(y)$ in Fig. 2a which is identified as a CE-1 given in Eq. (2).

[1] A Neyman orthogonal score is a function ϕ satisfying $\mathbb{E}_P[\phi(\mathbf{V}; \psi, \eta^*)] = 0$ and $\frac{\partial}{\partial \eta}\mathbb{E}_P[\phi(\mathbf{V}; \psi, \eta)]|_{\eta=\eta^*} = 0$, where η^* denotes the true nuisance.

[2] The cross-fitting technique uses distinct sets of samples in model training and estimator's evaluation.

[3] We use $\mathbf{W} = \{\mathbf{B}_1 < \cdots < \mathbf{B}_k\}$ to denote a set of ordered sets $\mathbf{W} = \{\mathbf{B}_1, \cdots, \mathbf{B}_k\}$ or a union of ordered sets $\mathbf{W} = \cup_{i=1}^{k} \mathbf{B}_i$ depending on the context.

Illustration 1 (UIF for $P_{x_1,x_2}(y)$ in Fig. 2a). *Let* $\mathbf{T} = \{C \prec B \prec A \prec X_1 \prec Z \prec X_2 \prec Y\}$, $\mathbf{C} = \{C \prec B \prec A \prec Z \prec Y\}$, *and* $\mathbf{X} = \{X_1 \prec X_2\}$. *We have* $\mathbf{C}_0 = \{C \prec B \prec A\}$, $\mathbf{C}_1 = \{Z\}$, *and* $\mathbf{C}_2 = \{Y\}$. *Then Lemma 1 gives a UIF for* ψ *as*

$$\mathcal{V}_{P_{\mathbf{x}}(y)} = \theta_{0,1} + \omega_1 (\theta_{1,1} - \theta_{1,2}) + \omega_2 (\theta_{2,1} - \theta_{2,2}), \quad (4)$$

where

$$\omega_1 = I_{x_1}(X_1)/P(X_1|pre(X_1))$$
$$\omega_2 = I_{x_1,x_2}(X_1, X_2)/P(X_1|pre(X_1))P(X_2|pre(X_2)),$$

and for

$$P_\pi \equiv I_{x_1,x_2}(X_1, X_2)P(A, B, C)P(Z|pre(Z))P(Y|pre(Y)),$$

$\theta_{0,1} = \mathbb{E}_{P_\pi}[I_y(Y)|pre(X_1)]$, $\theta_{1,1} = \mathbb{E}_{P_\pi}[I_y(Y)|pre(X_2)]$, $\theta_{1,2} = \mathbb{E}_{P_\pi}[I_y(Y)|pre(Z)]$; *and* $\theta_{2,1} = \mathbb{E}_{P_\pi}[I_y(Y)|\mathbf{T}] = I_y(Y)$ *and* $\theta_{2,2} = \mathbb{E}_{P_\pi}[I_y(Y)|pre(Y)]$.

3.2. Canonical expression 2

Definition 4 (Canonical expression 2 (CE-2)). Let $\mathcal{Q}_1$ and $\mathcal{Q}_2$ be two CE-1s, then the quantity $\mathcal{Q} = \sum_{\mathbf{z}} (\mathcal{Q}_1 \times \mathcal{Q}_2)$ for some $\mathbf{Z} \subseteq \mathbf{V}$ is said to be (in the form of) a *canonical expression 2 (CE-2)*.

A broad class of causal effects are identified as a CE-2, including all joint interventional distributions ($P_x(\mathbf{v})$) when X is singleton (Jaber et al., 2018b, Thm. 1), as well as in the following scenario which follows from Prop. 1:

Corollary 1. *Let a PTO in PAG* G *over* $\mathbf{V}$ *be* $\mathbf{B}_1 \prec \cdots \prec \mathbf{B}_m$. *Let* $\mathbf{X}, \mathbf{Y} \subset \mathbf{V}$ *with* $\mathbf{X}$ *being a bucket. If* $\mathcal{C}(\mathbf{X}) \cap Ch(\mathbf{X}) \subseteq \mathbf{X}$, *then* $P_{\mathbf{x}}(\mathbf{y})$ *is identifiable and given by*

$$P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{v}\setminus(\mathbf{x}\cup\mathbf{y})} \mathcal{Q}_{\mathbf{V}\setminus\mathbf{S}_\mathbf{X}} \times \mathcal{Q}_{\mathbf{S}_\mathbf{X}\setminus\mathbf{X}}, \quad (5)$$

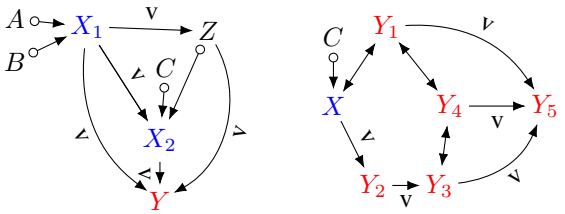
where $\mathcal{Q}_{\mathbf{V}\setminus\mathbf{S}_\mathbf{X}} \equiv \prod_{\mathbf{B}_i \subseteq \mathbf{V}\setminus\mathbf{S}_\mathbf{X}} P(\mathbf{b}_i|pre(\mathbf{b}_i))$, $\mathcal{Q}_{\mathbf{S}_\mathbf{X}\setminus\mathbf{X}} \equiv \sum_{\mathbf{x}} \prod_{\mathbf{B}_i \subseteq \mathbf{S}_\mathbf{X}} P(\mathbf{b}_i|pre(\mathbf{b}_i))$, *and* $\mathbf{S}_\mathbf{X} = \bigcup_{X\in\mathbf{X}} \mathbf{S}_X$ *with* $\mathbf{S}_X$ *being the DC-component of* X.

Eq. (5) is a CE-2 where $\mathcal{Q}_{\mathbf{V}\setminus\mathbf{S}_\mathbf{X}}$ and $\mathcal{Q}_{\mathbf{S}_\mathbf{X}\setminus\mathbf{X}}$ are CE-1s. As a concrete example, consider the PAG in Fig. 2b with a PTO $\mathbf{V} = \{C \prec X \prec Y_1 \prec Y_2 \prec Y_3 \prec Y_4 \prec Y_5\}$. Since X is a bucket and satisfies $\mathcal{C}(X) \cap Ch(X) = \{X\}$ with $\mathcal{C}(X) = \{X, C, Y_1, Y_4, Y_3\}$ and $Ch(X) = \{X, Y_2\}$, the causal effect $P_x(\mathbf{y})$ where $\mathbf{Y} = \{Y_1, \cdots, Y_5\}$ is identifiable by Coro. 1 and given by

$$P_x(\mathbf{y}) = \sum_c \mathcal{Q}_{\mathbf{V}\setminus\mathbf{S}_X} \mathcal{Q}_{\mathbf{S}_X\setminus X}, \quad (6)$$

where $\mathbf{S}_X = \{X, Y_1, Y_3, Y_4\}$, $\mathbf{V}\setminus\mathbf{S}_X = \{C, Y_2, Y_5\}$, $\mathcal{Q}_{\mathbf{V}\setminus\mathbf{S}_X} \equiv P(y_5|\text{pre}(y_5))P(y_2|\text{pre}(y_2))P(c)$, and $\mathcal{Q}_{\mathbf{S}_X\setminus X} \equiv \sum_{x'} P(y_3, y_4|y_1, y_2, x', c)P(y_1, x'|c)$.

We derive an IF for CE-2 as follows:



(a) (b)

Figure 2: Example PAGs. Causal effects $P_{\mathbf{x}}(\mathbf{y})$ are identifiable and given by **(a)** CE-1, **(b)** CE-2.

Lemma 2 (UIF for CE-2). *Let the target estimand* $\psi = \mathcal{Q}$ *be a CE-2 given in Def. 4. Let* $\mathcal{V}_i$ *be a UIF for the CE-1* $\mathcal{Q}_i$ *given in Lemma 1 and* $\mu_i \equiv \mathbb{E}_P[\mathcal{V}_i]$ *for* $i \in \{1, 2\}$. *Then,* $\mathcal{V}(\mathbf{V}; \eta)$ *below is a UIF for* ψ*:*

$$\mathcal{V}(\mathbf{V}; \eta) = \sum_{\mathbf{z}} (\mathcal{V}_1\mu_2 + (\mathcal{V}_2 - \mu_2)\mu_1). \quad (7)$$

Lemma 2 provides a UIF for any causal effects that are identifiable by Coro. 1. For a concrete example, we will use Lemma 2 to derive a UIF for $\psi \equiv P_x(\mathbf{y})$ in Fig. 2b identified by Coro. 1 as given in Eq. (6).

Illustration 2 (UIF for $P_x(\mathbf{y})$ in Fig. 2b). *A UIF for* $P_x(\mathbf{y})$ *in Eq. (6) is given by Lemma 2 as*

$$\mathcal{V}_{P_x(\mathbf{y})} = \sum_c \left(\mathcal{V}_{\mathbf{V}\setminus\mathbf{S}_\mathbf{X}}\mu_{\mathbf{S}_\mathbf{X}\setminus\mathbf{X}} + (\mathcal{V}_{\mathbf{S}_\mathbf{X}\setminus\mathbf{X}} - \mu_{\mathbf{S}_\mathbf{X}\setminus\mathbf{X}})\mu_{\mathbf{V}\setminus\mathbf{S}_\mathbf{X}}\right), \quad (8)$$

where $\mathcal{V}_{\mathbf{V}\setminus\mathbf{S}_\mathbf{X}}$ *is a UIF for* $\mathcal{Q}_{\mathbf{V}\setminus\mathbf{S}_X}$ *and, by Lemma 1, is given with* $\mathbf{V} = \{C \prec X \prec Y_1 \prec Y_2 \prec Y_3 \prec Y_4 \prec Y_5\}$ *as*

$$\mathcal{V}_{\mathbf{V}\setminus\mathbf{S}_\mathbf{X}} = \theta^a_{0,1} + \omega^a_1(\theta^a_{1,1} - \theta^a_{1,2}) + \omega^a_2(\theta^a_{2,1} - \theta^a_{2,2}),$$

where

$$\omega^a_1 = I_{x,y_1}(X, Y_1)/P(X|C)P(Y_1|X, C)$$
$$\omega^a_2 = \omega^a_1 \times I_{y_3,y_4}(Y_3, Y_4)/\left(P(Y_3|pre(Y_3))P(Y_4|pre(Y_4))\right),$$

and for

$$P_{\pi^a} \equiv I_{x,y_1,y_3,y_4}(X, Y_1, Y_3, Y_4)P(C) \times P(Y_2|pre(Y_2))P(Y_5|pre(Y_5)),$$

and $I^a \equiv I_{c,y_2,y_5}(C, Y_2, Y_5)$, $\theta^a_{0,1} = \mathbb{E}_{P_{\pi^a}}[I^a|C]$, $\theta^a_{1,1} = \mathbb{E}_{P_{\pi^a}}[I^a|Y_2, pre(Y_2)]$, $\theta^a_{1,2} = \mathbb{E}_{P_{\pi^a}}[I^a|pre(Y_2)]$, $\theta^a_{2,1} = I^a$, *and* $\theta^a_{2,2} = \mathbb{E}_{P_{\pi^a}}[I^a|pre(Y_5)]$.

Also, $\mathcal{V}_{\mathbf{S}_\mathbf{X}\setminus\mathbf{X}}$ *is a UIF for* $\mathcal{Q}_{\mathbf{S}_\mathbf{X}\setminus\mathbf{X}}$ *given by Lemma 1 as*

$$\mathcal{V}_{\mathbf{S}_\mathbf{X}\setminus\mathbf{X}} = \theta^b_{0,1} + \omega^b_1(\theta^b_{1,1} - \theta^b_{1,2}) + \omega^b_2(\theta^b_{2,1} - \theta^b_{2,2}),$$

where

$$\omega_1^b = I_c(C)/P(C)$$
$$\omega_2^b = \omega_1^b \times I_{y_2}(Y_2)/P(Y_2|pre(Y_2)),$$

and for

$$P_{\pi^b} \equiv I_{c,y_2}(C,Y_2)P(Y_3,Y_4|pre(Y_3))P(X,Y_1|C),$$

and $I^b \equiv I_{y_1,y_3,y_4}(Y_1,Y_3,Y_4)$, $\theta^b_{0,1} = \mathbb{E}_{P_{\pi^b}}[I^b]$, $\theta^b_{1,1} = \mathbb{E}_{P_{\pi^b}}[I^b|Y_1, pre(Y_1)]$, $\theta^b_{1,2} = \mathbb{E}_{P_{\pi^b}}[I^b|pre(X)]$, $\theta^b_{2,1} = I^b$, *and* $\theta^b_{2,2} = \mathbb{E}_{P_{\pi^2}}[I^b|pre(Y_3)]$.

Finally $\mu_{\mathbf{V}\setminus\mathbf{S_X}} \equiv \mathbb{E}_P[\mathcal{V}_{\mathbf{V}\setminus\mathbf{S_X}}]$, *and* $\mu_{\mathbf{S_X}\setminus\mathbf{X}} \equiv \mathbb{E}_P[\mathcal{V}_{\mathbf{S_X}\setminus\mathbf{X}}]$. *Refer Appendix A for derivation details.*

4. IFs for Causal Estimands

In this section, we derive IFs for any identifiable causal effects in PAGs, armed with IFs for the canonical expressions discussed in the previous section. We develop a complete algorithm for deriving IFs by recursively deriving IFs of C-factors $Q[\cdot]$ inspired by IDP algorithm (Jaber et al., 2019) which recursively identifies C-factors by repeated application of Prop. 1 or 2. We will first develop basic results for deriving IFs of C-factors corresponding to Prop. 1 and 2.

Prop. 1 computes $Q[\mathbf{T}\setminus\mathbf{X}]$ in terms of given $Q[\mathbf{T}]$. We rewrite Prop. 1 in a form more amenable for the purpose of deriving IFs:

Lemma 3. *Let G be a PAG over $\mathbf{V}$, $\mathbf{T} = \cup_{i=1}^m \mathbf{B}_i$ be the union of a set of buckets, and $\mathbf{X} \subseteq \mathbf{T}$ be a bucket. Given $Q[\mathbf{T}]$ and a PTO $\mathbf{B}_1 \prec \cdots \prec \mathbf{B}_m$ with respect to $G(\mathbf{T})$, $Q[\mathbf{T}\setminus\mathbf{X}]$ is identifiable if and only if $\mathcal{C}(\mathbf{X}) \cap Ch(\mathbf{X}) \subseteq \mathbf{X}$ in $G(\mathbf{T})$. When $Q[\mathbf{T}\setminus\mathbf{X}]$ is identifiable, letting $\mathbf{S_X} = \bigcup_{X\in\mathbf{X}} \mathbf{S}_X$ with $\mathbf{S}_X$ being the DC-component of X in $G(\mathbf{T})$, then $\mathbf{S_X}$ consists of a union of buckets. Denoting $\mathbf{S_X} = \{\mathbf{B}_{j_1}, \cdots, \mathbf{B}_{j_p}\}$ and $\mathbf{T}\setminus\mathbf{S_X} = \{\mathbf{B}_{i_1}, \cdots, \mathbf{B}_{i_q}\}$, $Q[\mathbf{T}\setminus\mathbf{X}]$ is given by*

$$Q[\mathbf{T}\setminus\mathbf{X}] = \mathcal{Q}_{\mathbf{T}\setminus\mathbf{S_X}} \times \mathcal{Q}_{\mathbf{S_X}\setminus\mathbf{X}}, \tag{9}$$

where $\mathcal{Q}_{\mathbf{T}\setminus\mathbf{S_X}} \equiv \prod_{\mathbf{B}_{i_r}\in\mathbf{T}\setminus\mathbf{S_X}} P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{i_r}|pre_{\mathbf{T}}(\mathbf{b}_{i_r}))$, *and* $\mathcal{Q}_{\mathbf{S_X}\setminus\mathbf{X}} \equiv \sum_{\mathbf{x}} \prod_{\mathbf{B}_{j_s}\in\mathbf{S_X}} P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{j_s}|pre_{\mathbf{T}}(\mathbf{b}_{j_s}))$.

For any $\mathbf{W} \subseteq \mathbf{V}$, we will use $\phi_{Q[\mathbf{W}]}$ to denote an IF for the C-factor $Q[\mathbf{W}]$, $\mathcal{V}_{Q[\mathbf{W}]}$ the corresponding UIF, and $\mu_{Q[\mathbf{W}]} \equiv \mathbb{E}_P[\mathcal{V}_{Q[\mathbf{W}]}]$. We derive an IF for $Q[\mathbf{T}\setminus\mathbf{X}]$ that is identified by Lemma 3 in terms of $\mathcal{V}_{Q[\mathbf{T}]}$ as follows:

Lemma 4 (IF of C-factors). *Suppose $\psi \equiv Q[\mathbf{T}\setminus\mathbf{X}]$ is identifiable via Lemma 3 and given by Eq. (9). Then, given $\mathcal{V}_{Q[\mathbf{T}]}$, $\mathcal{V} \equiv \mathcal{V}_{Q[\mathbf{T}\setminus\mathbf{X}]}$ below is a UIF for ψ:*

$$\mathcal{V} = \mathcal{V}_{\mathbf{S_X}\setminus\mathbf{X}}\mu_{\mathcal{V}_{\mathbf{T}\setminus\mathbf{S_X}}} + (\mathcal{V}_{\mathbf{T}\setminus\mathbf{S_X}} - \mu_{\mathcal{V}_{\mathbf{T}\setminus\mathbf{S_X}}})\mu_{\mathbf{S_X}\setminus\mathbf{X}}, \tag{10}$$

Algorithm 1 IFP$(\mathbf{x}, \mathbf{y}, G(\mathbf{V}), P)$

Input: Two disjoint sets $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$; A PAG G over $\mathbf{V}$; A distribution $P(\mathbf{v})$.
Output: Expression for UIF $\mathcal{V}_{P_{\mathbf{x}(\mathbf{y})}}$ or FAIL.
Let $\mathbf{D} = An(\mathbf{Y})_{G(\mathbf{V}\setminus\mathbf{X})}$.
$\mathcal{V}_{P_{\mathbf{x}(\mathbf{y})}} = \sum_{\mathbf{d}\setminus\mathbf{y}}$ DERIVEUIF $\left(\mathbf{D}, \mathbf{V}, P(\mathbf{V}), \mathcal{V}_{Q[\mathbf{V}]} = I_{\mathbf{v}}(\mathbf{V})\right)$
function DERIVEUIF $(\mathbf{C}, \mathbf{T}, Q = Q[\mathbf{T}], \mathcal{V} = \mathcal{V}_Q)$
  if $\mathbf{C} = \emptyset$, **then return** 1.
  if $\mathbf{C} = \mathbf{T}$, **then return** $\mathcal{V}$.
  $\{\mathbf{B}$ *denotes a bucket in* $G(\mathbf{T})$*;* $\mathcal{C}(\mathbf{B})$ *the PC-component of* $\mathbf{B}$ *in* $G(\mathbf{T})$*, and* $\mathcal{R}_{(\cdot)} \equiv \mathcal{R}^{\mathbf{C}}_{(\cdot)}$.$\}$
  if $\exists \mathbf{B} \subseteq \mathbf{T}\setminus\mathbf{C}$ s.t. $\mathcal{C}(\mathbf{B}) \cap Ch(\mathbf{B}) \subseteq \mathbf{B}$, **then**
    Compute $Q[\mathbf{T}\setminus\mathbf{B}]$ from Q via Lemma 3.
    if $Q[\mathbf{T}\setminus\mathbf{B}]$ is expressible as CE-1, **then**, Compute $\mathcal{V}_{Q[\mathbf{T}\setminus\mathbf{B}]}$ via Lemma 1.
    else if $Q[\mathbf{T}\setminus\mathbf{B}]$ is expressible as CE-2, **then**, Compute $\mathcal{V}_{Q[\mathbf{T}\setminus\mathbf{B}]}$ via Lemma 2.
    else, Compute $\mathcal{V}_{Q[\mathbf{T}\setminus\mathbf{B}]}$ via Lemma 4.
    return DERIVEUIF $\left(\mathbf{C}, \mathbf{T}\setminus\mathbf{B}, Q[\mathbf{T}\setminus\mathbf{B}], \mathcal{V}_{Q[\mathbf{T}\setminus\mathbf{B}]}\right)$.
  else if $\exists \mathbf{B} \subseteq \mathbf{C}$ s.t. $\mathcal{R}_{\mathbf{B}} \neq \mathbf{C}$, **then**
    return $(a) + (b) - (c)$, where
    $\{$ *Let* UIF$(\mathbf{W})$ = DERIVEUIF$(\mathbf{W}, \mathbf{T}, Q, \mathcal{V})$*;* IF$(\mathbf{W})$ = UIF$(\mathbf{W}) - \mathbb{E}_P[UIF(\mathbf{W})]$*;* ID$(\mathbf{W}) = \mathbb{E}_P[UIF(\mathbf{W})]\}$
    $(a) = \frac{\text{UIF}(\mathcal{R}_{\mathbf{B}})\cdot\text{ID}(\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{B}}})}{\text{ID}(\mathcal{R}_{\mathbf{B}}\cap\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{B}}})}$; $(b) = \frac{\text{IF}(\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{B}}})\cdot\text{ID}(\mathcal{R}_{\mathbf{B}})}{\text{ID}(\mathcal{R}_{\mathbf{B}}\cap\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{B}}})}$;
    $(c) = \frac{\text{ID}(\mathcal{R}_{\mathbf{B}})\cdot\text{ID}(\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{B}}})}{\text{ID}(\mathcal{R}_{\mathbf{B}}\cap\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{B}}})} \cdot \frac{\text{IF}(\mathcal{R}_{\mathbf{B}}\cap\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{B}}})}{\text{ID}(\mathcal{R}_{\mathbf{B}}\cap\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{B}}})}$.
  else return FAIL.
end function

where $(\mathcal{V}_{\mathbf{S_X}\setminus\mathbf{X}}, \mathcal{V}_{\mathbf{T}\setminus\mathbf{S_X}})$ *are UIFs for* $(\mathcal{Q}_{\mathbf{S_X}\setminus\mathbf{X}}, \mathcal{Q}_{\mathbf{T}\setminus\mathbf{S_X}})$ *respectively, given by*

$$\mathcal{V}_{\mathbf{S_X}\setminus\mathbf{X}} \equiv \sum_{\mathbf{x}}(\mathcal{V}_{j_1}\prod_{k=2}^{p}\mu_{j_k} + \sum_{k=2}^{p}\phi_{j_k}\prod_{\ell=1,\ell\neq k}^{p}\mu_{j_\ell}),$$
$$\mathcal{V}_{\mathbf{T}\setminus\mathbf{S_X}} \equiv \mathcal{V}_{i_1}\prod_{r=2}^{q}\mu_{i_r} + \sum_{r=2}^{q}\phi_{i_r}\prod_{\ell=1,\ell\neq r}^{q}\mu_{i_\ell},$$

where, for $c \in \{1, 2, \cdots, m\}$, $\mathcal{V}_c \equiv \frac{\sum_{\mathbf{t}\setminus\{\mathbf{b}_c, pre_{\mathbf{T}}(\mathbf{b}_c)\}}\mathcal{V}_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}\setminus pre_{\mathbf{T}}(\mathbf{b}_c)}\mu_{Q[\mathbf{T}]}} - \frac{\sum_{\mathbf{t}\setminus\{\mathbf{b}_c, pre_{\mathbf{T}}(\mathbf{b}_c)\}}\mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}\setminus pre_{\mathbf{T}}(\mathbf{b}_c)}\mu_{Q[\mathbf{T}]}} \cdot \frac{\sum_{\mathbf{t}\setminus pre_{\mathbf{T}}(\mathbf{b}_c)}\phi_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}\setminus pre_{\mathbf{T}}(\mathbf{b}_c)}\mu_{Q[\mathbf{T}]}}$, $\mu_c \equiv \mathbb{E}_P[\mathcal{V}_c]$, *and* $\phi_c \equiv \mathcal{V}_c - \mu_c$.

The following lemma derives an IF for the C-factor $Q[\mathbf{C}]$ from the IFs of C-factors over some subsets of C, corresponding to the C-factor decomposition in Prop. 2.

Lemma 5 (Decomposition of IFs). *For $\mathbf{A} \subseteq \mathbf{C} \subseteq \mathbf{V}$,*

$$\mathcal{V}_{Q[\mathbf{C}]} = (a) + (b) - (c), \tag{11}$$

where $(a) = \frac{\mathcal{V}_{Q[\mathcal{R}_{\mathbf{A}}]}\cdot\mu_{Q[\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{A}}}]}}{\mu_{Q[\mathcal{R}_{\mathbf{A}}\cap\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{A}}}]}}$, $(b) = \frac{\mu_{Q[\mathcal{R}_{\mathbf{A}}]}\cdot\phi_{Q[\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{A}}}]}}{\mu_{Q[\mathcal{R}_{\mathbf{A}}\cap\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{A}}}]}}$,
$(c) = \frac{\mu_{Q[\mathcal{R}_{\mathbf{A}}]}\cdot\mu_{Q[\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{A}}}]}}{\mu_{Q[\mathcal{R}_{\mathbf{A}}\cap\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{A}}}]}} \cdot \frac{\phi_{Q[\mathcal{R}_{\mathbf{A}}\cap\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{A}}}]}}{\mu_{Q[\mathcal{R}_{\mathbf{A}}\cap\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{A}}}]}}$ *with* $\mathcal{R}_{(\cdot)} = \mathcal{R}^{\mathbf{C}}_{(\cdot)}$.

Finally, we develop a systematic procedure named IFP (Influence Function for PAGs), given in Algo. 1, that derives a UIF for any identifiable causal effect in PAGs. IFP recursively applies Lemmas 4 and 5 until all needed C-factors are in CE-1 or CE-2 form, whose UIFs are given by Lemma 1 and 2, respectively, initially equipped with a UIF for $P(\mathbf{v})$, $\mathcal{V}_{Q[\mathbf{V}]} = I_{\mathbf{v}}(\mathbf{V})$.

Theorem 1 (Completeness of IFP). *Procedure IFP (Algo. 1) derives a UIF for any identifiable $P_{\mathbf{x}}(\mathbf{y})$ in a PAG G over $\mathbf{V}$ in $O(|\mathbf{V}|^4)$ time, where $|\mathbf{V}|$ is the number of variables. IFP returns* FAIL *if $P_{\mathbf{x}}(\mathbf{y})$ is not identifiable.*

For concreteness, we demonstrate the application of IFP by deriving a UIF for $\psi = P_x(\mathbf{y})$, where $\mathbf{Y} \equiv \{Y_1, Y_2, Y_3, Y_4\}$, in the PAG in Fig. 1.

Illustration 3 (UIF for $P_x(\mathbf{y})$ in Fig. 1 by IFP). *We start with $\mathbf{D} \equiv \mathbf{Y}$ (Line 3) and $\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})} =$ DERIVEUIF$(\mathbf{D}, \mathbf{V}, P(\mathbf{V}), \mathcal{V}_{Q[\mathbf{V}]} = I_{\mathbf{v}}(\mathbf{V}))$ (Line 4).* DERIVEUIF() *reaches line 14, where $\mathbf{B}_0 \equiv \{Y_2\}$ satisfies the condition with $\mathcal{R}_{\mathbf{B}_0} = \{Y_2, Y_3\}$, $\mathcal{R}_{\mathbf{D}\setminus\mathcal{R}_{\mathbf{B}_0}} = \{Y_1, Y_4\}$, and $\mathcal{R}_{\mathbf{B}_0} \cap \mathcal{R}_{\mathbf{D}\setminus\mathcal{R}_{\mathbf{B}_0}} = \emptyset$. Then, line 15 gives (using* $\mathrm{ID}(\emptyset) = 1$ *and* $\mathrm{IF}(\emptyset) = 0$*)*

$$\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})} = \mathrm{UIF}(\mathcal{R}_{\mathbf{B}_0}) \cdot \mathrm{ID}(\mathcal{R}_{\mathbf{D}\setminus\mathcal{R}_{\mathbf{B}_0}}) + \mathrm{IF}(\mathcal{R}_{\mathbf{D}\setminus\mathcal{R}_{\mathbf{B}_0}}) \cdot \mathrm{ID}(\mathcal{R}_{\mathbf{B}_0}).$$

Next we show a sketch derivation of $\mathrm{UIF}(\mathcal{R}_{\mathbf{B}_0})$ *and* $\mathrm{UIF}(\mathcal{R}_{\mathbf{D}\setminus\mathcal{R}_{\mathbf{B}_0}})$*. Refer Appendix A for details. First,*

$$\mathrm{UIF}(\mathcal{R}_{\mathbf{B}_0}) = \textsc{DeriveUIF}(\mathcal{R}_{\mathbf{B}_0}, \mathbf{V}, P(\mathbf{V}), I_{\mathbf{v}}(\mathbf{V})).$$

$\mathrm{UIF}(\mathcal{R}_{\mathbf{B}_0})$ *is derived by repeating Lines 8, 9, 10, and 13 as follows: Starting with $\mathbf{B} = Y_4$ at Line 8, let $\mathbf{T} = \mathbf{V} \setminus \mathbf{B} = \{Y_1, R, X, Y_2, Y_3\}$, compute $Q[\mathbf{T}]$ (Line 9) and $\mathcal{V}_{Q[\mathbf{T}]}$ (Line 10), call* DERIVEUIF$(\mathcal{R}_{\mathbf{B}_0}, \mathbf{T}, Q[\mathbf{T}], \mathcal{V}_{Q[\mathbf{T}]})$ *(Line 13). Then repeat the above by calling* DERIVEUIF$(\mathcal{R}_{\mathbf{B}_0}, \mathbf{T}, Q[\mathbf{T}], \mathcal{V}_{Q[\mathbf{T}]})$ *three more times with $\mathbf{B} = Y_1$ at line 8, $\mathbf{T} = \{R, X, Y_2, Y_3\}$; $\mathbf{B} = X$ at line 8, $\mathbf{T} = \{R, Y_2, Y_3\}$; and $\mathbf{B} = R$ at line 8, $\mathbf{T} = \{Y_2, Y_3\}$. Finally we obtain $Q[\mathcal{R}_{\mathbf{B}_0}] = Q[Y_2, Y_3] = \sum_r P(y_2, y_3|x, r)P(r)$, and* $\mathrm{UIF}(\mathcal{R}_{\mathbf{B}_0}) = \mathcal{V}_{Q[\mathcal{R}_{\mathbf{B}_0}]}$ *is given by Lemma 1 as*

$$\mathrm{UIF}(\mathcal{R}_{\mathbf{B}_0}) = \theta^a_{0,1} + \omega^a_1(\theta^a_{1,1} - \theta^a_{2,1})$$

where $\omega^a_1 = \frac{I_x(X)}{P(X|R)}$; and for $P_{\pi^a} = I_x(X)P(Y_2, Y_3|X, R)P(R)$, $\theta^a_{0,1} = \mathbb{E}_{P_{\pi^a}}[I^a|R]$, $\theta^a_{1,1} = I^a$, and $\theta^a_{1,2} = \mathbb{E}_{P_{\pi^a}}[I^a|X, R]$ where $I^a \equiv I_{y_2,y_3}(Y_2, Y_3)$.

Next, with a similar matter,

$$\begin{aligned}\mathrm{UIF}(\mathcal{R}_{\mathbf{D}\setminus\mathcal{R}_{\mathbf{B}_0}}) &= \textsc{DeriveUIF}(\mathcal{R}_{\mathbf{D}\setminus\mathcal{R}_{\mathbf{B}_0}}, \mathbf{V}, P(\mathbf{V}), I_{\mathbf{v}}(\mathbf{V})) \\ &= \theta^b_{0,1} + \omega^b_1(\theta^b_{1,1} - \theta^b_{2,1}),\end{aligned}$$

where $\omega^b_1 = \frac{I_{r,x,y_2,y_3}(R,X,Y_2,Y_3)}{P(R,X,Y_2,Y_3|Y_1)}$; and for $P_{\pi^b} = I_{r,x,y_2,y_3}(R, X, Y_2, Y_3)P(Y_4|pre(Y_4))P(Y_1)$, $\theta^b_{0,1} = \mathbb{E}_{P_{\pi^b}}[I^b|Y_1]$, $\theta^a_{1,1} = I^b$, and $\theta^b_{1,2} = \mathbb{E}_{P_{\pi^b}}[I^b|pre(Y_4)]$ where $I^b \equiv I_{y_1,y_4}(Y_1, Y_4)$.

For reference, $P_x(\mathbf{y})$ is identified as

$$P_x(\mathbf{y}) = Q[\mathbf{Y}] = Q[Y_2, Y_3]Q[Y_1, Y_4], \tag{12}$$

where $Q[Y_2, Y_3] = \sum_r P(y_2, y_3|x, r)P(r)$ and $Q[Y_1, Y_4] = P(y_4|pre(y_4))P(y_1)$.

5. DML Estimators

In this section, we construct DML estimators for causal effects $P_{\mathbf{x}}(\mathbf{y})$ from finite samples $\mathcal{D} = \{\mathbf{V}_{(i)}\}_{i=1}^N$ based on the UIF $\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})}(\mathbf{V}; \eta)$ derived by IFP algorithm. The resulting DML estimators have nice properties of debiasedness, as well as doubly robustness in the sense that an estimator T_N composed of the nuisances $\eta = (\eta_0, \eta_1)$ is said to be *doubly robust* if T_N is consistent whenever either η_0 or η_1 are consistent.

First we show that IFs derived by IFP are a Neyman orthogonal score, which is needed for the DML method.

Proposition 3. *Let $P_{\mathbf{x}}(\mathbf{y})$ be identified as $P_{\mathbf{x}}(\mathbf{y}) = \psi \equiv \Psi(P)$. Then, the IF $\phi_{P_{\mathbf{x}}(\mathbf{y})} = \mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})} - \mathbb{E}_P[\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})}]$, where $\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})}$ is derived by Algo. 1 IFP, is a Neyman orthogonal score for ψ.*

A DML estimator for $P_{\mathbf{x}}(\mathbf{y})$, named *DML-IDP* (DML estimator for IDentifiable causal effects in PAGs), is constructed according to (Chernozhukov et al., 2018) as follows:

Definition 5 (Double/Debiased Machine Learning estimator for identifiable causal effects (DML-IDP)). Let $\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})}(\mathbf{V}; \eta)$ be the UIF given by Algo. 1 IFP for the target functional $\psi = P_{\mathbf{x}}(\mathbf{y})$. Let $\mathcal{D} = \{\mathbf{V}_{(i)}\}_{i=1}^N$ denote samples drawn from $P(\mathbf{v})$. Then, the DML-IDP estimator T_N for $\psi = P_{\mathbf{x}}(\mathbf{y})$ is constructed as follows:
(1) Split $\mathcal{D}$ randomly into two halves: $\mathcal{D}_0$ and $\mathcal{D}_1$;
(2) For $p \in \{0, 1\}$, use $\mathcal{D}_p$ to construct models for $\widehat{\eta}_p$, the nuisance functions estimated from samples $\mathcal{D}_p$; and
(3) $T_N \equiv \sum_{p\in\{0,1\}} \frac{2}{N} \sum_{\mathbf{V}_{(i)}\in\mathcal{D}_p} \mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})}(\mathbf{V}_{(i)}, \widehat{\eta}_{1-p})$.

To witness the robustness properties of DML-IDP, we first note that the nuisances in $\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})}(\mathbf{V}; \eta)$ returned by IFP consist of the nuisances of UIFs for CE-1:

Lemma 6 (Nuisances of UIFs). *The UIF $\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})}(\mathbf{V}; \eta)$ returned by Algo. 1 IFP is an arithmetic combination (ratio, multiplication, and marginalization) of UIFs for functionals in the form of CE-1, denoted as $\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})}(\mathbf{V}; \eta = \{\boldsymbol{\omega}_j, \boldsymbol{\theta}_j\}_{j=1}^{\ell}) = \mathcal{A}\left(\{\mathcal{V}_j(\boldsymbol{\omega}_j, \boldsymbol{\theta}_j)\}_{j=1}^{\ell}\right)$ where $\mathcal{V}_j(\boldsymbol{\omega}_j, \boldsymbol{\theta}_j)$ denotes a UIF given by Lemma 1 with $\boldsymbol{\omega}_j = \{\omega_{j,k}\}_{k=1}^{m_j}$ and $\boldsymbol{\theta}_j = \{\theta_{j,0,1}\} \cup \{\theta_{j,k,1}, \theta_{j,k,2}\}_{k=1}^{m_j}$ being nuisances for $\mathcal{V}_j$, and $\mathcal{A}(\cdot)$ an arithmetic function.*

For example, the UIF for $P_x(\mathbf{y})$ in Fig. 2b given by Eq. (8)

is a function of UIFs $\mathcal{V}_{\mathbf{V}\setminus\mathbf{S_x}}$ and $\mathcal{V}_{\mathbf{S_x}\setminus\mathbf{x}}$ both of which are given by Lemma 1 as shown in Illustration 2.

We show that DML-IDP estimators attain debiasedness and doubly robustness, the main result of this section:

Theorem 2 (**Properties of DML-IDP**). *Let T_N be the DML-IDP estimator of $P_{\mathbf{x}}(\mathbf{y})$ defined in Def. 5 constructed based on the UIF $\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})}(\mathbf{V};\eta = \{\boldsymbol{\omega}_j, \boldsymbol{\theta}_j\}_{j=1}^{\ell})$ where $\boldsymbol{\omega}_j = \{\omega_{j,k}\}_{k=1}^{m_j}$ and $\boldsymbol{\theta}_j = \{\theta_{j,0,1}\} \cup \{\theta_{j,k,1}, \theta_{j,k,2}\}_{k=1}^{m_j}$ are nuisances as specified in Lemma 6. Suppose T_N is bounded from above by some constant $C \in \mathbb{R}$; i.e., $T_N < C < \infty$. Then,*

1. ***Debiasedness**: T_N is $\sqrt{N}$-consistent and asymptotically normal if estimates for all nuisances converge to the true nuisances at least at rate $o_P(N^{-1/4})$.*

2. ***Doubly robustness**: T_N is consistent if, for every $j = 1, \cdots, \ell$ and $k = 1, \cdots, m_j$, either estimates $\widehat{\omega}_{j,k}$ or $(\widehat{\theta}_{j,k-1,1}, \widehat{\theta}_{j,k,2})$ converge to the true nuisances at rate $o_P(1)$.*

By Thm. 2, DML-IDP estimators attain root-N consistency even when nuisances converge much slower at fourth-root-N rate or when some nuisances are misspecified. These properties allow one to employ flexible ML models (e.g., neural nets) that do not meet certain complexity restrictions (e.g., Donsker condition) for estimating nuisances in estimating causal effects (Klaassen, 1987; Robins & Ritov, 1997; Robins et al., 2008; Zheng & van der Laan, 2011; Chernozhukov et al., 2018). In contrast, plug-in estimators may fail to achieve $\sqrt{N}$-consistency if estimates for nuisances converges at $o_P(N^{-1/4})$ and are vulnerable to model misspecification.

For concreteness, we compare DML-IDP with plug-in estimators in the following examples (Refer to Appendix A for detailed derivations).

Illustration 4 (**DML-IDP vs. Plug-in (PI) estimators for $P_{\mathbf{x}}(\mathbf{y})$ in Fig. (2a,2b,1)**). *By Thm. 2, DML-IDP estimator for $P_{x_1,x_2}(y)$ in **Fig. (2a)** is consistent if estimates for either the following converges:*

$$\{P(v_i|pre(v_i))\}_{V_i\in\{X_1,X_2\}} \vee \{P(v_i|pre(v_i))\}_{V_i\in\{X_1,Y\}} \vee \{P(v_i|pre(v_i))\}_{V_i\in\{Z,Y\}},$$

while PI using Eq. (2) is consistent if estimates for $\{P(y|pre(y)), P(z|pre(z)), P(a|b,c), P(b|c), P(c)\}$ converge, where the variables are ordered as $\mathbf{V} = \{C \prec B \prec A \prec X_1 \prec Z \prec X_2 \prec Y\}$.

*DML-IDP estimator for $P_x(y_1, y_2, y_3, y_4, y_5)$ in **Fig. (2b)** is consistent if estimates*

$$\{P(v_i|pre(v_i)\}_{V_i\in\{X,Y_1,Y_3,Y_4\}} \vee \{P(v_i|pre(v_i)\}_{V_i\in\{X,Y_1,Y_5\}} \vee \{P(v_i|pre(v_i)\}_{V_i\in\{Y_2,Y_5\}},$$

and

$$\{P(v_i|pre(v_i))\}_{V_i\in\{C,Y_2\}} \vee \{P(v_i|pre(v_i))\}_{V_i\in\{C,Y_3,Y_4\}} \vee \{P(v_i|pre(v_i)\}_{V_i\in\{X,Y_1,Y_3,Y_4\}}$$

converge, while PI using Eq. (6) is consistent if estimates for $\{P(v_i|pre(v_i))\}_{V_i\in\mathbf{V}}$ converge, where the order over $\mathbf{V}$ is $C \prec X \prec Y_1 \prec Y_2 \prec Y_3 \prec Y_4 \prec Y_5$.

*DML-IDP for $P_x(y_1, y_2, y_3, y_4)$ in **Fig. (1)** is consistent if estimates for*

$$\{P(x|r)\} \vee \{P(y_2|x,r), P(y_3|y_2,x,r)\},$$

and

$$\{P(v_i|pre(v_i))\}_{V_i\in\{R,X,Y_2,Y_3\}} \vee \{P(y_4|pre(y_4))\}$$

converge, while PI using Eq. (12) is consistent if estimates for

$$\{P(y_2|x,r), P(y_3|y_2,x,r), P(r), P(y_4|pre(y_4)), P(y_1)\}$$

converge, where the order over $\mathbf{V}$ is $Y_1 \prec R \prec X \prec Y_2 \prec Y_3 \prec Y_4$.

6. Experiments

6.1. Experiments Setup

We evaluate DML-IDP for estimating $P_{\mathbf{x}}(\mathbf{y})$ in Fig. (2a,2b,1). We specify an SCM M for each PAG and generate datasets $\mathcal{D}$ from M. Details of the models and the data generating process are described in Appendix C. Throughout the experiments, the target causal effect is $\mu(\mathbf{x}) \equiv P_{\mathbf{x}}(\mathbf{Y} = 1)$, with ground-truth pre-computed. We compare DML-IDP with plug-in estimator (PI), the only available general-purpose estimator working for arbitrary causal functionals. Nuisance functions are estimated using standard techniques available in the literature (refer to Appendix C for details), e.g., conditional probabilities are estimated using a gradient boosting model XGBoost (Chen & Guestrin, 2016), which is known to be flexible.

Accuracy Measure Given a data set $\mathcal{D}$ with N samples, let $\widehat{\mu}_{\text{DML}}(\mathbf{x})$ and $\widehat{\mu}_{\text{PI}}(\mathbf{x})$ be the estimated $P_{\mathbf{x}}(\mathbf{Y} = 1)$ using DML-IDP and PI estimators. For each $\widehat{\mu} \in \{\widehat{\mu}_{\text{DML}}(\mathbf{x}), \widehat{\mu}_{\text{PI}}(\mathbf{x})\}$, we compute the average absolute error (AAE) as $|\mu(\mathbf{x}) - \hat{\mu}(\mathbf{x})|$ averaged over $\mathbf{x}$. We generate 100 datasets for each sample size N. We call the mean of the 100 AAEs the *mean average absolute error*, or MAAE, and its plot vs. the sample size N, the *MAAE plot*.

Simulation Strategy To show debiasedness ('DB') property, we add a 'converging noise' ϵ, decaying at a $N^{-\alpha}$ rate (i.e., $\epsilon \sim \text{Normal}(N^{-\alpha}, N^{-2\alpha})$) for $\alpha = 1/4$, to the estimated nuisance values to control the convergence rate of the estimators for nuisances, following the technique in

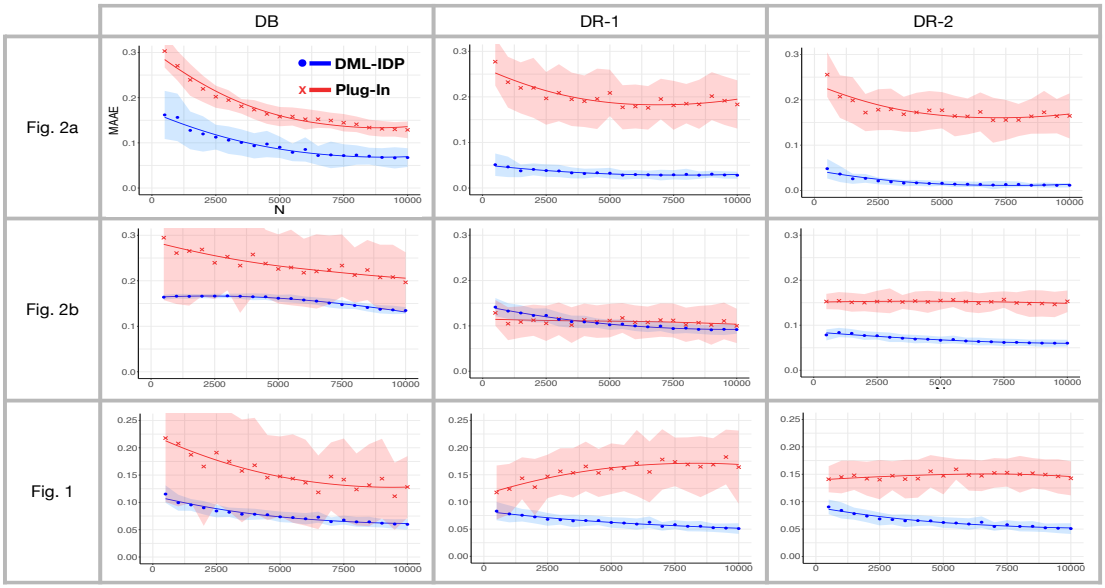


Figure 3: MAAE Plots for **(Top)** Fig. 2a, **(Middle)** Fig. 2b, and **(Bottom)** Fig. 1, under scenarios 'Debiasedness' ('DB') and 'Doubly Robustness' ('DR-1' and 'DR-2'). The solid lines represent MAAEs and shades represent one standard deviation.

(Kennedy, 2020b). We simulate a misspecified model for nuisance functions of the form $P(v_i|\cdot)$ by replacing samples for V_i with randomly generated samples V_i', training the model $\widehat{P}(v_i'|\cdot)$, and using this misspecified nuisance in computing the target functional, following (Kang et al., 2007).

6.2. Experimental Results

Debiasedness (DB) The MAAE plots for the debiasedness experiments for Fig. (2a,2b,1) are shown in the first column of Fig. 3. DML-IDP shows the debiasedness property against the converging noise decaying at $N^{-1/4}$ rates, while PI converges much slower for all three examples.

Doubly robustness (DR) The MAAE plots for the doubly robustness experiments are shown in the 2nd and 3rd columns of Fig. 3. Two misspecification scenarios are simulated for each example based on the results in Illustration 4. For Fig. 2a, nuisances $\{\widehat{P}(v_i|\text{pre}(v_i))\}$ for $V_i \in \{Y, Z\}$ in 'DR-1' and for $V_i \in \{Z, X_2\}$ in 'DR-2' are misspecified. For Fig. 2b, nuisances $\{\widehat{P}(v_i|\text{pre}(v_i))\}$ for $V_i \in \{Y_2, Y_5\}$ in 'DR-1' and for $V_i \in \{X, Y_1, Y_3, Y_4\}$ in 'DR-2' are misspecified. For Fig. 1, nuisances $\widehat{P}(y_4|\text{pre}(y_4))$ in 'DR-1' and $\{\widehat{P}(y_2|x, r), \widehat{P}(y_3|y_2, x, r)\}$ in 'DR-2' are misspecified. The results in all the scenarios support the doubly robustness of DML-IDP, whereas PI may fail to converge when misspecification is present.

7. Conclusions

We derived influence functions (Algo. 1, Thm. 1) and developed DML estimators named DML-IDP (Def. 5) for any causal effects identifiable given a Markov equivalence class of causal graphs represented as a PAG. DML-IDP estimators are guaranteed to have the property of debiasedness and doubly robustness (Thm. 2). Our experimental results demonstrate that these estimators are significantly more robust against model misspecification and slow convergence rate in learning nuisances compared to the only alternative estimator available in the literature, a plug-in estimator. We hope the new machinery developed here will allow more reliable and robust causal effect estimates by integrating modern ML methods that are capable of handling complex, high-dimensional data with causal learning and identification theory, paving the way towards a robust, data-driven, and end-to-end solution to causal effect estimation.

Acknowledgements

We thank the reviewers for their feedback and help to improve this manuscript. Elias Bareinboim and Yonghan Jung were partially supported by grants from NSF IIS-1750807 (CAREER). Jin Tian was partially supported by ONR grant N000141712140.

References

Bang, H. and Robins, J. M. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61 (4):962–973, 2005.

Bareinboim, E. and Pearl, J. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352, 2016.

Bengio, Y. *Learning deep architectures for AI*. Now Publishers Inc, 2009.

Benkeser, D., Carone, M., Laan, M. V. D., and Gilbert, P. Doubly robust nonparametric inference on the average treatment effect. *Biometrika*, 104(4):863–880, 2017.

Bhattacharya, R., Nabi, R., and Shpitser, I. Semiparametric inference for causal effects in graphical models with hidden variables. *arXiv preprint arXiv:2003.12659*, 2020.

Breiman, L. Random forests. *Machine learning*, 45(1): 5–32, 2001.

Casella, G. and Berger, R. L. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

Chen, T. and Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. Double/debiased machine learning for treatment and structural parameters: Double/debiased machine learning. *The Econometrics Journal*, 21(1), 2018.

Chernozhukov, V., Demirer, M., Lewis, G., and Syrgkanis, V. Semi-parametric efficient policy learning with continuous actions. In *Advances in Neural Information Processing Systems*, pp. 15065–15075, 2019.

Colangelo, K. and Lee, Y.-Y. Double debiased machine learning nonparametric inference with continuous treatments. *arXiv preprint arXiv:2004.03036*, 2020.

Díaz, I. and van der Laan, M. J. Targeted data adaptive estimation of the causal dose–response curve. *Journal of Causal Inference*, 1(2):171–192, 2013.

Freund, Y., Schapire, R. E., et al. Experiments with a new boosting algorithm. In *icml*, volume 96, pp. 148–156. Citeseer, 1996.

Fulcher, I. R., Shpitser, I., Marealle, S., and Tchetgen Tchetgen, E. J. Robust inference on population indirect causal effects: the generalized front door criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(1):199–214, 2020.

Hill, J. L. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240, 2011.

Huang, Y. and Valtorta, M. Pearl's calculus of intervention is complete. In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*, pp. 217–224. AUAI Press, 2006.

Jaber, A., Zhang, J., and Bareinboim, E. Causal identification under markov equivalence. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence*, 2018a.

Jaber, A., Zhang, J., and Bareinboim, E. A graphical criterion for effect identification in equivalence classes of causal diagrams. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018b.

Jaber, A., Zhang, J., and Bareinboim, E. Causal identification under markov equivalence: Completeness results. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 2981–2989, 2019.

Jung, Y., Tian, J., and Bareinboim, E. Estimating causal effects using weighting-based estimators. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, 2020a.

Jung, Y., Tian, J., and Bareinboim, E. Learning causal effects via weighted empirical risk minimization. *Proceedings of the 34th Annual Conference on Neural Information Processing Systems*, 2020b.

Jung, Y., Tian, J., and Bareinboim, E. Estimating identifiable causal effects through double machine learning. *Proceedings of the 35th AAAI Conference on Artificial Intelligence*, 2021.

Kang, J. D., Schafer, J. L., et al. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical science*, 22(4):523–539, 2007.

Kennedy, E. H. Efficient nonparametric causal inference with missing exposure information. *The international journal of biostatistics*, 16(1), 2020a.

Kennedy, E. H. Optimal doubly robust estimation of heterogeneous causal effects. *arXiv preprint arXiv:2004.14497*, 2020b.

Kennedy, E. H., Ma, Z., McHugh, M. D., and Small, D. S. Nonparametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 79(4): 1229, 2017.

Klaassen, C. A. Consistent estimation of the influence function of locally asymptotically linear estimators. *The Annals of Statistics*, pp. 1548–1562, 1987.

Lee, S. and Bareinboim, E. Causal effect identifiability under partial-observability. In *Proceedings of the 37th International Conference on Machine Learning*, 2020.

Molina, J., Rotnitzky, A., Sued, M., and Robins, J. Multiple robustness in factorized likelihood models. *Biometrika*, 104(3):561–581, 2017.

Neugebauer, R. and van der Laan, M. Nonparametric causal effects based on marginal structural models. *Journal of Statistical Planning and Inference*, 137(2):419–434, 2007.

Pearl, J. Causal diagrams for empirical research. *Biometrika*, 82(4):669–710, 1995.

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

Pearl, J. and Mackenzie, D. *The book of why: the new science of cause and effect*. Basic Books, 2018.

Pearl, J. and Robins, J. Probabilistic evaluation of sequential plans from causal models with hidden variables. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, pp. 444–453. Morgan Kaufmann Publishers Inc., 1995.

Perkovic, E., Textor, J., Kalisch, M., and Maathuis, M. H. Complete graphical characterization and construction of adjustment sets in markov equivalence classes of ancestral graphs. *The Journal of Machine Learning Research*, 18 (1):8132–8193, 2017.

Peters, J., Janzing, D., and Schölkopf, B. *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017.

Robins, J., Li, L., Tchetgen, E., van der Vaart, A., et al. Higher order influence functions and minimax estimation of nonlinear functionals. In *Probability and statistics: essays in honor of David A. Freedman*, pp. 335–421. Institute of Mathematical Statistics, 2008.

Robins, J. M. Marginal structural models versus structural nested models as tools for causal inference. In *Statistical models in epidemiology, the environment, and clinical trials*, pp. 95–133. Springer, 2000.

Robins, J. M. and Ritov, Y. Toward a curse of dimensionality appropriate (coda) asymptotic theory for semi-parametric models. *Statistics in medicine*, 16(3):285–319, 1997.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866, 1994.

Robins, J. M., Hernan, M. A., and Brumback, B. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5), 2000.

Rosenbaum, P. R. and Rubin, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

Rotnitzky, A. and Smucler, E. Efficient adjustment sets for population average causal treatment effect estimation in graphical models. *Journal of Machine Learning Research*, 21(188):1–86, 2020.

Rotnitzky, A., Robins, J., and Babino, L. On the multiply robust estimation of the mean of the g-functional. *arXiv preprint arXiv:1705.08582*, 2017.

Rudolph, K. E. and van der Laan, M. J. Robust estimation of encouragement-design intervention effects transported across sites. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 79(5):1509, 2017.

Shpitser, I. and Pearl, J. Identification of joint interventional distributions in recursive semi-markovian causal models. In *Proceedings of the 21st AAAI Conference on Artificial Intelligence*, pp. 1219, 2006.

Smucler, E., Sapienza, F., and Rotnitzky, A. Efficient adjustment sets in causal graphical models with hidden variables. *arXiv preprint arXiv:2004.10521*, 2020.

Spirtes, P., Glymour, C. N., Scheines, R., and Heckerman, D. *Causation, prediction, and search*. MIT press, 2000.

Tian, J. and Pearl, J. A general identification condition for causal effects. In *Proceedings of the 18th National Conference on Artificial Intelligence*, pp. 567–573, 2002.

Tian, J. and Pearl, J. On the identification of causal effects. Technical Report R-290-L, 2003.

Toth, B. and van der Laan, M. Tmle for marginal structural models based on an instrument. uc berkeley division of biostatistics working paper series. Technical report, working paper 350, 2016.

Van der Laan, M. J. and Rose, S. *Targeted learning: causal inference for observational and experimental data*. Springer Science & Business Media, 2011.

Van Der Laan, M. J. and Rubin, D. Targeted maximum likelihood learning. *The International Journal of Biostatistics*, 2(1), 2006.

Van der Vaart, A. W. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

Zhang, J. *Causal inference and reasoning in causally insufficient systems*. PhD thesis, Citeseer, 2006.

Zhang, J. A characterization of markov equivalence classes for directed acyclic graphs with latent variables. In *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence*, pp. 450–457, 2007.

Zhang, J. Causal reasoning with ancestral graphs. *Journal of Machine Learning Research*, 9(Jul):1437–1474, 2008a.

Zhang, J. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16-17): 1873–1896, 2008b.

Zheng, W. and van der Laan, M. J. Cross-validated targeted minimum-loss-based estimation. In *Targeted Learning*, pp. 459–474. Springer, 2011.

Appendix – Estimating Identifiable Causal Effects on Markov Equivalence Class through Double Machine Learning

A. Details

A.1. Background Results

Partial topological order (PTO) is a useful notion in PAGs which specifies a topological order over buckets and defines a partial order over the variables that is valid in all the causal graphs in the MEC represented by the PAG. An algorithm for assigning a valid PTO in a PAG has been developed in (Jaber et al., 2018a), presented in the following as Algo. A.1 for convenience.

Algorithm A.1 PTO($G(\mathbf{V})$) (Jaber et al., 2018a)

Input: A PAG G over $\mathbf{V}$,
Output: Partial topological order (PTO) over $\mathbf{V}$ in G.
Create a singleton bucket $\mathbf{B}_i$ such that $\mathbf{B}_i = V_i \in \mathbf{V}$.
Merge buckets $\mathbf{B}_i$ and $\mathbf{B}_j$ if there is a circle edge between them; i.e., $\mathbf{B}_i \ni X \circ\!\!-\!\!\circ Y \in \mathbf{B}_j$.
while A set of buckets $\mathbf{B}$ is not empty **do**
    Extract $\mathbf{B}_i$ with only arrowheads incident on it; and
    Remove edges between $\mathbf{B}_i$ and other buckets.
end while
Assign a partial order $\mathbf{B}_1 \prec \cdots \mathbf{B}_m$ in reverse order of bucket extraction; i.e., $\mathbf{B}_1$ is the last extracted bucket.

A complete algorithm for identifying causal effects in PAGs called IDP has been developed in (Jaber et al., 2019), presented in the following for convenience.

Algorithm A.2 IDP($\mathbf{x}, \mathbf{y}, G(\mathbf{V}), P$) (Jaber et al., 2019)

Input: Two disjoint sets $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$; A PAG G over $\mathbf{V}$; A distribution $P(\mathbf{v})$,
Output: Expression for $P_{\mathbf{x}}(\mathbf{y})$ or FAIL,
Let $\mathbf{D} = An(\mathbf{Y})_{G(\mathbf{V}\setminus\mathbf{X})}$.
$P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{d}\setminus\mathbf{y}}$ IDENTIFY $(\mathbf{D}, \mathbf{V}, P(\mathbf{V}))$
function IDENTIFY $(\mathbf{C}, \mathbf{T}, Q = Q[\mathbf{T}])$
    if $\mathbf{C} = \emptyset$, **then return** 1.
    if $\mathbf{C} = \mathbf{T}$, **then return** Q.
    {*In* $G(\mathbf{T})$*, let* $\mathbf{B}$ *denote a bucket,* $\mathcal{C}(\mathbf{B})$ *denote the* PC*-component of* $\mathbf{B}$*, and* $\mathcal{R}_{(\cdot)} = \mathcal{R}^{\mathbf{C}}_{(\cdot)}$ }
    if $\exists \mathbf{B} \subseteq \mathbf{T}\setminus\mathbf{C}$ s.t. $\mathcal{C}(\mathbf{B}) \cap Ch(\mathbf{B}) \subseteq \mathbf{B}$, **then**
        Compute $Q[\mathbf{T}\setminus\mathbf{B}]$ from Q via Prop. 1.
        return IDENTIFY $(\mathbf{C}, \mathbf{T}\setminus\mathbf{B}, Q[\mathbf{T}\setminus\mathbf{B}])$.
    else if $\exists \mathbf{B} \subseteq \mathbf{C}$ s.t. $\mathcal{R}_{\mathbf{B}} \neq \mathbf{C}$, **then**
        return $\frac{\text{IDENTIFY}(\mathcal{R}_{\mathbf{B}}, \mathbf{T}, Q) \cdot \text{IDENTIFY}(\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{B}}}, \mathbf{T}, Q)}{\text{IDENTIFY}(\mathcal{R}_{\mathbf{B}} \cap \mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{B}}}, \mathbf{T}, Q)}$.
    else return FAIL.
end function

A.2. Detailed Description of Illustrations

In this section, we provide a detailed description for Illustrations (2,3,4). For convenience, we restate our proposed Alg. 1 IFP for deriving UIFs of causal effect in PAGs as Algo. A.3.

Illustration 2 (UIF for $P_x(\mathbf{y})$ in Fig. 2b). *We will use Lemma 2 to derive a UIF for* $\psi \equiv P_x(\mathbf{y})$ *in Fig. 2b identified by*

Algorithm A.3 IFP$(\mathbf{x}, \mathbf{y}, G(\mathbf{V}), P)$ (Restated Alg. 1)

Input: Two disjoint sets $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$; A PAG G over $\mathbf{V}$; A distribution $P(\mathbf{v})$.
Output: Expression for UIF $\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})}$ or FAIL.
Let $\mathbf{D} = An(\mathbf{Y})_{G(\mathbf{V}\setminus\mathbf{X})}$.
$\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})} = \sum_{\mathbf{d}\setminus\mathbf{y}}$ DERIVEUIF $\left(\mathbf{D}, \mathbf{V}, P(\mathbf{V}), \mathcal{V}_{Q[\mathbf{V}]} = I_{\mathbf{v}}(\mathbf{V})\right)$
function DERIVEUIF $\left(\mathbf{C}, \mathbf{T}, Q = Q\left[\mathbf{T}\right], \mathcal{V} = \mathcal{V}_Q\right)$
  if $\mathbf{C} = \emptyset$, **then return** 1.
  if $\mathbf{C} = \mathbf{T}$, **then return** $\mathcal{V}$.
  $\{\mathbf{B}$ *denotes a bucket in* $G(\mathbf{T})$*;* $\mathcal{C}(\mathbf{B})$ *the* PC*-component of* $\mathbf{B}$ *in* $G(\mathbf{T})$*, and* $\mathcal{R}_{(\cdot)} \equiv \mathcal{R}^{\mathbf{C}}_{(\cdot)}$.$\}$
  if $\exists \mathbf{B} \subseteq \mathbf{T}\setminus\mathbf{C}$ s.t. $\mathcal{C}(\mathbf{B}) \cap Ch(\mathbf{B}) \subseteq \mathbf{B}$, **then**
    Compute $Q\left[\mathbf{T}\setminus\mathbf{B}\right]$ from Q via Lemma 3.
    if $Q\left[\mathbf{T}\setminus\mathbf{B}\right]$ is expressible as CE-1,
      then, Compute $\mathcal{V}_{Q[\mathbf{T}\setminus\mathbf{B}]}$ via Lemma 1.
    else if $Q\left[\mathbf{T}\setminus\mathbf{B}\right]$ is expressible as CE-2,
      then, Compute $\mathcal{V}_{Q[\mathbf{T}\setminus\mathbf{B}]}$ via Lemma 2.
    else, Compute $\mathcal{V}_{Q[\mathbf{T}\setminus\mathbf{B}]}$ via Lemma 4.
    return DERIVEUIF $\left(\mathbf{C}, \mathbf{T}\setminus\mathbf{B}, Q\left[\mathbf{T}\setminus\mathbf{B}\right], \mathcal{V}_{Q[\mathbf{T}\setminus\mathbf{B}]}\right)$.
  else if $\exists \mathbf{B} \subseteq \mathbf{C}$ s.t. $\mathcal{R}_{\mathbf{B}} \neq \mathbf{C}$, **then**
    return $(a) + (b) - (c)$, where
  $\{$ *Let* $\text{UIF}(\mathbf{W}) = \text{DERIVEUIF}(\mathbf{W}, \mathbf{T}, Q, \mathcal{V})$*;* $\text{IF}(\mathbf{W}) = \text{UIF}(\mathbf{W}) - \mathbb{E}_P[\text{UIF}(\mathbf{W})]$*;* $\text{ID}(\mathbf{W}) = \mathbb{E}_P[\text{UIF}(\mathbf{W})]\}$
  $(a) = \frac{\text{UIF}(\mathcal{R}_{\mathbf{B}})\cdot\text{ID}(\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{B}}})}{\text{ID}(\mathcal{R}_{\mathbf{B}}\cap\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{B}}})}$; $(b) = \frac{\text{IF}(\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{B}}})\cdot\text{ID}(\mathcal{R}_{\mathbf{B}})}{\text{ID}(\mathcal{R}_{\mathbf{B}}\cap\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{B}}})}$; $(c) = \frac{\text{ID}(\mathcal{R}_{\mathbf{B}})\cdot\text{ID}(\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{B}}})}{\text{ID}(\mathcal{R}_{\mathbf{B}}\cap\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{B}}})} \cdot \frac{\text{IF}(\mathcal{R}_{\mathbf{B}}\cap\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{B}}})}{\text{ID}(\mathcal{R}_{\mathbf{B}}\cap\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_{\mathbf{B}}})}$.
  else return FAIL.
end function

Coro. 1 as given in Eq. (6). The PAG in Fig. 2b has a PTO $\mathbf{V} = \{C \prec X \prec Y_1 \prec Y_2 \prec Y_3 \prec Y_4 \prec Y_5\}$*. By Coro. 1,*

$$P_{\mathbf{x}}(\mathbf{y}) = \mathcal{Q}_{\mathbf{V}\setminus\mathbf{S}_{\mathbf{X}}} \times \mathcal{Q}_{\mathbf{S}_{\mathbf{X}}\setminus\mathbf{X}},$$

where, for $\mathbf{S}_{\mathbf{X}} = \{X, Y_1, Y_4, Y_3\}$*,* $\mathbf{V}\setminus\mathbf{S}_{\mathbf{X}} = \{C, Y_2, Y_5\}$*,* $\mathcal{Q}_{\mathbf{V}\setminus\mathbf{S}_{\mathbf{X}}} = P(y_5|pre(y_5))P(y_2|pre(y_2))P(c)$ *and* $\mathcal{Q}_{\mathbf{S}_{\mathbf{X}}\setminus\mathbf{X}} = \sum_{x'} P(y_3, y_4|y_1, y_2, x', c)P(y_1, x'|c)$*. Then, a UIF for* $P_x(\mathbf{y})$ *in Eq. (6) is given by Lemma 2 as*

$$\mathcal{V}_{P_x(\mathbf{y})} = \sum_{c} \left(\mathcal{V}_{\mathbf{V}\setminus\mathbf{S}_{\mathbf{X}}}\mu_{\mathbf{S}_{\mathbf{X}}\setminus\mathbf{X}} + (\mathcal{V}_{\mathbf{S}_{\mathbf{X}}\setminus\mathbf{X}} - \mu_{\mathbf{S}_{\mathbf{X}}\setminus\mathbf{X}})\mu_{\mathbf{V}\setminus\mathbf{S}_{\mathbf{X}}}\right), \tag{A.1}$$

where $\mathcal{V}_{\mathbf{V}\setminus\mathbf{S}_{\mathbf{X}}}$*,* $\mathcal{V}_{\mathbf{S}_{\mathbf{X}}\setminus\mathbf{X}}$ *are UIFs for* $\mathcal{Q}_{\mathbf{V}\setminus\mathbf{S}_{\mathbf{X}}}$ *and* $\mathcal{Q}_{\mathbf{S}_{\mathbf{X}}\setminus\mathbf{X}}$*.* $\mathcal{Q}_{\mathbf{V}\setminus\mathbf{S}_{\mathbf{X}}}$ *is a CE-1, and hence,* $\mathcal{V}_{\mathbf{V}\setminus\mathbf{S}_{\mathbf{X}}}$ *is given by Lemma 1 as follows.*

Let $\mathbf{T} = \{\mathbf{B}_1 \prec \cdots \prec \mathbf{B}_7\} = \{C \prec X \prec Y_1 \prec Y_2 \prec Y_3 \prec Y_4 \prec Y_5\}$*. That is,* $\mathbf{B}_1 = C, \mathbf{B}_2 = X, \cdots, \mathbf{B}_7 = Y_5$*. Let* $\mathbf{C} = \{\mathbf{B}_1 \prec \mathbf{B}_4 \prec \mathbf{B}_7\}$*,* $\mathbf{X} = \{\mathbf{B}_2 \prec \mathbf{B}_3 \prec \mathbf{B}_5 \prec \mathbf{B}_6\}$*. Then,* $\mathbf{C}_0 = \{\mathbf{B}_1\} = \{C\}$*,* $\mathbf{C}_1 = \emptyset$*,* $\mathbf{C}_2 = \{\mathbf{B}_4\} = \{Y_2\}$*,* $\mathbf{C}_3 = \emptyset$*,* $\mathbf{C}_4 = \{\mathbf{B}_7\} = \{Y_5\}$*. Then,*

$$\mathcal{V}_{\mathbf{V}\setminus\mathbf{S}_{\mathbf{X}}} = \theta^a_{0,1} + \sum_{\substack{k=1 \\ \mathbf{C}_k\neq\emptyset}}^{4} \omega^a_k(\theta^a_{k,1} - \theta^a_{k,2}) = \theta^a_{0,1} + \omega^a_2(\theta^a_{2,1} - \theta^a_{2,2}) + \omega^a_4(\theta^a_{4,1} - \theta^a_{4,2}),$$

where, for $P_{\pi^a} \equiv I_{x,y_1,y_3,y_4}(X, Y_1, Y_3, Y_4)P(C)P(Y_2|pre(Y_2))P(Y_5|pre(Y_5))$ *and* $I^a \equiv I_{c,y_2,y_5}(C, Y_2, Y_5)$*,* $\theta^a_{0,1} = \mathbb{E}_{P_{\pi^a}}[I^a|\mathbf{B}_{0_{\max}}, pre_{\mathbf{T}}(\mathbf{B}_{0_{\max}})] = \mathbb{E}_{P_{\pi^a}}[I^a|C] = I_c(C)P(y_5|x, y_1, y_2, y_3, y_4, C)$*,* $\theta^a_{2,1} = \mathbb{E}_{P_{\pi^a}}[I^a|\mathbf{B}_{2_{\max}}, pre_{\mathbf{T}}(\mathbf{B}_{2_{\max}})] = \mathbb{E}_{P_{\pi^a}}[I^a|Y_2, pre(Y_2)] = I_{c,y_2}(C, Y_2)P(y_5|y_3, y_4, pre(Y_3))$*,* $\theta^a_{2,2} = \mathbb{E}_{P_{\pi^a}}[I^a|pre_{\mathbf{T}}(\mathbf{B}_{2_{\min}})] = \mathbb{E}_{P_{\pi^a}}[I^a|pre(Y_2)] = I_c(C)P(y_5|y_2, y_3, y_4, pre(Y_2))$*,* $\theta^a_{4,1} = \mathbb{E}_{P_{\pi^a}}[I^a|\mathbf{B}_{4_{\max}}, pre_{\mathbf{T}}(\mathbf{B}_{4_{\max}})] = \mathbb{E}_{P_{\pi^a}}[I^a|Y_5, pre(Y_5)] = I^a$*,* $\theta^a_{4,2} = \mathbb{E}_{P_{\pi^a}}[I^a|pre_{\mathbf{T}}(\mathbf{B}_{4_{\min}})] = \mathbb{E}_{P_{\pi^a}}[I^a|pre(Y_5)] = I_{c,y_2}(C, Y_2)P(y_5|pre(Y_5))$*. Also,* $\omega^a_2 = \frac{I_{x,y_1}(X,Y_1)}{P(X|C)P(Y_1|X,C)}$ *and* $\omega^a_4 = \omega^a_2 \times \frac{I_{y_3,y_4}(Y_3,Y_4)}{P(Y_3|pre(Y_3))P(Y_4|pre(Y_4))}$*. Without loss of generality, we set* $\theta^a_{k/2,1} \leftarrow \theta^a_{k,1}$*,* $\theta^a_{k/2,2} \leftarrow \theta^a_{k,2}$*, and* $\omega^a_{k/2} \leftarrow \omega^a_k$ *for* $k \in \{2, 4\}$*.*

We now derive the UIF $\mathcal{V}_{\mathbf{S}_{\mathbf{X}}\setminus\mathbf{X}}$ *by Lemma 1. Let* $\mathbf{C} = \{\mathbf{B}_2 \prec \mathbf{B}_3 \prec \mathbf{B}_5 \prec \mathbf{B}_6\}$*,* $\mathbf{X} = \{\mathbf{B}_1 \prec \mathbf{B}_4\}$*. Then,* $\mathbf{C}_0 = \emptyset$*,*

$\mathbf{C}_1 = \{\mathbf{B}_2, \mathbf{B}_3\} = \{X, Y_1\}$, $\mathbf{C}_2 = \{\mathbf{B}_5, \mathbf{B}_6\} = \{Y_3, Y_4\}$. *Then,*

$$\mathcal{V}_{\mathbf{S_X}\setminus\mathbf{X}} = \theta^b_{0,1} + \omega^b_1(\theta^b_{1,1} - \theta^b_{1,2}) + \omega^b_2(\theta^b_{2,1} - \theta^b_{2,2}),$$

where, for $P_{\pi^b} \equiv I_{c,y_2}(C, Y_2)P(Y_3, Y_4|pre(Y_3))P(X, Y_1|C)$ *and* $I^b \equiv I_{y_1,y_3,y_4}(Y_1, Y_3, Y_4)$, $\theta^b_{0,1} = \mathbb{E}_{P_{\pi^b}}[I^b|\mathbf{B}_{0_{\max}}, pre_{\mathbf{T}}(\mathbf{B}_{0_{\max}})] = \mathbb{E}_{P_{\pi^b}}[I^b] = \sum_{x'} P(y_3, y_4|y_2, y_1, x', C)P(x', y_1|c)$, $\theta^b_{1,1} = \mathbb{E}_{P_{\pi^b}}[I^b|\mathbf{B}_{1_{\max}}, pre_{\mathbf{T}}(\mathbf{B}_{1_{\max}})] = \mathbb{E}_{P_{\pi^b}}[I^b|C, X, Y_1] = P(y_3, y_4|y_2, pre(Y_2))I_{y_1}(Y_1)$, $\theta^b_{1,2} = \mathbb{E}_{P_{\pi^b}}[I^b|pre_{\mathbf{T}}(\mathbf{B}_{1_{\min}})] = \mathbb{E}_{P_{\pi^b}}[I^b|C] = \sum_{x'} P(y_3, y_4|y_2, y_1, x', C)P(x, y_1|C)$, $\theta^b_{2,1} = \mathbb{E}_{P_{\pi^b}}[I^b|\mathbf{B}_{2_{\max}}, pre_{\mathbf{T}}(\mathbf{B}_{2_{\max}})] = \mathbb{E}_{P_{\pi^b}}[I^b|Y_4, pre(Y_4)] = I_{y_1,y_3,y_3}(Y_1, Y_3, Y_4)$, $\theta^b_{2,2} = \mathbb{E}_{P_{\pi^b}}[I^b|pre_{\mathbf{T}}(\mathbf{B}_{2_{\min}})] = \mathbb{E}_{P_{\pi^b}}[I^b|pre(Y_3)] = P(y_3, y_4|pre(Y_3))I_{y_1}(Y_1)$. *Also,* $\omega^b_1 = \frac{I_c(C)}{P(C)}$ *and* $\omega^b_2 = \omega^b_1 \times \frac{I_{y_2}(Y_2)}{P(Y_2|pre(Y_2))}$.

Illustration 3 (UIF for $P_x(\mathbf{y})$ in Fig. 1 by IFP). *We demonstrate the application of IFP by deriving a UIF for* $\psi = P_x(\mathbf{y})$, *where* $\mathbf{Y} \equiv \{Y_1, Y_2, Y_3, Y_4\}$, *in the PAG in Fig. 1. We assume a PTO* $\mathbf{V} = \{Y_1 \prec R \prec X \prec Y_2 \prec Y_3 \prec Y_4\}$ *in the following. For reference,* $P_x(\mathbf{y})$ *is identified as*

$$P_x(\mathbf{y}) = Q[\mathbf{Y}] = Q[Y_2, Y_3]Q[Y_1, Y_4], \tag{A.2}$$

where $Q[Y_2, Y_3] = \sum_r P(y_2, y_3|x, r)P(r)$ *and* $Q[Y_1, Y_4] = P(y_4|pre(y_4))P(y_1)$.

We start with $\mathbf{D} \equiv \mathbf{Y}$ *(Line 3) and* $\mathcal{V}_{P_{\mathbf{x}(\mathbf{y})}} = \text{DERIVEUIF}(\mathbf{D}, \mathbf{V}, P(\mathbf{V}), \mathcal{V}_{Q[\mathbf{V}]} = I_{\mathbf{v}}(\mathbf{V}))$ *(Line 4).* DERIVEUIF() *reaches line 14, where* $\mathbf{B}_0 \equiv \{Y_2\}$ *satisfies the condition with* $\mathcal{R}_{\mathbf{B}_0} = \{Y_2, Y_3\}$, $\mathcal{R}_{\mathbf{D}\setminus\mathcal{R}_{\mathbf{B}_0}} = \{Y_1, Y_4\}$, *and* $\mathcal{R}_{\mathbf{B}_0} \cap \mathcal{R}_{\mathbf{D}\setminus\mathcal{R}_{\mathbf{B}_0}} = \emptyset$. *Then, line 15 gives (using* $\text{ID}(\emptyset) = 1$ *and* $\text{IF}(\emptyset) = 0$*)*

$$\mathcal{V}_{P_{\mathbf{x}(\mathbf{y})}} = \text{UIF}(\mathcal{R}_{\mathbf{B}_0}) \cdot \text{ID}(\mathcal{R}_{\mathbf{D}\setminus\mathcal{R}_{\mathbf{B}_0}}) + \text{IF}(\mathcal{R}_{\mathbf{D}\setminus\mathcal{R}_{\mathbf{B}_0}}) \cdot \text{ID}(\mathcal{R}_{\mathbf{B}_0}).$$

Next we derive $\text{UIF}(\mathcal{R}_{\mathbf{B}_0}) = \text{DERIVEUIF}(\mathcal{R}_{\mathbf{B}_0}, \mathbf{V}, P(\mathbf{V}), I_{\mathbf{v}}(\mathbf{V}))$ *by repeating Lines 8, 9, 10, and 13 as follows. Starting with* $\mathbf{B} = Y_4$ *at line 8, let* $\mathbf{T} = \mathbf{V} \setminus \mathbf{B} = \{Y_1, R, X, Y_2, Y_3\}$, *We compute* $Q[\mathbf{T}]$ *by invoking line 9: for* $Q[\mathbf{S}_{Y_4}] = P(y_4|pre(y_4))$

$$Q[\mathbf{T}] = \frac{P(\mathbf{v})}{Q[\mathbf{S}_{Y_4}]} \cancel{\sum_{y_4} Q[\mathbf{S}_{Y_4}]} = P(\mathbf{t}),$$

which is a CE-1 according to Def. 3, with $\mathbf{C} = \mathbf{T}, \mathbf{A} = \emptyset$. *Then,* $\mathcal{V}_{Q[\mathbf{T}]}$ *can be found as* $I_{\mathbf{t}}(\mathbf{T})$ *(Line 10). We then call* $\text{DERIVEUIF}(\mathcal{R}_{\mathbf{B}_0}, \mathbf{T}, Q[\mathbf{T}], \mathcal{V}_{Q[\mathbf{T}]})$.

In the 2nd round, with $\mathbf{B} \leftarrow \{Y_1\}$ *at line 8, let* $\mathbf{T} \leftarrow \mathbf{T}\setminus\{Y_1\} = \{R, X, Y_2, Y_3\}$. *We compute* $Q[\mathbf{T}]$ *by invoking line 9. For* $Q[\mathbf{S}_{Y_1}] = P(y_1, r, x, y_2, y_3)$, *we have*

$$Q[\mathbf{T}] = \frac{P(y_1, r, x, y_2, y_3)}{P(y_1, r, x, y_2, y_3)} \sum_{y_1} P(y_1, r, x, y_2, y_3) = P(r, x, y_2, y_3),$$

which is a CE-1, with $\mathbf{C} = \{R, X, Y_2, Y_3\}$. *Then,* $\mathcal{V}_{Q[\mathbf{T}]}$ *can be found as* $I_{\mathbf{t}}(\mathbf{T})$ *(Line 10). We then call* $\text{DERIVEUIF}(\mathcal{R}_{\mathbf{B}_0}, \mathbf{T}, Q[\mathbf{T}], \mathcal{V}_{Q[\mathbf{T}]})$.

In the 3rd round, with $\mathbf{B} \leftarrow X$ *at line 8, let* $\mathbf{T} \leftarrow \{R, Y_2, Y_3\}$. *We compute* $Q[\mathbf{T}]$ *by invoking line line 9. For* $Q[\mathbf{S}_X] = P(x|r)$, *we have*

$$Q[\mathbf{T}] = \frac{P(r, x, y_2, y_3)}{P(x|r)} \cancel{\sum_{x} P(x|r)} = P(r)P(y_2, y_3|x, r),$$

which is a CE-1, with $\mathbf{C} = \{R, Y_2, Y_3\}$. *Then,* $\mathcal{V}_{Q[\mathbf{T}]}$ *can be found at line 10, as* $\mathcal{V}_{Q[\mathbf{T}]} = \theta^a_{0,1} + \omega^a_1(\theta^a_{1,1} - \theta^a_{1,2})$ *with* $P_{\pi^a} = I_x(X)P(R)P(Y_2, Y_3|X, R)$, $\omega^a_1 = \frac{I_x(X)}{P(X|R)}$, $\theta^a_{0,1} = \mathbb{E}_{P_{\pi^a}}[I_{r,y_2,y_3}(R, Y_2, Y_3)|R] = I_r(R)P(y_2, y_3|x, R)$, $\theta^a_{1,1} = I_{r,y_2,y_3}(R, Y_2, Y_3)$, *and* $\theta^a_{1,2} = \mathbb{E}_{P_{\pi^a}}[I_{r,y_2,y_3}(R, Y_2, Y_3)|X, R] = I_r(R)P(y_2, y_3|X, R)$,

In the 4th round, with $\mathbf{B} \leftarrow \{R\}$ *at line 8, let* $\mathbf{T} \leftarrow \{Y_2, Y_3\} = \mathcal{R}_{\mathbf{B}_0}$*. We compute* $Q[\mathbf{T}]$ *by invoking line line 9. For* $Q[\mathbf{S}_R] = Q[R, Y_2, Y_3] = P(r)P(y_2, y_3|x, r)$*, we have*

$$Q[\mathbf{T}] = Q[\mathcal{R}_{\mathbf{B}_0}] = \frac{P(r)P(y_2, y_3|x, r)}{P(r)P(y_2, y_3|x, r)} \sum_r P(r)P(y_2, y_3|x, r)P(r),$$

which is a CE-1, with $\mathbf{C} = \{R, Y_2, Y_3\}$ *and* $\mathbf{A} = \{R\}$*. Then,* $\mathcal{V}_{Q[\mathbf{T}]}$ *can be found at line 10 as*

$$\mathcal{V}_{Q[\mathcal{R}_{\mathbf{B}_0}]} = \theta^a_{0,1} + \omega^a_1(\theta^a_{1,1} - \theta^a_{1,2}),$$

with $P_{\pi^a} = I_x(X)P(R)P(Y_2, Y_3|X, R)$*, where* $\omega^a_1 = \frac{I_x(X)}{P(X|R)}$*,* $\theta^a_{0,1} = \mathbb{E}_{P_{\pi^a}}[I_{y_2,y_3}(Y_2, Y_3)|R] = P(y_2, y_3|x, R)$*,* $\theta^a_{1,1} = I_{y_2,y_3}(Y_2, Y_3)$*,* $\theta^a_{1,2} = \mathbb{E}_{P_{\pi^a}}[I_{y_2,y_3}(Y_2, Y_3)|X, R] = P(y_2, y_3|X, R)$*.*

Let $\mathbf{R} \equiv \mathcal{R}_{\mathbf{D}\setminus\mathcal{R}_{\mathbf{B}_0}} = \{Y_1, Y_4\}$ *for notation convenience. Next we derive* $\text{UIF}(\mathbf{R}) = \text{DERIVEUIF}(\mathbf{R}, \mathbf{V}, P(\mathbf{V}), I_{\mathbf{v}}(\mathbf{V}))$*. Starting with* $\mathbf{B} = Y_3$ *at line 8, let* $\mathbf{T} = \mathbf{V} \setminus \mathbf{B} = \{Y_1, R, X, Y_2, Y_4\}$*, We compute* $Q[\mathbf{T}]$ *by invoking line 9: for* $Q[\mathbf{S}_{Y_3}] = P(y_1, r, x, y_2, y_3)$*, we have*

$$Q[\mathbf{T}] = \frac{P(\mathbf{v})}{P(y_1, r, x, y_2, y_3)} \sum_{y_3} P(y_1, r, x, y_2, y_3) = P(y_4|pre(y_4))P(y_1, r, x, y_2),$$

which is a CE-1, with $\mathbf{C} = \{Y_1, R, X, Y_2, Y_4\}$*. Then,* $\mathcal{V}_{Q[\mathbf{T}]}$ *can be found at line 10 as* $\mathcal{V}_{Q[\mathbf{T}]} = \theta^b_{0,1} + \omega^b_1(\theta^b_{1,1} - \theta^b_{1,2})$ *with* $P_{\pi^b} = I_{y_3}(Y_3)P(Y_1, R, X, Y_2)P(Y_4|pre(Y_4))$*,* $\omega^b_1 = \frac{I_{y_3}(Y_3)}{P(Y_3|pre(Y_3))}$*,* $\theta^b_{1,1} = I_{\mathbf{t}}(\mathbf{T})$*,* $\theta^b_{1,2} = \mathbb{E}_{P_{\pi^b}}[I_{\mathbf{t}}(\mathbf{T})|Y_3, pre(Y_3)] = P(y_4|pre(Y_4))I_{\mathbf{t}\setminus\{y_4\}}(\mathbf{T}\setminus\{Y_4\})$*,* $\theta^b_{0,1} = \mathbb{E}_{P_{\pi^b}}[I_{\mathbf{t}}(\mathbf{T})|pre(Y_3)] = P(y_4|y_3, pre(Y_3))I_{\mathbf{t}\setminus\{y_4\}}(\mathbf{T}\setminus\{Y_4\})$*. We then call* $\text{DERIVEUIF}(\mathbf{R}, \mathbf{T}, Q[\mathbf{T}], \mathcal{V}_{Q[\mathbf{T}]})$*.*

In the 2nd round, with $\mathbf{B} \leftarrow \{Y_2\}$ *at line 8, let* $\mathbf{T} \leftarrow \mathbf{T}\setminus\{Y_2\} = \{Y_1, R, X, Y_4\}$*. We compute* $Q[\mathbf{T}]$ *by invoking line 9. For* $Q[\mathbf{S}_{Y_2}] = P(y_1, r, x, y_2)$*, we have*

$$Q[\mathbf{T}] = \frac{P(y_4|pre(y_4))P(y_1, r, x, y_2)}{P(y_1, r, x, y_2)} \sum_{y_2} P(y_1, r, x, y_2) = P(y_4|pre(y_4))P(y_1, r, x),$$

which is a CE-1, with $\mathbf{C} = \{Y_1, R, X, Y_4\}$*. Then,* $\mathcal{V}_{Q[\mathbf{T}]}$ *can be found at line 10 as* $\mathcal{V}_{Q[\mathbf{T}]} = \theta^b_{0,1} + \omega^b_1(\theta^b_{1,1} - \theta^b_{1,2})$ *with* $P_{\pi^b} = I_{y_2,y_3}(Y_2, Y_3)P(Y_1, R, X)P(Y_4|pre(Y_4))$*,* $\omega^b_1 = \frac{I_{y_2,y_3}(Y_2,Y_3)}{P(Y_2|pre(Y_2))P(Y_3|pre(Y_3))}$*,* $\theta^b_{1,1} = I_{\mathbf{t}}(\mathbf{T})$*,* $\theta^b_{1,2} = \mathbb{E}_{P_{\pi^b}}[I_{\mathbf{t}}(\mathbf{T})|Y_3, pre(Y_3)] = P(y_4|pre(Y_4))I_{\mathbf{t}\setminus\{y_4\}}(\mathbf{Y}\setminus\{Y_4\})$*,* $\theta^b_{0,1} = \mathbb{E}_{P_{\pi^b}}[I_{\mathbf{t}}(\mathbf{T})|pre(Y_2)] = P(y_4|y_2, y_3, pre(Y_2))I_{\mathbf{t}\setminus\{y_4\}}(\mathbf{Y}\setminus\{Y_4\})$*. We then call* $\text{DERIVEUIF}(\mathbf{R}, \mathbf{T}, Q[\mathbf{T}], \mathcal{V}_{Q[\mathbf{T}]})$*.*

In the 3rd round, with $\mathbf{B} \leftarrow \{X\}$ *at line 8, let* $\mathbf{T} \leftarrow \mathbf{T}\setminus\{X\} = \{Y_1, R, Y_4\}$*. We compute* $Q[\mathbf{T}]$ *by invoking line 9. For* $Q[\mathbf{S}_X] = P(y_1, r, x)$*, we have*

$$Q[\mathbf{T}] = \frac{P(y_4|pre(y_4))P(y_1, r, x)}{P(y_1, r, x)} \sum_x P(y_1, r, x) = P(y_4|pre(y_4))P(y_1, r),$$

which is a CE-1, with $\mathbf{C} = \{Y_1, R, Y_4\}$*. Then,* $\mathcal{V}_{Q[\mathbf{T}]}$ *can be found at line 10 as* $\mathcal{V}_{Q[\mathbf{T}]} = \theta^b_{0,1} + \omega^b_1(\theta^b_{1,1} - \theta^b_{1,2})$ *with* $P_{\pi^b} = I_{x,y_2,y_3}(X, Y_2, Y_3)P(Y_1, R)P(Y_4|pre(Y_4))$*,* $\omega^b_1 = \frac{I_{x,y_2,y_3}(X,Y_2,Y_3)}{P(X|pre(X))P(Y_2|pre(Y_2))P(Y_3|pre(Y_3))}$*,* $\theta^b_{1,1} = I_{\mathbf{t}}(\mathbf{T})$*,* $\theta^b_{1,2} = \mathbb{E}_{P_{\pi^b}}[I_{\mathbf{t}}(\mathbf{T})|Y_3, pre(Y_3)] = P(y_4|pre(Y_4))I_{\mathbf{t}\setminus\{y_4\}}(\mathbf{T}\setminus\{Y_4\})$*,* $\theta^b_{0,1} = \mathbb{E}_{P_{\pi^b}}[I_{\mathbf{t}}(\mathbf{T})|pre(X)] = P(y_4|y_2, y_3, x, pre(X))I_{\mathbf{t}\setminus\{y_4\}}(\mathbf{T}\setminus\{Y_4\})$*. We then call* $\text{DERIVEUIF}(\mathbf{R}, \mathbf{T}, Q[\mathbf{T}], \mathcal{V}_{Q[\mathbf{T}]})$*.*

In the 4th round, with $\mathbf{B} \leftarrow \{R\}$ *at line 8, let* $\mathbf{T} \leftarrow \mathbf{T}\setminus\{R\} = \{Y_1, Y_4\}$*. We compute* $Q[\mathbf{T}]$ *by invoking line 9. For* $Q[\mathbf{S}_R] = P(y_1, r)$*, we have*

$$Q[\mathbf{T}] = \frac{P(y_4|pre(y_4))P(y_1, r)}{P(y_1, r)} \sum_r P(y_1, r) = P(y_4|pre(y_4))P(y_1),$$

which is a CE-1, with $\mathbf{C} = \{Y_1, Y_4\}$*. Then,* $\mathcal{V}_{Q[\mathbf{T}]}$ *can be found at line 10 as*

$$\mathcal{V}_{Q[\mathcal{R}_{\mathbf{D}\setminus\mathcal{R}_{\mathbf{B}_0}}]} = \theta^b_{0,1} + \omega^b_1(\theta^b_{1,1} - \theta^b_{1,2}),$$

with $P_{\pi^b} = I_{r,x,y_2,y_3}(R,X,Y_2,Y_3)P(Y_1)P(Y_4|pre(Y_4))$ *and* $I^b \equiv I_{y_1,y_4}(Y_1,Y_4)$, $\omega_1^b = \frac{I_{r,x,y_2,y_3}(R,X,Y_2,Y_3)}{P(R|pre(R))P(X|pre(X))P(Y_2|pre(Y_2))P(Y_3|pre(Y_3))}$, $\theta_{0,1}^b = \mathbb{E}_{P_{\pi^b}}[I^b|Y_1] = P(y_4|y_2,y_3,x,r,pre(R))I_{y_1}(Y_1)$, $\theta_{1,1}^b = I^b$, *and* $\theta_{1,2}^b = \mathbb{E}_{P_{\pi^b}}[I^b|pre(Y_4)] = P(y_4|pre(Y_4))I_{y_1}(Y_1)$.

Illustration 4 (DML-IDP vs. Plug-in (PI) estimators for $P_{\mathbf{x}}(\mathbf{y})$ in Fig. (2a,2b,1)).
*(**Fig. 2a**). Based on Illustration 1 with PTO* $\mathbf{V} = \{C \prec B \prec A \prec X_1 \prec Z \prec X_2 \prec Y\}$, *we have*

$$\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})} = \theta_{0,1} + \omega_1(\theta_{1,1} - \theta_{1,2}) + \omega_2(\theta_{2,1} - \theta_{2,2}),$$

where $\omega_1 = \frac{I_{x_1}(X_1)}{P(X_1|pre(X_1))}$, $\omega_2 = \frac{I_{x_1,x_2}(X_1,X_2)}{P(X_1|pre(X_1))P(X_2|pre(X_2))}$; *for* $P_\pi \equiv I_{x_1,x_2}(X_1,X_2)P(A,B,C)P(Z|pre(Z))P(Y|pre(Y))$, $\theta_{0,1} = \mathbb{E}_{P_\pi}[I_y(Y)|pre(X_1)] = \sum_z P(y|x_2,z,x_1,pre(X_1))P(z|x_1,pre(X_1))$, $\theta_{1,1} = \mathbb{E}_{P_\pi}[I_y(Y)|pre(X_2)] = P(y|x_2,pre(X_2))$, $\theta_{1,2} = \mathbb{E}_{P_\pi}[I_y(Y)|pre(Z)] = \sum_z P(y|x_2,z,pre(Z))P(z|pre(Z))$; *and* $\theta_{2,1} = \mathbb{E}_{P_\pi}[I_y(Y)|\mathbf{T}] = I_y(Y)$ *and* $\theta_{2,2} = \mathbb{E}_{P_\pi}[I_y(Y)|pre(Y)] = P(y|pre(Y))$.

By Thm. 2, DML-IDP estimator for $P_{x_1,x_2}(y)$ *is consistent if estimates for either* $P(x_1|pre(x_1))$ *for* ω_1 *or* $\sum_z P(y|pre(y))P(z|pre(z))$ *for* $(\theta_{0,1},\theta_{1,2})$; *and* $P(x_1|pre(x_1))P(x_2|pre(x_2))$ *for* ω_2 *or* $P(y|pre(y))$ *for* $(\theta_{1,1},\theta_{2,2})$ *converge. This implies that DML-IDP estimator is consistent if estimates for either* $\{P(v_i|pre(v_i))\}_{V_i\in\{X_1,X_2\}}$, *or* $\{P(v_i|pre(v_i))\}_{V_i\in\{X_1,Y\}}$, *or* $\{P(v_i|pre(v_i))\}_{V_i\in\{Z,Y\}}$ *converge. In contrast, PI using Eq. (2) is consistent if estimates for* $\{P(y|pre(y)), P(z|pre(z)), P(a|b,c), P(b|c), P(c)\}$ *converge.*

*(**Fig. 2b**). Based on Illustration 2 with PTO* $C \prec X \prec Y_1 \prec Y_2 \prec Y_3 \prec Y_4 \prec Y_5$. *we have*

$$\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})} = \sum_c \left(\mathcal{V}_{\mathbf{V}\setminus\mathbf{S_X}}\mu_{\mathbf{S_X}\setminus\mathbf{X}} + (\mathcal{V}_{\mathbf{S_X}\setminus\mathbf{X}} - \mu_{\mathbf{S_X}\setminus\mathbf{X}})\mu_{\mathbf{V}\setminus\mathbf{S_X}}\right),$$

where

$$\mathcal{V}_{\mathbf{V}\setminus\mathbf{S_X}} = \theta_{0,1}^a + \omega_1^a(\theta_{1,1}^a - \theta_{1,2}^a) + \omega_2^a(\theta_{2,1}^a - \theta_{2,2}^a),$$

for $P_{\pi^a} \equiv I_{x,y_1,y_3,y_4}(X,Y_1,Y_3,Y_4)P(C)P(Y_2|pre(Y_2))P(Y_5|pre(Y_5))$ *and* $I^a \equiv I_{c,y_2,y_5}(C,Y_2,Y_5)$, $\theta_{0,1}^a = \mathbb{E}_{P_{\pi^a}}[I^a|\mathbf{B}_{0_{\max}}, pre_{\mathbf{T}}(\mathbf{B}_{0_{\max}})] = \mathbb{E}_{P_{\pi^a}}[I^a|C] = I_c(C)P(y_5|y_3,y_4,y_2,y_1,x,C)P(y_2|y_1,x,C)$, $\theta_{1,1}^a = \mathbb{E}_{P_{\pi^a}}[I^a|Y_2,pre(Y_2)] = I_c(C)P(y_5|y_3,y_4,pre(Y_3))$, $\theta_{1,2}^a = \mathbb{E}_{P_{\pi^a}}[I^a|pre(Y_2)] = I_c(C)P(y_5|y_3,y_4,y_2,pre(Y_2))P(y_2|pre(Y_2))$, $\theta_{2,1}^a = I^a$, $\theta_{2,2}^a = \mathbb{E}_{P_{\pi^a}}[I^a|pre(Y_5)] = I_{c,y_2}(C,Y_2)P(y_5|pre(Y_5))$. *Also,* $\omega_1^a = \frac{I_{x,y_1}(X,Y_1)}{P(X|C)P(Y_1|X,C)}$ *and* $\omega_2^a = \omega_1^a \times \frac{I_{y_3,y_4}(Y_3,Y_4)}{P(Y_3|pre(Y_3))P(Y_4|pre(Y_4))}$; *and*

$$\mathcal{V}_{\mathbf{S_X}\setminus\mathbf{X}} = \theta_{0,1}^b + \omega_1^b(\theta_{1,1}^a - \theta_{1,2}^b) + \omega_2^b(\theta_{2,1}^a - \theta_{2,2}^b),$$

for $P_{\pi^b} \equiv I_{c,y_2}(C,Y_2)P(Y_3,Y_4|pre(Y_3))P(X,Y_1|C)$ *and* $I^b \equiv I_{y_1,y_3,y_4}(Y_1,Y_3,Y_4)$, $\theta_{0,1}^b = \mathbb{E}_{P_{\pi^b}}[I^b]$, $\theta_{1,1}^b = \mathbb{E}_{P_{\pi^b}}[I^b|C,X,Y_1] = I_{y_1}(Y_1)P(y_3,y_4|y_2,pre(Y_2))$, $\theta_{1,2}^b = \mathbb{E}_{P_{\pi^b}}[I^b|C] = \sum_{x'} P(y_3,y_4|y_2,y_1,x',C)P(x',y_1|C)$, $\theta_{2,1}^b = I_{y_1,y_3,y_3}(Y_1,Y_3,Y_4)$, $\theta_{2,2}^b = \mathbb{E}_{P_{\pi^b}}[I^b|pre(Y_3)] = I_{y_1,y_3}(Y_1,Y_3)P(y_3,y_4|pre(Y_3))$. *Also,* $\omega_1^b = \frac{I_c(C)}{P(C)}$ *and* $\omega_2^b = \omega_1^b \times \frac{I_{y_2}(Y_2)}{P(Y_2|pre(Y_2))}$.

By Thm. 2, DML-IDP estimator for $P_x(\mathbf{y})$ *is consistent if estimates for either* $P(x|pre(x))P(y_1|pre(y_1))$ *for* ω_1^a *or* $P(y_5|pre(y_5))P(y_2|pre(y_2))$ *for* $(\theta_{0,1}^a,\theta_{1,2}^a)$; *and* $P(x|pre(x))P(y_1|pre(y_1))P(y_3|pre(y_3))P(y_4|pre(y_4))$ *for* ω_2^a *or* $P(y_5|pre(y_5))$ *for* $(\theta_{1,1}^a,\theta_{2,1}^a)$; *and* $P(c)$ *for* ω_1^b *or* $\sum_{x'} P(y_4|pre(y_4))P(y_3|pre(y_3))P(y_1|pre(y_1))P(x'|pre(x'))$ *for* $(\theta_{0,1}^b,\theta_{1,2}^b)$; *and* $P(c)P(y_2|pre(y_2))$ *for* ω_2^b *or* $P(y_4|pre(y_4))P(y_3|pre(y_3))$ *for* $(\theta_{1,1}^b,\theta_{2,2}^b)$ *converge. This implies that DML-IDP estimator is consistent if estimates* $\{P(v_i|pre(v_i)\}_{V_i\in\{X,Y_1,Y_3,Y_4\}}$, *or* $\{P(v_i|pre(v_i)\}_{V_i\in\{X,Y_1,Y_5\}}$, *or* $\{P(v_i|pre(v_i)\}_{V_i\in\{Y_2,Y_5\}}$; *and* $\{P(v_i|pre(v_i))\}_{V_i\in\{C,Y_2\}}$, *or* $\{P(v_i|pre(v_i))\}_{V_i\in\{C,Y_3,Y_4\}}$, *or* $\{P(v_i|pre(v_i)\}_{V_i\in\{X,Y_1,Y_3,Y_4\}}$ *converge. In contrast, PI using Eq. (6) is consistent if estimates for* $\{P(v_i|pre(v_i))\}_{V_i\in\mathbf{V}}$ *converge.*

*(**Fig. 1**). Based on Illustration 3 with PTO* $Y_1 \prec R \prec X \prec Y_2 \prec Y_3 \prec Y_4$. *we have*

$$\begin{aligned}\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})} &= \mathcal{V}_{Q[\mathcal{R}_{Y_2}]} \cdot \mu_{Q[\mathcal{R}_{\mathbf{Y}\setminus\mathcal{R}_{Y_2}}]} + (\mathcal{V}_{Q[\mathcal{R}_{\mathbf{Y}\setminus\mathcal{R}_{Y_2}}]} - \mu_{Q[\mathcal{R}_{\mathbf{Y}\setminus\mathcal{R}_{Y_2}}]}) \cdot \mu_{Q[\mathcal{R}_{Y_2})]},\\ &= \mathcal{V}_{Q[Y_2,Y_3]} \cdot \mu_{Q[Y_1,Y_4]} + \left(\mathcal{V}_{Q[Y_1,Y_4]} - \mu_{Q[Y_1,Y_4]}\right)\mu_{Q[Y_2,Y_3]},\end{aligned}$$

where $\mathbf{Y} \equiv \{Y_1, Y_2, Y_3, Y_4\}$ *and* $\mathcal{R}_{Y_2}$ *is the region (Def. 2) of* Y_2 *with respect to* $\mathbf{Y}$*, and*

$$\mathcal{V}_{Q[Y_2,Y_3]} = \theta^a_{0,1} + \omega^a_1(\theta^a_{1,1} - \theta^a_{1,2}),$$

with $P_{\pi^a} = I_x(X)P(R)P(Y_2, Y_3|X, R)$*, where* $\omega^a_1 = \frac{I_x(X)}{P(X|R)}$*,* $\theta^a_{1,1} = I_{y_2,y_3}(Y_2, Y_3)$*,* $\theta^a_{1,2} = \mathbb{E}_{P_{\pi^a}}[I_{y_2,y_3}(Y_2, Y_3)|X, R] = P(y_2, y_3|X, R)$*,* $\theta^a_{0,1} = \mathbb{E}_{P_{\pi^a}}[I_{y_2,y_3}(Y_2, Y_3)|R] = P(y_2, y_3|x, R)$*, and*

$$\mathcal{V}_{Q[Y_1,Y_4]} = \theta^b_{0,1} + \omega^b_1(\theta^b_{1,1} - \theta^b_{1,2}),$$

with $P_{\pi^b} = I_{r,x,y_2,y_3}(R, X, Y_2, Y_3)P(Y_1)P(Y_4|pre(Y_4))$*,* $\omega^b_1 = \frac{I_{r,x,y_2,y_3}(R,X,Y_2,Y_3)}{P(R|pre(R))P(X|pre(X))P(Y_2|pre(Y_2))P(Y_3|pre(Y_3))}$*,* $\theta^b_{0,1} = \mathbb{E}_{P_{\pi^b}}[I^b|Y_1] = I_{y_1}(Y_1)P(y_4|y_3, y_2, x, r, Y_1)$*,* $\theta^b_{1,1} = I^b$*, and* $\theta^b_{1,2} = \mathbb{E}_{P_{\pi^b}}[I^b|pre(Y_4)] = I_{y_1}(Y_1)P(y_4|pre(Y_4))$ *where* $I^b \equiv I_{y_1,y_4}(Y_1, Y_4)$*.*

By Thm. 2, DML-IDP for $P_x(y_1, y_2, y_3, y_4)$ *is consistent if estimates for either* $P(x|r)$ *for* ω^a_1 *or* $P(y_2|x, r)P(y_3|y_2, x, r)$ *for* $(\theta^a_{0,1}, \theta^a_{1,2})$*; and* $P(r|pre(r))P(x|pre(x))P(y_2|pre(y_2))P(y_3|pre(y_3))$ *for* ω^b_1 *or* $P(y_4|pre(y_4))$ *for* $(\theta^b_{0,1}, \theta^b_{1,2})$ *converge. This condition implies that DML-IDP estimator is consistent if estimates for* $P(x|r)$ *or* $\{P(y_2|x, r), P(y_3|y_2, x, r)\}$*; and* $\{P(v_i|pre(v_i))\}_{V_i \in \{R,X,Y_2,Y_3\}}$ *or* $P(y_4|pre(y_4))$ *converge. In contrast, PI using Eq. (12) is consistent if estimates for* $\{P(y_2|x, r), P(y_3|y_2, x, r), P(r), P(y_4|pre(y_4)), P(y_1)\}$ *converge.*

B. Proofs.

Notations. We will use P_γ to denote parametric submodel $P_\gamma \equiv P(\mathbf{v})(1 + \gamma g(\mathbf{v}))$ for any $\gamma \in \mathbb{R}$ and bounded mean-zero function $g(\cdot)$ over random variables $\mathbf{V}$. For a functional $F(P)$ of a joint distribution P , we will use $F(P_\gamma)$ to denote the functional with respect to P_γ. We will use $\nabla_\gamma F(P_\gamma) \equiv \frac{\partial}{\partial \gamma} F(P_\gamma)|_{\gamma=0}$. We denote $S_\gamma(V_i|\mathbf{W}_i; \gamma = 0) \equiv \nabla_\gamma \log P_\gamma(V_i|\mathbf{W}_i)$. For $\mathbf{T} \subseteq \mathbf{V}$, we use $S(\mathbf{T}) \equiv \nabla_\gamma \log P_\gamma(\mathbf{T})$. Suppose $F(P)$ is composed of conditional probabilities $P(\mathbf{a}_i|\mathbf{b}_i)$. Then, we will use $\nabla_{P_\gamma(\mathbf{a}_i|\mathbf{b}_i)} F(P_\gamma) \equiv (\nabla_\gamma P_\gamma(\mathbf{a}_i|\mathbf{b}_i)) \cdot \frac{\partial F(P)}{\partial P(\mathbf{a}_i|\mathbf{b}_i)}$.

Preliminaries. Once we have a UIF, the corresponding IF for ψ can be expressed as $\phi(\mathbf{V}; \psi, \eta) = \mathcal{V}(\mathbf{V}; \eta) - \mathbb{E}_P[\mathcal{V}(\mathbf{V}; \eta)]$, since $\mathbb{E}_P[\mathcal{V}(\mathbf{V}; \eta)] = \psi$.

B.1. Proofs for Sec. 3

Lemma S.1 (Gateaux derivative of conditional distributions). *Let* $\mathbf{V}$ *be a set of ordered variables (with an order* $\prec$*), and* $\mathbf{T} \subseteq \mathbf{V}$*. For* $V_i \in \mathbf{T}$*, the following holds:*

$$\nabla_\gamma P_\gamma(V_i|pre_{\mathbf{T}}(V_i)) = \left(\mathbb{E}_{P(\mathbf{T})}\left[S(\mathbf{T})|V_i, pre_{\mathbf{T}}(V_i)\right] - \mathbb{E}_{P(\mathbf{T})}\left[S(\mathbf{T})|pre_{\mathbf{T}}(V_i)\right]\right) P(V_i|pre_{\mathbf{T}}(V_i)).$$

Proof. Let $S_\gamma(V_i|\text{pre}_{\mathbf{T}}(V_i); \gamma = 0)$ be shortly denoted as $S(V_i|\text{pre}_{\mathbf{T}}(V_i))$. Then, we first note that $\nabla_\gamma P_\gamma(V_i|\text{pre}_{\mathbf{T}}(V_i)) = S(V_i|\text{pre}_{\mathbf{T}}(V_i))P(V_i|\text{pre}_{\mathbf{T}}(V_i))$, since

$$S(V_i|\text{pre}_{\mathbf{T}}(V_i)) \equiv \nabla_\gamma \log P_\gamma(V_i|\text{pre}_{\mathbf{T}}(V_i)) = \nabla_\gamma P(V_i|\text{pre}_{\mathbf{T}}(V_i)) \underbrace{\frac{\partial}{\partial P(V_i|\text{pre}_{\mathbf{T}}(V_i))} \log P(V_i|\text{pre}_{\mathbf{T}}(V_i))}_{=1/P(V_i|\text{pre}_{\mathbf{T}}(V_i))}.$$

Then, it suffices to show $S(V_i|\text{pre}_{\mathbf{T}}(V_i)) = \left(\mathbb{E}_{P(\mathbf{T})}\left[S(\mathbf{T})|V_i, \text{pre}_{\mathbf{T}}(V_i)\right] - \mathbb{E}_{P(\mathbf{T})}\left[S(\mathbf{T})|\text{pre}_{\mathbf{T}}(V_i)\right]\right)$. We will use the property that the mean of the score function is zero; i.e.,

$$\mathbb{E}_{P(V_i|\text{pre}_{\mathbf{T}}(V_i))}\left[S(V_i|\text{pre}_{\mathbf{T}}(V_i))\right] = \sum_{v_i} \cancel{P(v_i|\text{pre}_{\mathbf{T}}(V_i))} \frac{1}{\cancel{P(v_i|\text{pre}_{\mathbf{T}}(V_i))}} \frac{\partial P_\gamma(v_i|\text{pre}_{\mathbf{T}}(V_i))}{\partial \gamma}|_{\gamma=0} = \frac{\partial}{\partial \gamma} \sum_{v_i} P_\gamma(V_i|\text{pre}_{\mathbf{T}}(V_i)) = 0.$$

Also, from the fact that $P(\mathbf{T}) = \prod_{V_i \in \mathbf{T}} P(V_i|\text{pre}_{\mathbf{T}}(V_i))$, we note $S(\mathbf{T}) = \sum_{V_i \in \mathbf{T}} S(V_i|\text{pre}_{\mathbf{T}}(V_i))$.

For any $V_j \succ V_i$ for $(V_i, V_j) \in \mathbf{T}$, $\mathbb{E}_{P(\mathbf{T})}[S(V_j|\text{pre}_{\mathbf{T}}(V_j))|V_i, \text{pre}_{\mathbf{T}}(V_i)] = \mathbb{E}_{P(\mathbf{T})}[S(V_j|\text{pre}_{\mathbf{T}}(V_j))|\text{pre}_{\mathbf{T}}(V_i)] = 0$, by $\sum_{v_j} S(v_j|\text{pre}_{\mathbf{T}}(v_j))P(v_j|\text{pre}_{\mathbf{T}}(v_j)) = 0$.

For any $V_j \prec V_i$, $\mathbb{E}_{P(\mathbf{T})}[S(V_j|\text{pre}_{\mathbf{T}}(V_j))|V_i, \text{pre}_{\mathbf{T}}(V_i)] = \mathbb{E}_{P(\mathbf{T})}[S(V_j|\text{pre}_{\mathbf{T}}(V_j))|\text{pre}_{\mathbf{T}}(V_i)] = S(V_j|\text{pre}_{\mathbf{T}}(V_j))$ since $\{V_j, \text{pre}_{\mathbf{T}}(V_j)\} \subseteq \text{pre}_{\mathbf{T}}(V_i)$. Hence, $\mathbb{E}_{P(\mathbf{T})}[S(V_j|\text{pre}_{\mathbf{T}}(V_j))|V_i, \text{pre}_{\mathbf{T}}(V_i)] - \mathbb{E}_{P(\mathbf{T})}[S(V_j|\text{pre}_{\mathbf{T}}(V_j))|\text{pre}_{\mathbf{T}}(V_i)] = 0$.

For $V_j = V_i$, $\mathbb{E}_{P(\mathbf{T})}[S(V_i|\text{pre}_{\mathbf{T}}(V_i))|V_i, \text{pre}_{\mathbf{T}}(V_i)] = S(V_i|\text{pre}_{\mathbf{T}}(V_i))$ and $\mathbb{E}_{P(\mathbf{T})}[S(V_i|\text{pre}_{\mathbf{T}}(V_i))|\text{pre}_{\mathbf{T}}(V_i)] = 0$. Therefore,

$$S(V_i|\text{pre}_{\mathbf{T}}(V_i)) = \left(\mathbb{E}_{P(\mathbf{T})}\left[S(\mathbf{T})|V_i, \text{pre}_{\mathbf{T}}(V_i)\right] - \mathbb{E}_{P(\mathbf{T})}\left[S(\mathbf{T})|\text{pre}_{\mathbf{T}}(V_i)\right]\right).$$

This completes the proof. □

Lemma S.2. *Let $\mathbf{V}$ be a set of ordered variables and $\mathbf{T} = (\mathbf{Z} \cup \mathbf{X} \cup \mathbf{Y}) \subseteq \mathbf{V}$. Let $\mathbf{W}_{V_i} \equiv pre_{\mathbf{T}}(V_i)$ (shortly, $\mathbf{W}_i$). Let the observational model P and the interventional model $P_{\pi(\mathbf{x})}$ over $\mathbf{T}$ be defined as follows:*

$$P(\mathbf{T}) \equiv \prod_{V_i \in \{\mathbf{Z},\mathbf{Y}\}} P(v_i|\mathbf{w}_i) \prod_{X_j \in \mathbf{X}} P(x_j|\mathbf{w}_{X_j}),$$
$$P_{\pi(\mathbf{x})}(\mathbf{T}) \equiv \prod_{V_i \in \{\mathbf{Z},\mathbf{Y}\}} P(v_i|\mathbf{w}_i) \prod_{X_j \in \mathbf{X}} I_{x_j}(X_j).$$

Let $q(\mathbf{X}) \equiv \prod_{X_i \in \mathbf{X}} P(X_i|\mathbf{W}_{X_i})$ and $\pi(\mathbf{X}) \equiv \prod_{X_i \in \mathbf{X}} I_{x_i}(X_i)$. Let $\mathbb{E}_\pi[\cdot]$ be an expectation with respect to the distribution $P_{\pi(\mathbf{x})}$. For $V_i \notin \mathbf{X}$, the following holds:

$$\nabla_\gamma P_\gamma(V_i|\mathbf{W}_i) = \left(\mathbb{E}_\pi\left[\frac{q(\mathbf{X})}{\pi(\mathbf{X})}\frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)}S(\mathbf{T}) \;\middle|\; V_i, \mathbf{W}_i\right] - \mathbb{E}_\pi\left[\frac{q(\mathbf{X})}{\pi(\mathbf{X})}\frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)}S(\mathbf{T}) \;\middle|\; \mathbf{W}_i\right]\right) \cdot P(V_i|\mathbf{W}_i).$$

Proof. By Lemma S.1, we have

$$\nabla_\gamma P_\gamma(V_i|\mathbf{W}_i) = \left(\mathbb{E}_{P(\mathbf{T})}\left[S(\mathbf{T})|V_i, \text{pre}_{\mathbf{T}}(V_i)\right] - \mathbb{E}_{P(\mathbf{T})}\left[S(\mathbf{T})|\text{pre}_{\mathbf{T}}(V_i)\right]\right) P(V_i|\text{pre}_{\mathbf{T}}(V_i)).$$

We first consider $\mathbb{E}_{P(\mathbf{T})}\left[S(\mathbf{T})|V_i, \text{pre}_{\mathbf{T}}(V_i)\right]$ and show that

$$\mathbb{E}_\pi\left[\frac{q(\mathbf{X})}{\pi(\mathbf{X})}\frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)}S(\mathbf{T}) \;\middle|\; V_i, \mathbf{W}_i\right] = \mathbb{E}_{P(\mathbf{T})}\left[S(\mathbf{T})|V_i, \text{pre}_{\mathbf{T}}(V_i)\right].$$

The equality can be shown as follows:

$$\begin{aligned}
\mathbb{E}_\pi\left[\frac{q(\mathbf{X})}{\pi(\mathbf{X})}\frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)}S(\mathbf{T}) \;\middle|\; V_i, \mathbf{W}_i\right] &= \int_{\mathbf{t}} \frac{q(\mathbf{x})}{\pi(\mathbf{x})}\frac{P_\pi(\mathbf{w}_i)}{P(\mathbf{w}_i)}\frac{P_\pi(\mathbf{t})}{P_\pi(v_i|\mathbf{w}_i)P_\pi(\mathbf{w}_i)}s(\mathbf{t})d\mathbf{t} \\
&= \int_{\mathbf{t}} \frac{1}{\cancel{\pi(\mathbf{x})}}\frac{\cancel{P_\pi(\mathbf{w}_i)}}{P(\mathbf{w}_i)}\underbrace{q(\mathbf{x})P_\pi(\mathbf{t}\backslash\mathbf{x})}_{=P(\mathbf{t})}\cancel{\pi(\mathbf{x})}s(\mathbf{t})\frac{1}{\underbrace{P_\pi(v_i|\mathbf{w}_i)}_{=P(v_i|\mathbf{w}_i)}\cancel{P_\pi(\mathbf{w}_i)}}d\mathbf{t} \\
&= \int_{\mathbf{t}} \frac{P(\mathbf{t})}{P(\mathbf{w}_i)P(v_i|\mathbf{w}_i)}S(\mathbf{t})d\mathbf{t} \\
&= \mathbb{E}_{P(\mathbf{T})}\left[S(\mathbf{T})|V_i, \mathbf{W}_i\right],
\end{aligned}$$

where we have used $P_\pi(v_i|\mathbf{w}_i) = P(v_i|\mathbf{w}_i)$ for $V_i \in \mathbf{T}\backslash\mathbf{X}$. To witness, we have the following:

$$\begin{aligned}
P_\pi(V_i|\mathbf{W}_i) &= \frac{P_\pi(V_i, \mathbf{W}_i)}{P_\pi(\mathbf{W}_i)} \\
&= \frac{\prod_{V_j \in \{V_i, \text{pre}_{\mathbf{T}}(V_i)\}\backslash\mathbf{X}} P(V_j|\mathbf{W}_j)\prod_{X_k \in \{V_i, \text{pre}_{\mathbf{T}}(V_i)\}\cap\mathbf{X}} I_{x_k}(X_k)}{\prod_{V_j \in \{\text{pre}_{\mathbf{T}}(V_i)\}\backslash\mathbf{X}} P(V_j|\mathbf{W}_j)\prod_{X_k \in \{\text{pre}_{\mathbf{T}}(V_i)\}\cap\mathbf{X}} I_{x_k}(X_k)} \\
&= \frac{\prod_{V_j \in \{V_i, \text{pre}_{\mathbf{T}}(V_i)\}\backslash\mathbf{X}} P(V_j|\mathbf{W}_j)\cancel{\prod_{X_k \in \{\text{pre}_{\mathbf{T}}(V_i)\}\cap\mathbf{X}} I_{x_k}(X_k)}}{\prod_{V_j \in \{\text{pre}_{\mathbf{T}}(V_i)\}\backslash\mathbf{X}} P(V_j|\mathbf{W}_j)\cancel{\prod_{X_k \in \{\text{pre}_{\mathbf{T}}(V_i)\}\cap\mathbf{X}} I_{x_k}(X_k)}} \\
&= P(V_i|\mathbf{W}_i),
\end{aligned}$$

where the third equality holds since $V_k \notin \mathbf{X}$, and the second equality holds, since

$$P_\pi(V_i, \mathbf{W}_i) = \sum_{\mathbf{t}\setminus\{v_i,\mathbf{w}_i\}} P_\pi(\mathbf{t}) = \sum_{\mathbf{t}\setminus\{v_i,\mathbf{w}_i\}} \prod_{V_i\in\{\mathbf{Z},\mathbf{Y}\}} P(v_i|\mathbf{w}_i) \prod_{X_j\in\mathbf{X}} I_{x_j}(X_j)$$
$$= \prod_{V_j\in\{V_i,\text{pre}_\mathbf{T}(V_i)\}\setminus\mathbf{X}} P(v_j|\mathbf{w}_j) \prod_{X_k\in\{V_i,\text{pre}_\mathbf{T}(V_i)\}\cap\mathbf{X}} I_{x_k}(X_k),$$

because conditional probabilities $P(v_k|\text{pre}_\mathbf{T}(v_i))$ for $V_k \notin \{V_i, \text{pre}_\mathbf{T}(V_i)\}$ and $I_{x_p}(X_p)$ for $X_p \notin \{V_i, \text{pre}_\mathbf{T}(V_i)\}$ are summed out. Similarly, we have $P_\pi(\mathbf{W}_i) = \prod_{V_j\in\{\text{pre}_\mathbf{T}(V_i)\}\setminus\mathbf{X}} P(V_j|\mathbf{W}_j) \prod_{X_k\in\{\text{pre}_\mathbf{T}(V_i)\}\cap\mathbf{X}} I_{x_k}(X_k)$.

We now show

$$\mathbb{E}_\pi\left[\frac{q(\mathbf{X})}{\pi(\mathbf{X})}\frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)}S(\mathbf{T}) \;\middle|\; \mathbf{W}_i\right] = \mathbb{E}_{P(\mathbf{T})}\left[S(\mathbf{T})|\text{pre}_\mathbf{T}(V_i)\right].$$

The equality can be shown as follows:

$$\mathbb{E}_\pi\left[\frac{q(\mathbf{X})}{\pi(\mathbf{X})}\frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)}S(\mathbf{T}) \;\middle|\; \mathbf{W}_i\right] = \int_\mathbf{t} \frac{q(\mathbf{x})}{\pi(\mathbf{x})}\frac{P_\pi(\mathbf{w}_i)}{P(\mathbf{w}_i)}\frac{P_\pi(\mathbf{t})}{P_\pi(\mathbf{w}_i)}s(\mathbf{t})d\mathbf{t}$$
$$= \int_\mathbf{t} \frac{1}{\cancel{\pi(\mathbf{x})}}\frac{\cancel{P_\pi(\mathbf{w}_i)}}{P(\mathbf{w}_i)}\underbrace{q(\mathbf{x})P_\pi(\mathbf{t}\setminus\mathbf{x})}_{=P(\mathbf{t})}\cancel{\pi(\mathbf{x})}s(\mathbf{t})\frac{1}{\cancel{P_\pi(\mathbf{w}_i)}}d\mathbf{t}$$
$$= \int_\mathbf{t} \frac{P(\mathbf{t})}{P(\mathbf{w}_i)}S(\mathbf{t})d\mathbf{t}$$
$$= \mathbb{E}_{P(\mathbf{T})}\left[S(\mathbf{T})|\mathbf{W}_i\right].$$

Therefore,

$$\nabla_\gamma P_\gamma(V_i|\mathbf{W}_i) = \left(\mathbb{E}_{P(\mathbf{T})}\left[S(\mathbf{T})|V_i,\mathbf{W}_i\right] - \mathbb{E}_{P(\mathbf{T})}\left[S(\mathbf{T})|\mathbf{W}_i\right]\right)P(V_i|\mathbf{W}_i).$$
$$= \left(\mathbb{E}_\pi\left[\frac{q(\mathbf{X})}{\pi(\mathbf{X})}\frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)}S(\mathbf{T}) \;\middle|\; V_i,\mathbf{W}_i\right] - \mathbb{E}_\pi\left[\frac{q(\mathbf{X})}{\pi(\mathbf{X})}\frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)}S(\mathbf{T}) \;\middle|\; \mathbf{W}_i\right]\right)\cdot P(V_i|\mathbf{W}_i).$$

□

Lemma S.3. *Let $\mathbf{V}$ be a set of ordered variables and $\mathbf{T} = (\mathbf{Z}\cup\mathbf{X}\cup\mathbf{Y}) \subseteq \mathbf{V}$. Let $\mathbf{W}_{V_i} \equiv pre_\mathbf{T}(V_i)$ (shortly, $\mathbf{W}_i$). Let $P(\mathbf{T}) \equiv \prod_{V_i\in\{\mathbf{Z},\mathbf{Y}\}} P(v_i|\mathbf{w}_i)\prod_{X_j\in\mathbf{X}} P(x_j|\mathbf{w}_{x_j})$, and $P_{\pi(\mathbf{x})}(\mathbf{T}) \equiv \prod_{V_i\in\{\mathbf{Z},\mathbf{Y}\}} P(v_i|\mathbf{w}_i)\prod_{X_j\in\mathbf{X}} I_{x_j}(X_j)$. Let $\mathbb{E}_\pi[\cdot]$ be an expectation with respect to the distribution $P_{\pi(\mathbf{x})}(\mathbf{T})$. Let $\psi \equiv \Psi(P) \equiv \mathbb{E}_\pi[I_\mathbf{y}(\mathbf{Y})] = \sum_\mathbf{z}\prod_{Y_j\in\mathbf{Y}} P(y_j|pre_\mathbf{T}(y_j))\prod_{Z_k\in\mathbf{Z}} P(z_k|pre_\mathbf{T}(z_k))$. Then, an influence function for ψ is given as*

$$\phi = \sum_{V_k\in\{\mathbf{Y},\mathbf{Z}\}} \frac{P_\pi(\mathbf{W}_k)}{P(\mathbf{W}_k)}\left\{\mathbb{E}_\pi\left[I_\mathbf{y}(\mathbf{Y})|V_k,\mathbf{W}_k\right] - \mathbb{E}_\pi\left[I_\mathbf{y}(\mathbf{Y})|\mathbf{W}_k\right]\right\}. \tag{B.1}$$

where we use $\sum_{V_k\in\{\mathbf{Y},\mathbf{Z}\}}$ as a shorthand for $\sum_{k:V_k\in\{\mathbf{Y},\mathbf{Z}\}}$.

Proof. Let $\pi(\mathbf{X}) \equiv \prod_{X_i\in\mathbf{X}} I_{x_i}(X_i)$. Let $q(\mathbf{X}) \equiv \prod_{X_i\in\mathbf{X}} P(X_i|\mathbf{W}_{X_i})$. We will prove the following:

$$\nabla_\gamma\Psi(P_\gamma) = \mathbb{E}_{P(\mathbf{T})}\left[\left(\sum_{V_k\in\{\mathbf{Y},\mathbf{Z}\}} \frac{P_\pi(\mathbf{W}_k)}{P(\mathbf{W}_k)}\left\{\mathbb{E}_\pi\left[I_\mathbf{y}(\mathbf{Y})|V_k,\mathbf{W}_k\right] - \mathbb{E}_\pi\left[I_\mathbf{y}(\mathbf{Y})|\mathbf{W}_k\right]\right\}\right)\cdot S(\mathbf{T})\right] \tag{B.2}$$
$$= \mathbb{E}_{P(\mathbf{V})}\left[\left(\sum_{V_k\in\{\mathbf{Y},\mathbf{Z}\}} \frac{P_\pi(\mathbf{W}_k)}{P(\mathbf{W}_k)}\left\{\mathbb{E}_\pi\left[I_\mathbf{y}(\mathbf{Y})|V_k,\mathbf{W}_k\right] - \mathbb{E}_\pi\left[I_\mathbf{y}(\mathbf{Y})|\mathbf{W}_k\right]\right\}\right)\cdot S(\mathbf{V})\right], \tag{B.3}$$

where $S(\mathbf{T}) \equiv \sum_{V_k \in \mathbf{T}} S_\gamma(V_k|\mathbf{W}_k; \gamma = 0)$. The second equality holds as follow: Note $\left(\sum_{V_k \in \{\mathbf{Y},\mathbf{Z}\}} \frac{P_\pi(\mathbf{W}_k)}{P(\mathbf{W}_k)} \{\mathbb{E}_\pi [I_\mathbf{y}(\mathbf{Y})|V_k, \mathbf{W}_k] - \mathbb{E}_\pi [I_\mathbf{y}(\mathbf{Y})|\mathbf{W}_k]\}\right)$ is only a function of $\mathbf{T}$ (say, $f(\mathbf{T})$). We first note that

$$\mathbb{E}_{P(\mathbf{T})} [[f(\mathbf{T})S(\mathbf{T})]] = \sum_{\mathbf{t}} f(\mathbf{t})S(\mathbf{t})P(\mathbf{t}) = \sum_{\mathbf{v}\backslash\mathbf{t},\mathbf{t}} f(\mathbf{t})S(\mathbf{t})P(\mathbf{t}, \mathbf{v}\backslash\mathbf{t}) = \mathbb{E}_{P(\mathbf{V})} [f(\mathbf{T})S(\mathbf{T})] .$$

Then, we will show that $\mathbb{E}_{P(\mathbf{V})} [f(\mathbf{T})S(\mathbf{V}\backslash\mathbf{T})] = 0$, where $S(\mathbf{V}\backslash\mathbf{T})$ is a score function of $P(\mathbf{v}\backslash\mathbf{t}|\mathbf{t})$, namely, $S(\mathbf{V}\backslash\mathbf{T}) \equiv \sum_{V_k \in \mathbf{V}\backslash\mathbf{T}} S_\gamma(V_k|\text{pre}_{\mathbf{V}\backslash\mathbf{T}}(V_k), \mathbf{T})$, where $S_\gamma(V_k|\text{pre}_{\mathbf{V}\backslash\mathbf{T}}(V_k), \mathbf{T})$ is a score function of $P(v_k|\text{pre}_{\mathbf{V}\backslash\mathbf{T}}(v_k), \mathbf{t})$. We note $P(\mathbf{v}) = P(\mathbf{t})P(\mathbf{v}\backslash\mathbf{t}|\mathbf{t}) = \prod_{V_j \in \mathbf{T}} P(v_j|\text{pre}_\mathbf{T}(v_j)) \prod_{V_k \in \mathbf{V}\backslash\mathbf{T}} P(v_k|\text{pre}_{\mathbf{V}\backslash\mathbf{T}}(v_k))$, and this implies that $S(\mathbf{V}) = S(\mathbf{T}) + S(\mathbf{V}\backslash\mathbf{T})$. Then, given the equality $\mathbb{E}_{P(\mathbf{V})} [f(\mathbf{T})S(\mathbf{V}\backslash\mathbf{T})] = 0$, we can show the equality in Eq. (B.3) as follow:

$$\begin{aligned}
\mathbb{E}_{P(\mathbf{T})} [f(\mathbf{T})S(\mathbf{T})] &= \mathbb{E}_{P(\mathbf{V})} [f(\mathbf{T})S(\mathbf{T})] + \mathbb{E}_{P(\mathbf{V})} [f(\mathbf{T})S(\mathbf{V}\backslash\mathbf{T})] \\
&= \mathbb{E}_{P(\mathbf{V})} [f(\mathbf{T}) (S(\mathbf{T}) + S(\mathbf{V}\backslash\mathbf{T}))] \\
&= \mathbb{E}_{P(\mathbf{V})} [f(\mathbf{T})S(\mathbf{V})] .
\end{aligned}$$

We witness $\mathbb{E}_{P(\mathbf{V})} [f(\mathbf{T})S(\mathbf{V}\backslash\mathbf{T})] = 0$ as follow:

$$\mathbb{E}_{P(\mathbf{V})} [f(\mathbf{T})S(\mathbf{V}\backslash\mathbf{T})] = \sum_{\mathbf{v}} f(\mathbf{t})S(\mathbf{v}\backslash\mathbf{t})P(\mathbf{v}) = \sum_{\mathbf{t}} (\underbrace{\sum_{\mathbf{v}\backslash\mathbf{t}} S(\mathbf{v}\backslash\mathbf{t})P(\mathbf{v}\backslash\mathbf{t}|\mathbf{t}))}_{=0} P(\mathbf{t})f(\mathbf{t}) = 0.$$

The second equality implies that Eq. (B.1) is an IF for ψ. The second equality implies that Eq. (B.1) is an IF for ψ.

We now prove the first equality. Let $\nabla_{P_\gamma(v_k|\mathbf{w}_k)} \Psi(P_\gamma) \equiv \nabla_\gamma P_\gamma(v_k|\mathbf{w}_k) \cdot \frac{\partial \Psi(P)}{\partial P(v_k|\mathbf{w}_k)}$, for $V_k \in \{\mathbf{Y}, \mathbf{Z}\}$. Then,

$$\nabla_\gamma \Psi(P_\gamma) = \sum_{V_k \in \{\mathbf{Y},\mathbf{Z}\}} \nabla_{P_\gamma(v_k|\mathbf{w}_k)} \Psi(P_\gamma),$$

by the chain rule. A closed form of $\nabla_{P_\gamma(v_k|\mathbf{w}_k)} \Psi(P_\gamma)$ is given as follows using Lemma S.2.

$$\begin{aligned}
&\nabla_{P_\gamma(v_k|\mathbf{w}_k)} \Psi(P_\gamma) \\
&= \nabla_{P_\gamma(v_k|\mathbf{w}_k)} \sum_{\mathbf{x},\mathbf{y},\mathbf{z}} I_\mathbf{y}(\mathbf{Y}) \prod_{V_i \in \{\mathbf{Z},\mathbf{Y}\}} P(v_i|\mathbf{w}) \prod_{X_j \in \mathbf{X}} I_{x_j}(X_j) \\
&= \sum_{\mathbf{x},\mathbf{y},\mathbf{z}} I_\mathbf{y}(\mathbf{Y}) \left(\mathbb{E}_\pi \left[S(\mathbf{T}) \frac{q(\mathbf{X})}{\pi(\mathbf{X})} \frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)} \,\middle|\, V_k, \mathbf{W}_k\right] - \mathbb{E}_\pi \left[S(\mathbf{T}) \frac{q(\mathbf{X})}{\pi(\mathbf{X})} \frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)} \,\middle|\, \mathbf{W}_k\right]\right) P_\pi (\mathbf{z}, \mathbf{x}, \mathbf{y}) , \\
&= \mathbb{E}_\pi \left[I_\mathbf{y}(\mathbf{Y}) \left(\mathbb{E}_\pi \left[S(\mathbf{T}) \frac{q(\mathbf{X})}{\pi(\mathbf{X})} \frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)} \,\middle|\, V_k, \mathbf{W}_k\right] - \mathbb{E}_\pi \left[S(\mathbf{T}) \frac{q(\mathbf{X})}{\pi(\mathbf{X})} \frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)} \,\middle|\, \mathbf{W}_k\right]\right)\right] .
\end{aligned}$$

Note

$$\mathbb{E}_\pi \left[I_\mathbf{y}(\mathbf{Y}) \mathbb{E}_\pi \left[S(\mathbf{T}) \frac{q(\mathbf{X})}{\pi(\mathbf{X})} \frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)} \,\middle|\, V_k, \mathbf{W}_k\right]\right] = \mathbb{E}_\pi \left[\mathbb{E}_\pi \left[I_\mathbf{y}(\mathbf{Y}) \frac{q(\mathbf{X})}{\pi(\mathbf{X})} \frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)} \,\middle|\, V_k, \mathbf{W}_k\right] \cdot S(\mathbf{T})\right] ,$$

and

$$\mathbb{E}_\pi \left[I_\mathbf{y}(\mathbf{Y}) \mathbb{E}_\pi \left[S(\mathbf{T}) \frac{q(\mathbf{X})}{\pi(\mathbf{X})} \frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)} \,\middle|\, \mathbf{W}_k\right]\right] = \mathbb{E}_\pi \left[\mathbb{E}_\pi \left[I_\mathbf{y}(\mathbf{Y}) \frac{q(\mathbf{X})}{\pi(\mathbf{X})} \frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)} \,\middle|\, \mathbf{W}_k\right] \cdot S(\mathbf{T})\right] .$$

To witness, let $f(\mathbf{X}, \mathbf{W}_k) \equiv \frac{q(\mathbf{X})}{\pi(\mathbf{X})} \frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)}$. Then,

$$
\begin{aligned}
&\mathbb{E}_\pi \left[I_{\mathbf{y}}(\mathbf{Y}) \mathbb{E}_\pi \left[S(\mathbf{T}) f(\mathbf{X}, \mathbf{W}_k) | V_k, \mathbf{W}_k \right] \right] \\
&= \sum_{\mathbf{z}', \mathbf{x}', \mathbf{y}'} I_{\mathbf{y}}(\mathbf{y}') \mathbb{E}_\pi \left[S(\mathbf{T}) f(\mathbf{X}, \mathbf{W}_k) | v'_k, \mathbf{w}'_k \right] \prod_{V_i \in \{\mathbf{Z}, \mathbf{Y}\}} P(v'_i | \mathbf{w}'_i) \prod_{X_j \in \mathbf{X}} I_{x_j}(x'_j) \\
&= \sum_{\mathbf{z}', \mathbf{x}', \mathbf{y}'} I_{\mathbf{y}}(\mathbf{y}') \left(\sum_{\mathbf{z}'', \mathbf{x}'', \mathbf{y}'' \setminus \{v'_k, \mathbf{w}'_k\}} S(\mathbf{t}'') f(\mathbf{x}'', \mathbf{w}'_k) \frac{P_\pi(\mathbf{z}'', \mathbf{x}'', \mathbf{y}'')}{P_\pi(v'_k, \mathbf{w}'_k)} \right) \prod_{V_i \in \{\mathbf{Z}, \mathbf{Y}\}} P(v'_i | \mathbf{w}'_i) \prod_{X_j \in \mathbf{X}} I_{x_j}(x'_j) \\
&= \sum_{\mathbf{z}', \mathbf{x}', \mathbf{y}' \setminus \{v'_k, \mathbf{w}'_k\}} I_{\mathbf{y}}(\mathbf{y}') \left(\sum_{\mathbf{z}'', \mathbf{x}'', \mathbf{y}''} S(\mathbf{t}'') f(\mathbf{x}'', \mathbf{w}'_k) \frac{P_\pi(\mathbf{z}'', \mathbf{x}'', \mathbf{y}'')}{\cancel{P_\pi(v'_k, \mathbf{w}'_k)}} \cancel{P_\pi(v'_k, \mathbf{w}'_k)} \right) \frac{P_\pi(\mathbf{t}')}{P_\pi(v'_k, \mathbf{w}'_k)} \\
&= \sum_{\mathbf{z}'', \mathbf{x}'', \mathbf{y}''} \left(\sum_{\mathbf{z}', \mathbf{x}', \mathbf{y}' \setminus \{v'_k, \mathbf{w}'_k\}} I_{\mathbf{y}}(\mathbf{y}') f(\mathbf{x}'', \mathbf{w}'_k) \frac{P_\pi(\mathbf{t}')}{P_\pi(v'_k, \mathbf{w}'_k)} \right) S(\mathbf{t}'') P_\pi(\mathbf{t}'') \\
&= \mathbb{E}_\pi \left[\mathbb{E}_\pi \left[I_{\mathbf{y}}(\mathbf{Y}) \frac{q(\mathbf{X})}{\pi(\mathbf{X})} \frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)} \,\middle|\, V_k, \mathbf{W}_k \right] \cdot S(\mathbf{T}) \right].
\end{aligned}
$$

Also,

$$
\begin{aligned}
&\mathbb{E}_\pi \left[I_{\mathbf{y}}(\mathbf{Y}) \mathbb{E}_\pi \left[S(\mathbf{T}) f(\mathbf{X}, \mathbf{W}_k) | \mathbf{W}_k \right] \right] \\
&= \sum_{\mathbf{z}', \mathbf{x}', \mathbf{y}'} I_{\mathbf{y}}(\mathbf{y}') \mathbb{E}_\pi \left[S(\mathbf{T}) f(\mathbf{X}, \mathbf{W}_k) | \mathbf{w}'_k \right] \prod_{V_i \in \{\mathbf{Z}, \mathbf{Y}\}} P(v'_i | \mathbf{w}'_i) \prod_{X_j \in \mathbf{X}} I_{x_j}(x'_j) \\
&= \sum_{\mathbf{z}', \mathbf{x}', \mathbf{y}'} I_{\mathbf{y}}(\mathbf{y}') \left(\sum_{\mathbf{z}'', \mathbf{x}'', \mathbf{y}'' \setminus \{\mathbf{w}'_k\}} S(\mathbf{t}'') f(\mathbf{x}'', \mathbf{w}'_k) \frac{P_\pi(\mathbf{z}'', \mathbf{x}'', \mathbf{y}'')}{P_\pi(\mathbf{w}'_k)} \right) \prod_{V_i \in \{\mathbf{Z}, \mathbf{Y}\}} P(v'_i | \mathbf{w}'_i) \prod_{X_j \in \mathbf{X}} I_{x_j}(x'_j) \\
&= \sum_{\mathbf{z}', \mathbf{x}', \mathbf{y}' \setminus \{\mathbf{w}'_k\}} I_{\mathbf{y}}(\mathbf{y}') \left(\sum_{\mathbf{z}'', \mathbf{x}'', \mathbf{y}''} S(\mathbf{t}'') f(\mathbf{x}'', \mathbf{w}'_k) \frac{P_\pi(\mathbf{z}'', \mathbf{x}'', \mathbf{y}'')}{\cancel{P_\pi(\mathbf{w}'_k)}} \cancel{P_\pi(\mathbf{w}'_k)} \right) \frac{P_\pi(\mathbf{t}')}{P_\pi(\mathbf{w}'_k)} \\
&= \sum_{\mathbf{z}'', \mathbf{x}'', \mathbf{y}''} \left(\sum_{\mathbf{z}', \mathbf{x}', \mathbf{y}' \setminus \{\mathbf{w}'_k\}} I_{\mathbf{y}}(\mathbf{y}') f(\mathbf{x}'', \mathbf{w}'_k) \frac{P_\pi(\mathbf{t}')}{P_\pi(\mathbf{w}'_k)} \right) S(\mathbf{t}'') P_\pi(\mathbf{t}'') \\
&= \mathbb{E}_\pi \left[\mathbb{E}_\pi \left[I_{\mathbf{y}}(\mathbf{Y}) \frac{q(\mathbf{X})}{\pi(\mathbf{X})} \frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)} \,\middle|\, \mathbf{W}_k \right] \cdot S(\mathbf{T}) \right].
\end{aligned}
$$

Then, for $V_k \in \{\mathbf{Y}, \mathbf{Z}\}$,

$$
\begin{aligned}
\nabla_{P_\gamma(v_k | \mathbf{w}_k)} \Psi(P) &= \mathbb{E}_\pi \left[\left(\mathbb{E}_\pi \left[I_{\mathbf{y}}(\mathbf{Y}) \frac{q(\mathbf{X})}{\pi(\mathbf{X})} \frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)} \,\middle|\, V_k, \mathbf{W}_k \right] - \mathbb{E}_\pi \left[I_{\mathbf{y}}(\mathbf{Y}) \frac{q(\mathbf{X})}{\pi(\mathbf{X})} \frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)} \,\middle|\, \mathbf{W}_k \right] \right) \cdot S(\mathbf{T}) \right] \\
&= \mathbb{E}_{P(\mathbf{T})} \left[\frac{P_\pi(\mathbf{T})}{P(\mathbf{T})} \left(\mathbb{E}_\pi \left[I_{\mathbf{y}}(\mathbf{Y}) \frac{q(\mathbf{X})}{\pi(\mathbf{X})} \frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)} \,\middle|\, V_k, \mathbf{W}_k \right] - \mathbb{E}_\pi \left[I_{\mathbf{y}}(\mathbf{Y}) \frac{q(\mathbf{X})}{\pi(\mathbf{X})} \frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)} \,\middle|\, \mathbf{W}_k \right] \right) \cdot S(\mathbf{T}) \right] \\
&= \mathbb{E}_{P(\mathbf{T})} \left[\frac{\pi(\mathbf{X})}{q(\mathbf{X})} \left(\mathbb{E}_\pi \left[I_{\mathbf{y}}(\mathbf{Y}) \frac{q(\mathbf{X})}{\pi(\mathbf{X})} \frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)} \,\middle|\, V_k, \mathbf{W}_k \right] - \mathbb{E}_\pi \left[I_{\mathbf{y}}(\mathbf{Y}) \frac{q(\mathbf{X})}{\pi(\mathbf{X})} \frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)} \,\middle|\, \mathbf{W}_k \right] \right) \cdot S(\mathbf{T}) \right] \\
&= \mathbb{E}_{P(\mathbf{T})} \left[\frac{P_\pi(\mathbf{W}_k)}{P(\mathbf{W}_k)} \left\{ \mathbb{E}_\pi \left[I_{\mathbf{y}}(\mathbf{Y}) | V_k, \mathbf{W}_k \right] - \mathbb{E}_\pi \left[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{W}_k \right] \right\} \cdot S(\mathbf{T}) \right].
\end{aligned}
$$

This proves Eq. (B.2). □

Lemma B.1 (UIF for CE-1 (Restated Lemma 1)**).** *Let a target estimand $\psi = \mathcal{Q}$ be a CE-1 given by Eq. (1) in Def. 3. Let $\mathbf{Y} \equiv \mathbf{C} \backslash \mathbf{A}$, and $\mathbf{X} \equiv \mathbf{T} \setminus \mathbf{C} \equiv \{\mathbf{B}_{j_1} < \cdots < \mathbf{B}_{j_m}\}$ where $\mathbf{B}_{j_s} \in \mathbf{T}$. Let $\mathbf{C}$ be partitioned with respect to $\mathbf{X}$ as $\mathbf{C} = \bigcup_{k=0}^{m} \mathbf{C}_k$, where $\mathbf{C}_k \equiv \{\mathbf{B}_r \in \mathbf{C} : j_k < r < j_{k+1}\} \equiv \{\mathbf{B}_{k_{\min}} < \cdots < \mathbf{B}_{k_{\max}}\}$ with $j_0 \equiv 0$ and $j_{m+1} \equiv n+1$. Let*

P_π be a distribution over $\mathbf{T}$ given by $P_\pi \equiv I_{\mathbf{x}}(\mathbf{X}) \prod_{\mathbf{B}_i \in \mathbf{C}} P(\mathbf{B}_i | pre_{\mathbf{T}}(\mathbf{B}_i))$. Then, $\mathcal{V}(\mathbf{T}; \eta = (\boldsymbol{\omega}, \boldsymbol{\theta}))$ in the following is a UIF for ψ:

$$\mathcal{V}(\mathbf{T}; \eta = (\boldsymbol{\omega}, \boldsymbol{\theta})) = \theta_{0,1} + \sum_{\substack{k=1 \\ \mathbf{C}_k \neq \emptyset}}^{m} \omega_k \left(\theta_{k,1} - \theta_{k,2}\right), \tag{B.4}$$

where $\boldsymbol{\omega} \equiv \{\omega_k |\ \mathbf{C}_k \neq \emptyset,\ k \in \{1, \cdots, m\}\}$ and $\boldsymbol{\theta} \equiv \{\theta_{0,1}\} \cup \{(\theta_{k,1}, \theta_{k,2}) |\ \mathbf{C}_k \neq \emptyset,\ k \in \{1, \cdots, m\}\}$ are nuisances given by $\omega_k \equiv \prod_{r=1}^{k} \frac{I_{\mathbf{b}_{j_r}}(\mathbf{B}_{j_r})}{P(\mathbf{B}_{j_r} | pre_{\mathbf{T}}(\mathbf{B}_{j_r}))}$, $\theta_{k,1} \equiv \mathbb{E}_{P_\pi}\left[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{B}_{k_{\max}}, pre_{\mathbf{T}}(\mathbf{B}_{k_{\max}})\right]$, $\theta_{k,2} \equiv \mathbb{E}_{P_\pi}\left[I_{\mathbf{y}}(\mathbf{Y}) | pre_{\mathbf{T}}(\mathbf{B}_{k_{\min}})\right]$ where $\theta_{0,1} = \mathbb{E}_{P_\pi}[I_{\mathbf{y}}(\mathbf{Y})]$ if $\mathbf{C}_0 = \emptyset$.

Proof. We first note that Lemma S.3 holds when each $V_k \in \mathbf{T}$ is a bucket instead of being a singleton, with the definition $\mathbf{W}_k \equiv \text{pre}_{\mathbf{T}}(\mathbf{B}_k)$. By the proof of Lemma S.3, an IF for ψ could be written as

$$\phi = \sum_{k:\mathbf{B}_k \in \mathbf{C}} \frac{P_\pi(\mathbf{W}_k)}{P(\mathbf{W}_k)} \left\{\mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{B}_k, \mathbf{W}_k] - \mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{W}_k]\right\}.$$

Using that $\frac{P_\pi(\mathbf{B}_j | \mathbf{W}_j)}{P(\mathbf{B}_j | \mathbf{W}_j)} = 1$ if $\mathbf{B}_j \in \mathbf{C}$, and $\frac{P_\pi(\mathbf{B}_j | \mathbf{W}_j)}{P(\mathbf{B}_j | \mathbf{W}_j)} = \frac{I_{\mathbf{b}_j}(\mathbf{B}_j)}{P(\mathbf{B}_j | \mathbf{W}_j)}$ if $\mathbf{B}_j \in \mathbf{X}$, we rewrite the IF as

$$\begin{aligned}
\phi &= \sum_{i:\mathbf{B}_i \in \mathbf{C}} \frac{P_\pi(\mathbf{W}_i)}{P(\mathbf{W}_i)} \left\{\mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{B}_i, \mathbf{W}_i] - \mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{W}_i]\right\} \\
&= \sum_{r:\mathbf{B}_r \in \mathbf{C}_0} \left\{\mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{B}_r, \mathbf{W}_r] - \mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{W}_r]\right\} + \sum_{\substack{k=1 \\ \mathbf{C}_k \neq \emptyset}}^{m} \sum_{r:\mathbf{B}_r \in \mathbf{C}_k} \frac{P_\pi(\mathbf{W}_r)}{P(\mathbf{W}_r)} \left\{\mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{B}_r, \mathbf{W}_r] - \mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{W}_r]\right\},
\end{aligned} \tag{B.5}$$

where the second equality holds since $\frac{P_\pi(\mathbf{W}_r)}{P(\mathbf{W}_r)} = 1$ for $\mathbf{B}_r \in \mathbf{C}_0$.

We first simplify the second term of Eq. (B.5):

$$\begin{aligned}
&\sum_{\substack{k=1 \\ \mathbf{C}_k \neq \emptyset}}^{m} \sum_{r:\mathbf{B}_r \in \mathbf{C}_k} \frac{P_\pi(\mathbf{W}_r)}{P(\mathbf{W}_r)} \left\{\mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{B}_r, \mathbf{W}_r] - \mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{W}_r]\right\} \\
&= \sum_{\substack{k=1 \\ \mathbf{C}_k \neq \emptyset}}^{m} \sum_{r:\mathbf{B}_r \in \mathbf{C}_k} \frac{\prod_{\mathbf{B}_a \in \text{pre}_{\mathbf{T}}(\mathbf{B}_r) \setminus \mathbf{X}} P_\pi(\mathbf{B}_a | \text{pre}_{\mathbf{T}}(\mathbf{B}_a)) \prod_{\mathbf{B}_c \in \text{pre}_T(\mathbf{B}_r) \cap \mathbf{X}} I_{\mathbf{b}_c}(\mathbf{B}_c)}{\prod_{\mathbf{B}_a \in \text{pre}_{\mathbf{T}}(\mathbf{B}_r) \setminus \mathbf{X}} P(\mathbf{B}_a | \text{pre}_{\mathbf{T}}(\mathbf{B}_a)) \prod_{\mathbf{B}_c \in \text{pre}_T(\mathbf{B}_r) \cap \mathbf{X}} P(\mathbf{B}_c | \text{pre}_{\mathbf{T}}(\mathbf{B}_c))} \left\{\mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{B}_r, \mathbf{W}_r] - \mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{W}_r]\right\} \\
&= \sum_{\substack{k=1 \\ \mathbf{C}_k \neq \emptyset}}^{m} \sum_{r:\mathbf{B}_r \in \mathbf{C}_k} \left(\prod_{\mathbf{B}_c \in \text{pre}_T(\mathbf{B}_r) \cap \mathbf{X}} \frac{I_{\mathbf{b}_c}(\mathbf{B}_c)}{P(\mathbf{B}_c | \text{pre}_{\mathbf{T}}(\mathbf{B}_c))} \right) \left\{\mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{B}_r, \mathbf{W}_r] - \mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{W}_r]\right\} \\
&= \sum_{\substack{k=1 \\ \mathbf{C}_k \neq \emptyset}}^{m} \left(\prod_{j_p = \{1, \cdots, j_k\}} \frac{I_{\mathbf{b}_{j_p}}(\mathbf{B}_{j_p})}{P(\mathbf{B}_{j_p} | \mathbf{W}_{j_p})} \right) \sum_{r | \mathbf{B}_r \in \mathbf{C}_k} \left(\left\{\mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{B}_r, \mathbf{W}_r] - \mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{W}_r]\right\}\right) \\
&= \sum_{\substack{k=1 \\ \mathbf{C}_k \neq \emptyset}}^{m} \underbrace{\left(\prod_{p=1}^{k} \frac{I_{\mathbf{b}_{j_p}}(\mathbf{B}_{j_p})}{P(\mathbf{B}_{j_p} | \mathbf{W}_{j_p})} \right)}_{=\omega_k} \sum_{r | \mathbf{B}_r \in \mathbf{C}_k} \left(\left\{\mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{B}_r, \mathbf{W}_r] - \mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{W}_r]\right\}\right) \\
&= \sum_{\substack{k=1 \\ \mathbf{C}_k \neq \emptyset}}^{m} \omega_k \underbrace{\left\{\mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{B}_{k_{\max}}, \mathbf{W}_{k_{\max}}] - \mathbb{E}_\pi[I_{\mathbf{y}}(\mathbf{Y}) | \mathbf{W}_{k_{\min}}]\right\}}_{=\theta_{k,1} - \theta_{k,2}} \\
&= \sum_{\substack{k=1 \\ \mathbf{C}_k \neq \emptyset}}^{m} \omega_k (\theta_{k,1} - \theta_{k,2}).
\end{aligned}$$

Then, we consider the first term of Eq. (B.5).

$$\sum_{r:\mathbf{B}_r\in\mathbf{C}_0}\{\mathbb{E}_\pi[I_\mathbf{y}(\mathbf{Y})|\mathbf{B}_r,\mathbf{W}_r]-\mathbb{E}_\pi[I_\mathbf{y}(\mathbf{Y})|\mathbf{W}_r]\}=\underbrace{\mathbb{E}_\pi[I_\mathbf{y}(\mathbf{Y})|\mathbf{B}_{0_{\max}},\mathbf{W}_{0_{\max}}]}_{=\theta_{0,1}}-\underbrace{\mathbb{E}_\pi[I_\mathbf{y}(\mathbf{Y})|\mathbf{W}_{0_{\min}}]}_{=\theta_{0,2}} \tag{B.6}$$

$$=\theta_{0,1}-\psi, \tag{B.7}$$

where the second equality holds since $\theta_{0,2}=\mathbb{E}_\pi[I_\mathbf{y}(\mathbf{Y})|\mathbf{W}_{0_{\min}}]=\mathbb{E}_\pi[I_\mathbf{y}(\mathbf{Y})]=\psi$. That is,

$$\phi=\theta_{0,1}-\psi+\sum_{\substack{k=1\\ \mathbf{C}_k\neq\emptyset}}^{m}\omega_k(\theta_{k,1}-\theta_{k,2}). \tag{B.8}$$

Then, the UIF is given as Eq. (B.4). This completes the proof. □

Lemma S.4. *Let G be a PAG over $\mathbf{V}$ and $\mathbf{T}$ be the union of a set of buckets in G. Let $\mathbf{X}$ be a bucket in $\mathbf{T}$. Let $\mathbf{S}_\mathbf{X}\equiv\bigcup_{X_i\in\mathbf{X}}\mathbf{S}_{X_i}$ where $\mathbf{S}_{X_i}$ denote the DC-component of X_i in $G(\mathbf{T})$. Then, for a bucket $\mathbf{B}_i$ in $\mathbf{T}$, $\mathbf{B}_i\not\subseteq\mathbf{S}_\mathbf{X}$ if and only if $\mathbf{B}_i\subseteq\mathbf{T}\backslash\mathbf{S}_\mathbf{X}$.*

Proof. We will consider $G=G(\mathbf{T})$. We first prove $\mathbf{B}_i\not\subseteq\mathbf{S}_\mathbf{X}\implies\mathbf{B}_i\subseteq\mathbf{T}\backslash\mathbf{S}_\mathbf{X}$. $\mathbf{B}_i\not\subseteq\mathbf{S}_\mathbf{X}$ means that there exists a node $V_k\in\mathbf{B}_i$ such that $V_k\notin\mathbf{S}_\mathbf{X}$. As a contradictory claim, suppose $\mathbf{B}_i\not\subseteq\mathbf{S}_\mathbf{X}\not\Longrightarrow\mathbf{B}_i\subseteq\mathbf{V}\backslash\mathbf{S}_\mathbf{X}$. That is, there exists $V_j\in\mathbf{B}_i$ such that $V_j\neq V_k$ and $V_j\in\mathbf{S}_{X_p}$ for some $X_p\in\mathbf{X}$. If $\left|\mathbf{S}_{X_p}\right|>1$, then there exists a node $C\in\mathbf{S}_{X_p}$ such that $C\leftrightarrow V_j$. Since V_j and V_k are in the same bucket, there is a circle path (a path composing $\circ\!\!-\!\!\circ$) between V_j and V_k. By (Zhang, 2006, Lemma 3.3.2), this implies that $C\leftrightarrow V_k$ (i.e., $V_k\in\mathbf{S}_{X_p}$), which contradicts that $V_k\notin\mathbf{S}_\mathbf{X}$. If $\left|\mathbf{S}_{X_p}\right|=1$ (i.e., $X_p=V_j$), then this implies that $\mathbf{X}=\mathbf{B}_i$, since both $\mathbf{X}$ and $\mathbf{B}_i$ are a bucket. Then this contradicts with $\mathbf{B}_i\not\subseteq\mathbf{S}_\mathbf{X}$. Therefore, there exists no $V_j\in\mathbf{B}_i$ such that $V_j\in\mathbf{S}_{X_p}$, when there exists $V_k\in\mathbf{B}_i$ such that $V_k\notin\mathbf{S}_\mathbf{X}$.

We now prove $\mathbf{B}_i\subseteq\mathbf{S}_\mathbf{X}\implies\mathbf{B}_i\not\subseteq\mathbf{T}\backslash\mathbf{S}_\mathbf{X}$. This is immediate, since $\mathbf{S}_\mathbf{X}\cap(\mathbf{T}\backslash\mathbf{S}_\mathbf{X})=\emptyset$. This completes the proof. □

Corollary B.1 (Restated Coro. 1). *Let a PTO in PAG G over $\mathbf{V}$ be $\mathbf{B}_1\prec\cdots\prec\mathbf{B}_m$. Let $\mathbf{X},\mathbf{Y}\subset\mathbf{V}$ with $\mathbf{X}$ being a bucket. Then, if $\mathcal{C}(\mathbf{X})\cap Ch(\mathbf{X})\subseteq\mathbf{X}$, $P_\mathbf{x}(\mathbf{y})$ is identifiable and given by*

$$P_\mathbf{x}(\mathbf{y})=\sum_{\mathbf{v}\backslash(\mathbf{x}\cup\mathbf{y})}\mathcal{Q}_{\mathbf{V}\backslash\mathbf{S}_\mathbf{X}}\times\mathcal{Q}_{\mathbf{S}_\mathbf{X}\backslash\mathbf{X}}, \tag{B.9}$$

where $\mathcal{Q}_{\mathbf{V}\backslash\mathbf{S}_\mathbf{X}}\equiv\prod_{\mathbf{B}_i\subseteq\mathbf{V}\backslash\mathbf{S}_\mathbf{X}}P(\mathbf{b}_i|pre(\mathbf{b}_i))$, $\mathcal{Q}_{\mathbf{S}_\mathbf{X}\backslash\mathbf{X}}\equiv\sum_\mathbf{x}\prod_{\mathbf{B}_i\subseteq\mathbf{S}_\mathbf{X}}P(\mathbf{b}_i|pre(\mathbf{b}_i))$, and $\mathbf{S}_\mathbf{X}=\bigcup_{X\in\mathbf{X}}\mathbf{S}_X$ with $\mathbf{S}_X$ being the DC-component of X.

Proof. By Prop. 1 with $\mathbf{T}=\mathbf{V}$, it suffices to show that $\frac{P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{t})}{\prod_{\mathbf{B}_i\subseteq\mathbf{S}_\mathbf{X}}P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_i|\text{pre}_\mathbf{t}(\mathbf{b}_i))}=\prod_{\mathbf{B}_i\subseteq\mathbf{T}\backslash\mathbf{S}_\mathbf{X}}P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_i|\text{pre}_\mathbf{T}(\mathbf{b}_i))$, which is equivalent to show $\prod_{\mathbf{B}_i\not\subseteq\mathbf{S}_\mathbf{X}}P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_i|\text{pre}_\mathbf{T}(\mathbf{b}_i))=\prod_{\mathbf{B}_i\subseteq\mathbf{T}\backslash\mathbf{S}_\mathbf{X}}P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_i|\text{pre}_\mathbf{T}(\mathbf{b}_i))$. To witness the equality, a sufficient condition is that $\mathbf{B}_i\not\subseteq\mathbf{S}_\mathbf{X}\Leftrightarrow\mathbf{B}_i\subseteq\mathbf{T}\backslash\mathbf{S}_\mathbf{X}$. This holds by Lemma S.4. □

Definition B.1 (Restated Def. 4). Let $\mathcal{Q}_1$ and $\mathcal{Q}_2$ be two CE-1s, then the quantity $\mathcal{Q}=\sum_\mathbf{z}(\mathcal{Q}_1\times\mathcal{Q}_2)$ is said to be (in the form of) a *canonical expression 2 (CE-2)*.

Lemma B.2 (Restated Lemma 2). *Let a target estimand $\psi=\mathcal{Q}$ be a CE-2 given in Def. 4. Let $\mathcal{V}_i$ be a UIF for the CE-1 $\mathcal{Q}_i$ given in Lemma 1 and $\mu_i\equiv\mathbb{E}_P[\mathcal{V}_i]$ for $i\in\{1,2\}$. Then, $\mathcal{V}(\mathbf{V};\eta)$ below is a UIF for ψ:*

$$\mathcal{V}(\mathbf{V};\eta)=\sum_\mathbf{z}(\mathcal{V}_1\mu_2+(\mathcal{V}_2-\mu_2)\mu_1).$$

Proof. Let $\mathcal{V}_i,\phi_i$ denote a UIF and an IF for $\mathcal{Q}_i=\mathcal{Q}_i(P)$ for $i\in\{1,2\}$. Let $\Psi(P_\gamma)\equiv\sum_\mathbf{z}(\mathcal{Q}_1(\gamma)\times\mathcal{Q}_2(\gamma))$ where $\mathcal{Q}_i(\gamma)$

are $\mathcal{Q}_i(P)$ with respect to P_γ (i.e., written w.r.t. $P_\gamma(\mathbf{b}_i|\text{pre}_{\mathbf{T}}(\mathbf{b}_i))$) for $i = 1, 2$. Note $\mathcal{Q}_i = \mathcal{Q}_i(\gamma = 0) = \mathcal{Q}_i(P)$. Then,

$$\begin{aligned}
\nabla_\gamma \Psi(P_\gamma) &\equiv \sum_{\mathbf{z}} \{(\nabla_\gamma(\mathcal{Q}_1(\gamma))\mathcal{Q}_2 + \mathcal{Q}_1 \nabla_\gamma(\mathcal{Q}_2(\gamma))\} \\
&= \sum_{\mathbf{z}} \{\mathbb{E}_P[\phi_1 \cdot S(\mathbf{V})]\,\mathcal{Q}_2 + \mathbb{E}_P[\phi_2 \cdot S(\mathbf{V})]\,\mathcal{Q}_1\} \\
&= \mathbb{E}_P\left[(\sum_{\mathbf{z}}(\phi_1\mathcal{Q}_2 + \phi_2\mathcal{Q}_1)) \cdot S(\mathbf{V})\right] \\
&= \mathbb{E}_P\left[(\sum_{\mathbf{z}}((\mathcal{V}_1 - \mathcal{Q}_1)\mathcal{Q}_2 + (\mathcal{V}_2 - \mathcal{Q}_2)\mathcal{Q}_1)) \cdot S(\mathbf{V})\right] \\
&= \mathbb{E}_P\left[\left(\sum_{\mathbf{z}}(\mathcal{V}_1\mathcal{Q}_2 + (\mathcal{V}_2 - \mathcal{Q}_2)\mathcal{Q}_1) - \psi\right) \cdot S(\mathbf{V})\right]
\end{aligned}$$

where the second equality holds by the definition of an IF, the fourth by the definition of a UIF. This implies that $\sum_{\mathbf{z}}(\mathcal{V}_1\mathcal{Q}_2 + (\mathcal{V}_2 - \mathcal{Q}_2)\mathcal{Q}_1) - \psi$ is an IF, and $\sum_{\mathbf{z}}(\mathcal{V}_1\mathcal{Q}_2 + (\mathcal{V}_2 - \mathcal{Q}_2)\mathcal{Q}_1)$ is a UIF. Since $\phi_2 = \mathcal{V}_2 - \mathcal{Q}_2 = \mathcal{V}_2 - \mu_2$ and $\mathcal{Q}_i = \mathbb{E}_P[\mathcal{V}_i]$ for $i = 1, 2$, this completes the proof. □

B.2. Proofs for Sec. 4

Lemma B.3 (Restated Lemma 3). *Let G be a PAG over $\mathbf{V}$, $\mathbf{T} = \cup_{i=1}^m \mathbf{B}_i$ be the union of a set of buckets, and $\mathbf{X} \subseteq \mathbf{T}$ be a bucket. Given $Q[\mathbf{T}]$ and a PTO $\mathbf{B}_1 \prec \cdots \prec \mathbf{B}_m$ with respect to $G(\mathbf{T})$, $Q[\mathbf{T}\backslash\mathbf{X}]$ is identifiable if and only if $\mathcal{C}(\mathbf{X}) \cap Ch(\mathbf{X}) \subseteq \mathbf{X}$ in $G(\mathbf{T})$. When $Q[\mathbf{T}\backslash\mathbf{X}]$ is identifiable, letting $\mathbf{S}_\mathbf{X} = \bigcup_{X\in\mathbf{X}} \mathbf{S}_X$ with $\mathbf{S}_X$ being the DC-component of X in $G(\mathbf{T})$, then $\mathbf{S}_\mathbf{X}$ consists of a union of buckets. Denoting $\mathbf{S}_\mathbf{X} = \{\mathbf{B}_{j_1}, \cdots, \mathbf{B}_{j_p}\}$ and $\mathbf{T}\backslash\mathbf{S}_\mathbf{X} = \{\mathbf{B}_{i_1}, \cdots, \mathbf{B}_{i_q}\}$, $Q[\mathbf{T}\backslash\mathbf{X}]$ is given by*

$$Q[\mathbf{T}\backslash\mathbf{X}] = \mathcal{Q}_{\mathbf{T}\backslash\mathbf{S}_\mathbf{X}} \times \mathcal{Q}_{\mathbf{S}_\mathbf{X}\backslash\mathbf{X}}, \tag{B.10}$$

where $\mathcal{Q}_{\mathbf{T}\backslash\mathbf{S}_\mathbf{X}} \equiv \prod_{\mathbf{B}_{i_r}\in\mathbf{T}\backslash\mathbf{S}_\mathbf{X}} P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_{i_r}|pre_{\mathbf{T}}(\mathbf{b}_{i_r}))$, and $\mathcal{Q}_{\mathbf{S}_\mathbf{X}\backslash\mathbf{X}} \equiv \sum_{\mathbf{x}} \prod_{\mathbf{B}_{j_s}\in\mathbf{S}_\mathbf{X}} P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_{j_s}|pre_{\mathbf{T}}(\mathbf{b}_{j_s}))$.

Proof. By Prop. 1 and $\frac{P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{t})}{\prod_{\mathbf{B}_i\subseteq\mathbf{S}_\mathbf{X}} P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_i|\text{pre}_{\mathbf{t}}(\mathbf{b}_i))} = \prod_{\mathbf{B}_i\subseteq\mathbf{T}\backslash\mathbf{S}_\mathbf{X}} P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_i|\text{pre}_{\mathbf{T}}(\mathbf{b}_i))$ which has been shown in the proof of Coro. 1, we have

$$Q[\mathbf{T}\backslash\mathbf{X}] = \mathcal{Q}'_{\mathbf{T}\backslash\mathbf{S}_\mathbf{X}} \times \mathcal{Q}'_{\mathbf{S}_\mathbf{X}\backslash\mathbf{X}},$$

where $\mathcal{Q}'_{\mathbf{T}\backslash\mathbf{S}_\mathbf{X}} \equiv \prod_{\mathbf{B}_{i_r}\subseteq\mathbf{T}\backslash\mathbf{S}_\mathbf{X}} P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_{i_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{i_r}))$, and $\mathcal{Q}'_{\mathbf{S}_\mathbf{X}\backslash\mathbf{X}} \equiv \sum_{\mathbf{x}} \prod_{\mathbf{B}_{j_s}\subseteq\mathbf{S}_\mathbf{X}} P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_{j_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{j_s}))$. Then, it suffices to show that $\mathbf{S}_\mathbf{X}$ is a union of buckets. Suppose $\mathbf{S}_\mathbf{X}$ is not a union of buckets. That is, there is some $\mathbf{B}_k \in \mathbf{T}$ and two variables $(V_1, V_2) \in \mathbf{B}_k$ such that $V_1 \in \mathbf{S}_\mathbf{X}$ but $V_2 \in \mathbf{T}\backslash\mathbf{S}_\mathbf{X}$. But this case doesn't exist by Lemma S.4. This shows that $\mathbf{S}_\mathbf{X}$ is a union of buckets, and completes the proof. □

Lemma S.5. *Let $\mathbf{T} = \{\mathbf{B}_1 < \cdots < \mathbf{B}_m\} \subseteq \mathbf{V}$. Let $\mathbf{C} \equiv \{\mathbf{B}_{c_1}, \cdots, \mathbf{B}_{c_n}\}$ where $\mathbf{B}_{c_r} \in \mathbf{T}$. Let $\Psi(P) \equiv \prod_{r=1}^n P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_{c_r}|pre_{\mathbf{T}}(\mathbf{b}_{c_r}))$. Let $\mathcal{V}_{Q[\mathbf{T}]}, \phi_{Q[\mathbf{T}]}$ denote a UIF and an IF for $Q[\mathbf{T}]$. Let $\mu_{Q[\mathbf{T}]} \equiv \mathbb{E}_P[\mathcal{V}_{Q[\mathbf{T}]}]$. For $\mathbf{B}_{c_r} \in \mathbf{C}$, let $\mathbf{T}_r^1 \equiv \mathbf{T}\backslash\{\mathbf{B}_{c_r}, pre_{\mathbf{T}}(\mathbf{B}_{c_r})\}$ and $\mathbf{T}_r^2 \equiv \mathbf{T}\backslash pre_{\mathbf{T}}(\mathbf{B}_{c_r})$. Then, an IF ϕ and a UIF $\mathcal{V}$ for $\Psi(P)$ is given as follows:*

$$\phi = \sum_{r=1}^n \left(\prod_{\substack{s=1\\s\neq r}}^n \mu_{P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_{c_s}|pre_{\mathbf{T}}(\mathbf{b}_{c_s}))}\right) \phi_{P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_{c_r}|pre_{\mathbf{T}}(\mathbf{b}_{c_r}))},$$

$$\mathcal{V} = \left(\prod_{s=2}^n \mu_{P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_{c_s}|pre_{\mathbf{T}}(\mathbf{b}_{c_s}))}\right) \mathcal{V}_{P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_{c_1}|pre_{\mathbf{T}}(\mathbf{b}_{c_1}))} + \sum_{r=2}^n \left(\prod_{\substack{s=1\\s\neq r}}^n \mu_{P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_{c_s}|pre_{\mathbf{T}}(\mathbf{b}_{c_s}))}\right) \phi_{P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_{c_r}|pre_{\mathbf{T}}(\mathbf{b}_{c_r}))},$$

where

$$\phi_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|pre_{\mathbf{T}}(\mathbf{b}_{c_r}))} = \frac{1}{\sum_{\mathbf{t}^2_{c_r}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^1_{c_r}}\phi_{Q[\mathbf{T}]} - \frac{\sum_{\mathbf{t}^1_{c_r}}\mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{c_r}}\mu_{Q[\mathbf{T}]}}\frac{1}{\sum_{\mathbf{t}^2_{c_r}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^2_{c_r}}\phi_{Q[\mathbf{T}]}$$

$$\mathcal{V}_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|pre_{\mathbf{T}}(\mathbf{b}_{c_r}))} = \frac{1}{\sum_{\mathbf{t}^2_{c_r}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^1_{c_r}}\mathcal{V}_{Q[\mathbf{T}]} - \frac{\sum_{\mathbf{t}^1_{c_r}}\mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{c_r}}\mu_{Q[\mathbf{T}]}}\frac{1}{\sum_{\mathbf{t}^2_{c_r}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^2_{c_r}}\phi_{Q[\mathbf{T}]},$$

and $\mu_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|pre_{\mathbf{T}}(\mathbf{b}_{c_r}))} \equiv \mathbb{E}_P[\mathcal{V}_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|pre_{\mathbf{T}}(\mathbf{b}_{c_r}))}]$.

Proof. Let $\mathcal{V}_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))}, \phi_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))}$ for $\mathbf{B}_{c_r} \in \mathbf{C}$ denote an IF and a UIF for $P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))$. Let $\mu_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))} \equiv \mathbb{E}_P[\mathcal{V}_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))}]$. We derive an IF for $\Psi(P)$, denoted ϕ, by taking a derivative of $\Psi(P_\gamma)$:

$$\begin{aligned}
\nabla_\gamma \Psi(P_\gamma) &= \nabla_\gamma \prod_{r=1}^{n} P_{\gamma,\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r})) \\
&= \sum_{r=1}^{n}\left(\nabla_\gamma P_{\gamma,\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))\right)\prod_{\substack{s=1\\s\neq r}}^{n} P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_s})) \\
&= \sum_{r=1}^{n}\prod_{\substack{s=1\\s\neq r}}^{n} P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_s}))\mathbb{E}_P[\phi_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))}\cdot S(\mathbf{V})] \\
&= \mathbb{E}_P\left[\left\{\sum_{r=1}^{n}\left(\prod_{\substack{s=1\\s\neq r}}^{n}\mu_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_s}))}\right)\phi_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))}\right\}\cdot S(\mathbf{V})\right],
\end{aligned}$$

implying that an IF ϕ is given as

$$\phi = \sum_{r=1}^{n}\left(\prod_{\substack{s=1\\s\neq r}}^{n}\mu_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_s}))}\right)\phi_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))}.$$

We derive the UIF by rewriting the IF:

$$\begin{aligned}
\phi &= \sum_{r=2}^{n}\left(\prod_{\substack{s=1\\s\neq r}}^{n}\mu_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_s}))}\right)\phi_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))} + \left(\prod_{s=2}^{n}\mu_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_s}))}\right)\left(\mathcal{V}_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_1}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_1}))} - \mu_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_1}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_1}))}\right) \\
&= \sum_{r=2}^{n}\left(\prod_{\substack{s=1\\s\neq r}}^{n}\mu_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_s}))}\right)\phi_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))} + \left(\prod_{s=2}^{n}\mu_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_s}))}\right)\mathcal{V}_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_1}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_1}))} - \Psi(P),
\end{aligned}$$

since $\prod_{s=1}^{n}\mu_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_s}))} = \prod_{s=1}^{n} P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_s})) = \Psi(P)$. This implies that

$$\mathcal{V} = \left(\prod_{s=2}^{n}\mu_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_s}))}\right)\mathcal{V}_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_1}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_1}))} + \sum_{r=2}^{n}\left(\prod_{\substack{s=1\\s\neq r}}^{n}\mu_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_s}))}\right)\phi_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))}.$$

We now derive $\phi_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))}$ and $\mathcal{V}_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))}$. We derive $\phi_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))}$ by taking derivative

$\nabla_\gamma P_{\gamma,\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))$. Recall for $\mathbf{B}_{c_r} \in \mathbf{C}$, $\mathbf{T}^1_{c_r} \equiv \mathbf{T}\setminus\{\mathbf{B}_{c_r}, \text{pre}_{\mathbf{T}}(\mathbf{B}_{c_r})\}$ and $\mathbf{T}^2_{c_r} \equiv \mathbf{T}\setminus\text{pre}_{\mathbf{T}}(\mathbf{B}_{c_r})$.

$$\begin{aligned}
\nabla_\gamma P_{\gamma,\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r})) &= \nabla_\gamma \frac{\sum_{\mathbf{t}^1_{c_r}} Q[\mathbf{T}](\gamma)}{\sum_{\mathbf{t}^2_{c_r}} Q[\mathbf{T}](\gamma)} \\
&= \frac{1}{\sum_{\mathbf{t}^2_{c_r}} Q[\mathbf{T}]} \sum_{\mathbf{t}^1_{c_r}} \nabla_\gamma Q[\mathbf{T}](\gamma) - \frac{\sum_{\mathbf{t}^1_{c_r}} Q[\mathbf{T}]}{\sum_{\mathbf{t}^2_{c_r}} Q[\mathbf{T}]} \frac{1}{\sum_{\mathbf{t}^2_{c_r}} Q[\mathbf{T}]} \sum_{\mathbf{t}^2_{c_r}} \nabla_\gamma Q[\mathbf{T}](\gamma) \\
&= \frac{1}{\sum_{\mathbf{t}^2_{c_r}} Q[\mathbf{T}]} \sum_{\mathbf{t}^1_{c_r}} \mathbb{E}_P[\phi_{Q[\mathbf{T}]} \cdot S(\mathbf{V})] - \frac{\sum_{\mathbf{t}^1_{c_r}} Q[\mathbf{T}]}{\sum_{\mathbf{t}^2_{c_r}} Q[\mathbf{T}]} \frac{1}{\sum_{\mathbf{t}^2_{c_r}} Q[\mathbf{T}]} \sum_{\mathbf{t}^2_{c_r}} \mathbb{E}_P[\phi_{Q[\mathbf{T}]} \cdot S(\mathbf{V})] \\
&= \mathbb{E}_P\left[\left\{\frac{1}{\sum_{\mathbf{t}^2_{c_r}} Q[\mathbf{T}]} \sum_{\mathbf{t}^1_{c_r}} \phi_{Q[\mathbf{T}]} - \frac{\sum_{\mathbf{t}^1_{c_r}} Q[\mathbf{T}]}{\sum_{\mathbf{t}^2_{c_r}} Q[\mathbf{T}]} \frac{1}{\sum_{\mathbf{t}^2_{c_r}} Q[\mathbf{T}]} \sum_{\mathbf{t}^2_{c_r}} \phi_{Q[\mathbf{T}]}\right\} \cdot S(\mathbf{V})\right] \\
&= \mathbb{E}_P\left[\left\{\frac{1}{\sum_{\mathbf{t}^2_{c_r}} \mu_{Q[\mathbf{T}]}} \sum_{\mathbf{t}^1_{c_r}} \phi_{Q[\mathbf{T}]} - \frac{\sum_{\mathbf{t}^1_{c_r}} \mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{c_r}} \mu_{Q[\mathbf{T}]}} \frac{1}{\sum_{\mathbf{t}^2_{c_r}} \mu_{Q[\mathbf{T}]}} \sum_{\mathbf{t}^2_{c_r}} \phi_{Q[\mathbf{T}]}\right\} \cdot S(\mathbf{V})\right],
\end{aligned}$$

implying that

$$\phi_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))} = \frac{1}{\sum_{\mathbf{t}^2_{c_r}} \mu_{Q[\mathbf{T}]}} \sum_{\mathbf{t}^1_{c_r}} \phi_{Q[\mathbf{T}]} - \frac{\sum_{\mathbf{t}^1_{c_r}} \mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{c_r}} \mu_{Q[\mathbf{T}]}} \frac{1}{\sum_{\mathbf{t}^2_{c_r}} \mu_{Q[\mathbf{T}]}} \sum_{\mathbf{t}^2_{c_r}} \phi_{Q[\mathbf{T}]}.$$

By rewriting,

$$\begin{aligned}
\phi_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))} &= \frac{1}{\sum_{\mathbf{t}^2_{c_r}} \mu_{Q[\mathbf{T}]}} \sum_{\mathbf{t}^1_{c_r}} \left(\mathcal{V}_{Q[\mathbf{T}]} - \mu_{Q[\mathbf{T}]}\right) - \frac{\sum_{\mathbf{t}^1_{c_r}} \mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{c_r}} \mu_{Q[\mathbf{T}]}} \frac{1}{\sum_{\mathbf{t}^2_{c_r}} \mu_{Q[\mathbf{T}]}} \sum_{\mathbf{t}^2_{c_r}} \phi_{Q[\mathbf{T}]} \\
&= \frac{1}{\sum_{\mathbf{t}^2_{c_r}} \mu_{Q[\mathbf{T}]}} \sum_{\mathbf{t}^1_{c_r}} \mathcal{V}_{Q[\mathbf{T}]} - \frac{\sum_{\mathbf{t}^1_{c_r}} \mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{c_r}} \mu_{Q[\mathbf{T}]}} \frac{1}{\sum_{\mathbf{t}^2_{c_r}} \mu_{Q[\mathbf{T}]}} \sum_{\mathbf{t}^2_{c_r}} \phi_{Q[\mathbf{T}]} - \underbrace{\frac{\sum_{\mathbf{t}^1_{c_r}} \mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{c_r}} \mu_{Q[\mathbf{T}]}}}_{=P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))},
\end{aligned}$$

we derive

$$\mathcal{V}_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{c_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{c_r}))} = \frac{1}{\sum_{\mathbf{t}^2_{c_r}} \mu_{Q[\mathbf{T}]}} \sum_{\mathbf{t}^1_{c_r}} \mathcal{V}_{Q[\mathbf{T}]} - \frac{\sum_{\mathbf{t}^1_{c_r}} \mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{c_r}} \mu_{Q[\mathbf{T}]}} \frac{1}{\sum_{\mathbf{t}^2_{c_r}} \mu_{Q[\mathbf{T}]}} \sum_{\mathbf{t}^2_{c_r}} \phi_{Q[\mathbf{T}]}.$$

□

Lemma B.4 (Restated Lemma 4). *Suppose $\psi \equiv Q[\mathbf{T}\setminus\mathbf{X}]$ is identifiable via Lemma 3 and given by Eq. (9). Then, given $\mathcal{V}_{Q[\mathbf{T}]}$, $\mathcal{V} \equiv \mathcal{V}_{Q[\mathbf{T}\setminus\mathbf{X}]}$ below is a UIF for ψ:*

$$\mathcal{V} = \mathcal{V}_{\mathbf{S}_\mathbf{X}\setminus\mathbf{X}} \mu_{\mathcal{V}_{\mathbf{T}\setminus\mathbf{S}_\mathbf{X}}} + (\mathcal{V}_{\mathbf{T}\setminus\mathbf{S}_\mathbf{X}} - \mu_{\mathcal{V}_{\mathbf{T}\setminus\mathbf{S}_\mathbf{X}}}) \mu_{\mathbf{S}_\mathbf{X}\setminus\mathbf{X}}, \tag{B.11}$$

where $(\mathcal{V}_{\mathbf{S}_\mathbf{X}\setminus\mathbf{X}}, \mathcal{V}_{\mathbf{T}\setminus\mathbf{S}_\mathbf{X}})$ are UIFs for $(\mathcal{Q}_{\mathbf{S}_\mathbf{X}\setminus\mathbf{X}}, \mathcal{Q}_{\mathbf{T}\setminus\mathbf{S}_\mathbf{X}})$ respectively, given by

$$\mathcal{V}_{\mathbf{S}_\mathbf{X}\setminus\mathbf{X}} \equiv \sum_{\mathbf{x}} (\mathcal{V}_{j_1} \prod_{k=2}^{p} \mu_{j_k} + \sum_{k=2}^{p} \phi_{j_k} \prod_{\ell=1,\ell\neq k}^{p} \mu_{j_\ell}), \tag{B.12}$$

$$\mathcal{V}_{\mathbf{T}\setminus\mathbf{S}_\mathbf{X}} \equiv \mathcal{V}_{i_1} \prod_{r=2}^{q} \mu_{i_r} + \sum_{r=2}^{q} \phi_{i_r} \prod_{\ell=1,\ell\neq r}^{q} \mu_{i_\ell}, \tag{B.13}$$

where, for $c \in \{1, 2, \cdots, m\}$, $\mathcal{V}_c \equiv \frac{\sum_{\mathbf{t}\setminus\{\mathbf{b}_c, pre_{\mathbf{T}}(\mathbf{b}_c)\}} \mathcal{V}_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}\setminus pre_{\mathbf{T}}(\mathbf{b}_c)} \mu_{Q[\mathbf{T}]}} - \frac{\sum_{\mathbf{t}\setminus\{\mathbf{b}_c, pre_{\mathbf{T}}(\mathbf{b}_c)\}} \mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}\setminus pre_{\mathbf{T}}(\mathbf{b}_c)} \mu_{Q[\mathbf{T}]}} \cdot \frac{\sum_{\mathbf{t}\setminus pre_{\mathbf{T}}(\mathbf{b}_c)} \phi_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}\setminus pre_{\mathbf{T}}(\mathbf{b}_c)} \mu_{Q[\mathbf{T}]}}$, $\mu_c \equiv \mathbb{E}_P[\mathcal{V}_c]$, and $\phi_c \equiv \mathcal{V}_c - \mu_c$.

Proof. We invoke the notation $\mathbf{S}_{\mathbf{X}} = \{\mathbf{B}_{j_1}, \cdots, \mathbf{B}_{j_p}\}$ and $\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}} = \{\mathbf{B}_{i_1}, \cdots, \mathbf{B}_{i_q}\}$ from Lemma 3. We first derive an IF for $\mathcal{Q}_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}}$. Let $\phi_{\mathbf{S}_{\mathbf{X}}}$ denote an IF for $\mathcal{Q}_{\mathbf{S}_{\mathbf{X}}}$. Since $\nabla_\gamma \mathcal{Q}_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}}(\gamma) = \sum_{\mathbf{x}} \nabla_\gamma \mathcal{Q}_{\mathbf{S}_{\mathbf{x}}}(\gamma) = \sum_{\mathbf{x}} \mathbb{E}_P[\phi_{\mathbf{S}_{\mathbf{x}}} \cdot S(\mathbf{V})] = \mathbb{E}_P[(\sum_{\mathbf{x}} \phi_{\mathbf{S}_{\mathbf{x}}}) \cdot S(\mathbf{V})]$, an IF for $\mathcal{Q}_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}}$ is given as $\phi_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} = \sum_{\mathbf{x}} \phi_{\mathbf{S}_{\mathbf{x}}}$, with its corresponding UIF $\mathcal{V}_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} = \sum_{\mathbf{x}} \mathcal{V}_{\mathbf{S}_{\mathbf{x}}}$. Since $\mathcal{Q}_{\mathbf{S}_{\mathbf{X}}} = \prod_{s=1}^{p} P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_{j_s}|\text{pre}_{\mathbf{T}}(\mathbf{B}_{j_s}))$, a UIF $\mathcal{V}_{\mathbf{S}_{\mathbf{X}}}$ is given by Lemma S.5 as

$$\mathcal{V}_{\mathbf{S}_{\mathbf{X}}} \equiv \mathcal{V}_{j_1} \prod_{k=2}^{p} \mu_{j_k} + \sum_{k=2}^{p} \phi_{j_k} \prod_{\ell=1, \ell\neq k}^{p} \mu_{j_\ell}.$$

Then an UIF for $\mathcal{Q}_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}}$ is given in Eq. (B.12).

Also, since $\mathcal{Q}_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}} = \prod_{s=1}^{q} P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_{i_s}|\text{pre}_{\mathbf{T}}(\mathbf{B}_{i_s}))$, a UIF for $\mathcal{Q}_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}}$ are given by Lemma S.5 as

$$\mathcal{V}_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}} \equiv \mathcal{V}_{i_1} \prod_{k=2}^{q} \mu_{i_k} + \sum_{k=2}^{q} \phi_{i_k} \prod_{\ell=1, \ell\neq k}^{q} \mu_{i_\ell},$$

which is equal to Eq. (B.13).

Next we derive an IF for $Q[\mathbf{T}\backslash\mathbf{X}]$. The derivative $\nabla_\gamma Q[\mathbf{T}\backslash\mathbf{X}](\gamma)$ is given as follows:

$$\begin{aligned}
\nabla_\gamma Q[\mathbf{T}\backslash\mathbf{X}](\gamma) &= \nabla_\gamma \left(\mathcal{Q}_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}}(\gamma) \cdot \mathcal{Q}_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}}(\gamma)\right) \\
&= \nabla_\gamma \mathcal{Q}_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}}(\gamma) \cdot \mathcal{Q}_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} + \nabla_\gamma \mathcal{Q}_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}}(\gamma) \mathcal{Q}_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}} \\
&= \mathbb{E}_P\left[\phi_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}} \cdot S(\mathbf{V})\right] \cdot \mathcal{Q}_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} + \mathbb{E}_P[\phi_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} \cdot S(\mathbf{V})] \mathcal{Q}_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}} \\
&= \mathbb{E}_P\left[\left\{\phi_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}} \cdot \mathcal{Q}_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} + \phi_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} \cdot \mathcal{Q}_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}}\right\} \cdot S(\mathbf{V})\right],
\end{aligned}$$

implying that an IF and a UIF for $Q[\mathbf{T}\backslash\mathbf{X}]$ is

$$\begin{aligned}
\phi_{Q[\mathbf{T}\backslash\mathbf{X}]} &= \phi_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}} \cdot \mathcal{Q}_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} + \phi_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} \cdot \mathcal{Q}_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}} \\
&= (\mathcal{V}_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}} - \mu_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}}) \cdot \mathcal{Q}_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} + \phi_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} \cdot \mathcal{Q}_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}} \\
&= (\mathcal{V}_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}} - \mu_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}}) \cdot \mathcal{Q}_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} + \mathcal{V}_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} \cdot \mathcal{Q}_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}} - Q[\mathbf{T}\backslash\mathbf{X}] \\
&= (\mathcal{V}_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}} - \mu_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}}) \cdot \mu_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} + \mathcal{V}_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} \cdot \mu_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}} - Q[\mathbf{T}\backslash\mathbf{X}] \\
\mathcal{V}_{Q[\mathbf{T}\backslash\mathbf{X}]} &= (\mathcal{V}_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}} - \mu_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}}) \cdot \mu_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} + \mathcal{V}_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} \cdot \mu_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}}
\end{aligned}$$

which completes the proof.

□

Lemma B.5 (Restated Lemma 5). *For* $\mathbf{A} \subseteq \mathbf{C} \subseteq \mathbf{V}$,

$$\mathcal{V}_{Q[\mathbf{C}]} = (a) + (b) - (c),$$

where $(a) = \frac{\mathcal{V}_{Q[\mathcal{R}_{\mathbf{A}}]} \cdot \mu_{Q[\mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}}{\mu_{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}}$, $(b) = \frac{\mu_{Q[\mathcal{R}_{\mathbf{A}}]} \cdot \phi_{Q[\mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}}{\mu_{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}}$, $(c) = \frac{\mu_{Q[\mathcal{R}_{\mathbf{A}}]} \cdot \mu_{Q[\mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}}{\mu_{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}} \cdot \frac{\phi_{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}}{\mu_{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}}$ *with* $\mathcal{R}_{(\cdot)} = \mathcal{R}^{\mathbf{C}}_{(\cdot)}$.

Proof. We have $Q[\mathbf{C}] = \frac{Q[\mathcal{R}_{\mathbf{A}}] \cdot Q[\mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}$ by Prop. 2. We derive an IF for $Q[\mathbf{C}]$ by computing the following derivative

$$\begin{aligned}
&\nabla_\gamma Q[\mathbf{C}](\gamma) \\
&= \frac{\nabla_\gamma Q[\mathcal{R}_{\mathbf{A}}](\gamma) \cdot Q[\mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]} + \frac{Q[\mathcal{R}_{\mathbf{A}}] \cdot \nabla_\gamma Q[\mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}](\gamma)}{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]} - \frac{Q[\mathcal{R}_{\mathbf{A}}] \cdot Q[\mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]} \frac{\nabla_\gamma Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}](\gamma)}{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]} \\
&= \mathbb{E}_P\left[\left\{\frac{\phi_{Q[\mathcal{R}_{\mathbf{A}}]} \cdot Q[\mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]} + \frac{Q[\mathcal{R}_{\mathbf{A}}] \cdot \phi_{Q[\mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}}{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]} - \frac{Q[\mathcal{R}_{\mathbf{A}}] \cdot Q[\mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]} \frac{\phi_{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}}{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}\right\} \cdot S(\mathbf{V})\right],
\end{aligned}$$

which implies that an IF for $Q[\mathbf{C}]$ is given by

$$\begin{aligned}
\phi_{Q[\mathbf{C}]} &= \frac{(\mathcal{V}_{Q[\mathcal{R}_{\mathbf{A}}]} - Q[\mathcal{R}_{\mathbf{A}}]) \cdot Q[\mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]} + \frac{Q[\mathcal{R}_{\mathbf{A}}] \cdot \phi_{Q[\mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}}{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]} - \frac{Q[\mathcal{R}_{\mathbf{A}}] \cdot Q[\mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]} \frac{\phi_{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}}{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]} \\
&= \frac{\mathcal{V}_{Q[\mathcal{R}_{\mathbf{A}}]} \cdot Q[\mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]} + \frac{Q[\mathcal{R}_{\mathbf{A}}] \cdot \phi_{Q[\mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}}{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]} - \frac{Q[\mathcal{R}_{\mathbf{A}}] \cdot Q[\mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]} \frac{\phi_{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]}}{Q[\mathcal{R}_{\mathbf{A}} \cap \mathcal{R}_{\mathbf{C}\backslash\mathcal{R}_{\mathbf{A}}}]} - Q[\mathbf{C}],
\end{aligned}$$

and its corresponding UIF for $Q[\mathbf{C}]$ is given by

$$\mathcal{V}_{Q[\mathbf{C}]} = \frac{\mathcal{V}_{Q[\mathcal{R}_\mathbf{A}]} \cdot Q[\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_\mathbf{A}}]}{Q[\mathcal{R}_\mathbf{A} \cap \mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_\mathbf{A}}]} + \frac{Q[\mathcal{R}_\mathbf{A}] \cdot \phi_{Q[\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_\mathbf{A}}]}}{Q[\mathcal{R}_\mathbf{A} \cap \mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_\mathbf{A}}]} - \frac{Q[\mathcal{R}_\mathbf{A}] \cdot Q[\mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_\mathbf{A}}]}{Q[\mathcal{R}_\mathbf{A} \cap \mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_\mathbf{A}}]} \frac{\phi_{Q[\mathcal{R}_\mathbf{A} \cap \mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_\mathbf{A}}]}}{Q[\mathcal{R}_\mathbf{A} \cap \mathcal{R}_{\mathbf{C}\setminus\mathcal{R}_\mathbf{A}}]}.$$

□

Lemma S.6. *Algo. 1 IFP derives a UIF for any identifiable $P_\mathbf{x}(\mathbf{y})$ in a PAG G over $\mathbf{V}$ in $O(|\mathbf{V}|^4)$ time where $|\mathbf{V}|$ denotes the number of variables.*

Proof. Line 3 takes $O(|\mathbf{V}|^2)$. We now derive the complexity for line 4, DERIVEUIF $\big(\mathbf{D}, \mathbf{V}, P(\mathbf{V}), \mathcal{V}_{Q[\mathbf{V}]} = I_\mathbf{v}(\mathbf{V})\big)$.

Let $\mathbf{T}$ be the input of DERIVEUIF. Let $N \equiv |\mathbf{T}|$. Suppose running DERIVEUIF$(\mathbf{C}, \mathbf{T}, Q[\mathbf{T}], \mathcal{V}_{Q[\mathbf{T}]})$ takes $T(N)$ time. To check the condition in line 8, for each bucket in $\mathbf{T}$ (possibly N buckets), one checks the graphical condition ($O(N^2)$). Therefore, it takes $O(N^3)$ to run line 8. Line 9-12 take $O(N^2)$ since it takes $O(N^2)$ to identify the DC-component for deriving $Q[\mathbf{T}\setminus\mathbf{B}]$. The recursive call with DERIVEUIF$(\mathbf{C}, \mathbf{T}\setminus\mathbf{B}, Q[\mathbf{T}\setminus\mathbf{B}], \mathcal{V}_{Q[\mathbf{T}\setminus\mathbf{B}]})$ takes at most $T(N-1)$. If line 15 is called, then it takes $T(N-M)+T(M)$ where $M \equiv |\mathcal{R}_\mathbf{B}|$. That is,

$$T(N) = \begin{cases} T(N-1) + O(N^3); & \text{or} \\ T(N-M) + T(M) + O(N^3). \end{cases} \tag{B.14}$$

For sufficiently large $N > C$ for some constant C, let $O(N^3) \leq a \cdot N^3$ and assume $T(N) \leq aN^4$ for some constants $a > 0$ (If this holds, this means that $T(N) = O(N^4)$). First, consider $T(N-1) + O(N^3)$. For some sufficiently large N,

$$\begin{aligned} T(N-1) + O(N^3) &\leq a(N-1)^4 + aN^3 \\ &= a(N^4 - 4N^3 + 6N^2 - 4N + 1) + aN^3 \\ &= a\left(N^4 - 3N^3 + 6N^2 - 4N + 1\right) \\ &\leq a \cdot N^4. \end{aligned}$$

$$\begin{aligned} T(N-M) + T(M) + O(N^3) &\leq a(N-M)^4 + aM^4 + aN^3 \\ &\leq a\left(N^4 - (4M-1)N^3 + 6N^2M^2 - 4NM^3 + 2M^4\right) \\ &\leq aN^4. \end{aligned}$$

This implies that $T(N) = O(N^4)$, i.e., DERIVEUIF$(\mathbf{D}, \mathbf{T}, Q[\mathbf{T}], \mathcal{V}_{Q[\mathbf{T}]})$ runs in $O(N^4)$. That is, line 4 (DERIVEUIF $\big(\mathbf{D}, \mathbf{V}, P(\mathbf{V}), \mathcal{V}_{Q[\mathbf{V}]} = I_\mathbf{v}(\mathbf{V})\big)$) runs in $O(|\mathbf{V}|^4)$. Therefore, Algo. 1 IFP runs in $O(|\mathbf{V}|^4)$. □

Theorem B.1 (Restated Thm. 1)**.** *Algo. 1 IFP derives a UIF for any identifiable $P_\mathbf{x}(\mathbf{y})$ in a PAG G over $\mathbf{V}$ in $O(|\mathbf{V}|^4)$ time where $|\mathbf{V}|$ denotes the number of variables. IFP returns* FAIL *if $P_\mathbf{x}(\mathbf{y})$ is not identifiable.*

Proof. Soundness of IFP follows from Lemmas (1 - 5) and the soundness of the IDP (Algo. A.2) (Jaber et al., 2019). Completeness follows from the completeness of IDP, since IFP fails if and only if IDP fails given that IDP and IFP share the same failure conditions. Finally, IFP runs in $O(|\mathbf{V}|^4)$ by Lemma S.6. □

B.3. Proofs for Sec. 5

Assumption. In analyzing the properties of estimators T_N, we assume *strict boundedness* for nuisances $(\omega_{j,k}, \theta_{j,k,i}) \in \eta$ and their estimates $(\widehat{\omega}_{j,k}, \widehat{\theta}_{j,k,i}) \in \widehat{\eta}$: there exist constants $M_1, M_2 > 0$ such that $M_1 < (\omega_{j,k}, \theta_{j,k,i}) < M_2$ and $M_1 < (\widehat{\omega}_{j,k}, \widehat{\theta}_{j,k,i}) < M_2$.

Lemma B.6 (Restated Lemma 6). *The UIF $\mathcal{V}_{P_\mathbf{x}(\mathbf{y})}(\mathbf{V}; \eta)$ returned by Algo. 1 IFP is an arithmetic combination (ratio, multiplication, and marginalization) of UIFs for functionals in the form of CE-1, denoted as $\mathcal{V}_{P_\mathbf{x}(\mathbf{y})}(\mathbf{V}; \eta = \{\boldsymbol{\omega}_j, \boldsymbol{\theta}_j\}_{j=1}^{\ell}) = \mathcal{A}\left(\{\mathcal{V}_j(\boldsymbol{\omega}_j, \boldsymbol{\theta}_j)\}_{j=1}^{\ell}\right)$ where $\mathcal{V}_j(\boldsymbol{\omega}_j, \boldsymbol{\theta}_j)$ denotes a UIF given by Lemma 1 with $\boldsymbol{\omega}_j = \{\omega_{j,k}\}_{k=1}^{m_j}$ and $\boldsymbol{\theta}_j = \{\theta_{j,0,1}\} \cup \{\theta_{j,k,1}, \theta_{j,k,2}\}_{k=1}^{m_j}$ being nuisances for $\mathcal{V}_j$, and $\mathcal{A}(\cdot)$ an arithmetic function.*

Proof. IFP recursively calls DERIVEUIF, initially equipped with a UIF for $P(\mathbf{v})$ which is a special case of CE-1. We show that DERIVEUIF always returns an arithmetic function of the UIFs for CE-1. Suppose $\mathcal{V}_{Q[\mathbf{T}]}$ in DERIVEUIF is given as an arithmetic function of UIFs for CE-1. Let $\mathbf{B}$ be a bucket satisfying line 8 of IFP in Algo. 1. Then, the UIF for $Q[\mathbf{T}\backslash\mathbf{B}]$, denoted $\mathcal{V}_{Q[\mathbf{T}\backslash\mathbf{B}]}$, is given either by Lemma 1 (as a UIF for CE-1), or Lemma 2 (as a function of UIFs for CE-1), or by Lemma 4 (Eq. (B.11,B.12,B.13)) as a function of $\mathcal{V}_{Q[\mathbf{T}]}$ (and $\phi_{Q[\mathbf{T}]}, \mu_{Q[\mathbf{T}]}$) which is a function of UIFs for CE-1. If Line 15 is invoked, then the output is expressed as an arithmetic function of the outputs of several DERIVEUIF calls which are functions of UIFs for CE-1.

□

Lemma S.7. *Let $P_{\mathbf{x}}(\mathbf{y})$ be identified as $P_{\mathbf{x}}(\mathbf{y}) = \psi \equiv \Psi(P)$. Let $\mathbf{D}$ be the set of variables defined in line 3 of Algo. 1. Then, the IF for ψ given by Algo. 1 is in the form of a linear combination of IFs for CE-1, denoted as*

$$\phi_{P_{\mathbf{x}}(\mathbf{y})} = \sum_{\mathbf{d}\backslash\mathbf{y}} \sum_{j=1}^{\ell} f_j \sum_{\mathbf{h}_j} \phi_j(\boldsymbol{\omega}_j, \boldsymbol{\theta}_j),$$

where $\mathbf{H}_j \subseteq \mathbf{V}$ are sets of variables, $f_j > 0$ are constants, and $\phi_j(\boldsymbol{\omega}_j, \boldsymbol{\theta}_j)$ are IFs for CE-1.

Proof. Let $\phi_{Q[\mathbf{D}]}$ denote the IF for $Q[\mathbf{D}]$. It suffices to show that $\phi_{Q[\mathbf{D}]}$ is in the form $\phi_{Q[\mathbf{D}]} = \sum_{j=1}^{\ell} f_j \sum_{\mathbf{h}_j} \phi_j(\boldsymbol{\omega}_j, \boldsymbol{\theta}_j)$ since $\phi_{P_{\mathbf{x}}(\mathbf{y})} = \sum_{\mathbf{d}\backslash\mathbf{y}} \phi_{Q[\mathbf{D}]}$. It is obvious that $\phi_{Q[\mathbf{D}]}$ is given as a linear combination of IFs for CE-1 if $Q[\mathbf{D}]$ is in CE-1 or CE-2.

First assume Line 8 in IFP is invoked. $\phi_{Q[\mathbf{T}\backslash\mathbf{X}]}$ is obviously a linear combination of IFs for CE-1 if Line 10 or 11 is invoked. Next, we focus on the case of Line 12. Let $\mathbf{T} \subseteq \mathbf{V}$ be a set of nodes defined in Lemma 3. We invoke the notation $\mathbf{S}_{\mathbf{X}} = \{\mathbf{B}_{j_1}, \cdots, \mathbf{B}_{j_p}\}$ and $\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}} = \{\mathbf{B}_{i_1}, \cdots, \mathbf{B}_{i_q}\}$ from Lemma 3. For $\mathbf{B}_{i_r} \in \mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}$, let $\mathbf{T}^1_{i_r} \equiv \mathbf{T}\backslash\{\mathbf{B}_{i_r}, \text{pre}_{\mathbf{T}}(\mathbf{B}_{i_r})\}$ and $\mathbf{T}^2_{i_r} \equiv \mathbf{T}\backslash\text{pre}_{\mathbf{T}}(\mathbf{B}_{i_r})$. For $\mathbf{B}_{j_s} \in \mathbf{S}_{\mathbf{X}}$, let $\mathbf{T}^1_{j_s} \equiv \mathbf{T}\backslash\{\mathbf{B}_{j_s}, \text{pre}_{\mathbf{T}}(\mathbf{B}_{j_s})\}$ and $\mathbf{T}^2_{j_s} \equiv \mathbf{T}\backslash\text{pre}_{\mathbf{T}}(\mathbf{B}_{j_s})$.

Let $\mathbf{X}$ be a bucket satisfying the criterion in Lemma 3. Suppose

$$\phi_{Q[\mathbf{T}]} = \sum_{j=1}^{\ell} f_j \sum_{\mathbf{h}_j} \phi_j(\boldsymbol{\omega}_j, \boldsymbol{\theta}_j).$$

Then, we will derive $\phi_{Q[\mathbf{T}\backslash\mathbf{X}]}$. We will use $\phi_{(\cdot)}$, $\mathcal{V}_{(\cdot)}$ and $\mu_{(\cdot)}$ to denote an IF, a UIF and $\mathbb{E}_P[\mathcal{V}_{(\cdot)}]$ of $\mathcal{Q}_{(\cdot)}$.

$$\phi_{Q[\mathbf{T}\backslash\mathbf{X}]} = \phi_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}} \cdot \mathcal{Q}_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} + \phi_{\mathbf{S}_{\mathbf{X}}\backslash\mathbf{X}} \cdot \mathcal{Q}_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}},$$

where by Lemma S.5,

$$\phi_{\mathbf{T}\backslash\mathbf{S}_{\mathbf{X}}} = \sum_{r=1}^{q} \left(\prod_{\substack{s=1 \\ s\neq r}}^{q} \mu_{P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_{i_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{i_s}))} \right) \phi_{P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_{i_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{i_r}))}$$

where

$$\phi_{P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_{i_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{i_s}))} = \frac{1}{\sum_{\mathbf{t}^2_{i_s}} \mu_{Q[\mathbf{T}]}} \sum_{\mathbf{t}^1_{i_s}} \phi_{Q[\mathbf{T}]} - \frac{\sum_{\mathbf{t}^1_{i_s}} \mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{i_s}} \mu_{Q[\mathbf{T}]}} \frac{1}{\sum_{\mathbf{t}^2_{i_s}} \mu_{Q[\mathbf{T}]}} \sum_{\mathbf{t}^2_{i_s}} \phi_{Q[\mathbf{T}]}$$

$$\mathcal{V}_{P_{\mathbf{v}\backslash\mathbf{t}}(\mathbf{b}_{i_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{i_s}))} = \frac{1}{\sum_{\mathbf{t}^2_{i_s}} \mu_{Q[\mathbf{T}]}} \sum_{\mathbf{t}^1_{i_s}} \mathcal{V}_{Q[\mathbf{T}]} - \frac{\sum_{\mathbf{t}^1_{i_s}} \mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{i_s}} \mu_{Q[\mathbf{T}]}} \frac{1}{\sum_{\mathbf{t}^2_{i_s}} \mu_{Q[\mathbf{T}]}} \sum_{\mathbf{t}^2_{i_s}} \phi_{Q[\mathbf{T}]},$$

and $\mu_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{i_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{i_s}))} \equiv \mathbb{E}_P[\mathcal{V}_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{i_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{i_s}))}]$. That is,

$$\begin{aligned}\phi_{\mathbf{T}\setminus\mathbf{S_X}} &= \sum_{r=1}^{q}\left(\underbrace{\prod_{\substack{s=1\\s\neq r}}^{q}\mu_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{i_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{i_s}))}}_{=c^r_{\mathbf{T}\setminus\mathbf{S_X}}}\right)\left(\frac{1}{\sum_{\mathbf{t}^2_{i_r}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^1_{i_r}}\phi_{Q[\mathbf{T}]} - \frac{\sum_{\mathbf{t}^1_{i_r}}\mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{i_r}}\mu_{Q[\mathbf{T}]}}\frac{1}{\sum_{\mathbf{t}^2_{i_r}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^2_{i_r}}\phi_{Q[\mathbf{T}]}\right)\\ &= \sum_{r=1}^{q}\frac{c^r_{\mathbf{T}\setminus\mathbf{S_X}}}{\sum_{\mathbf{t}^2_{i_r}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^2_{i_r}}\phi_{Q[\mathbf{T}]} - \sum_{r=1}^{q}\frac{\sum_{\mathbf{t}^1_{i_r}}\mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{i_r}}\mu_{Q[\mathbf{T}]}}\frac{c^r_{\mathbf{T}\setminus\mathbf{S_X}}}{\sum_{\mathbf{t}^2_{i_r}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^2_{i_r}}\phi_{Q[\mathbf{T}]}.\end{aligned}$$

Also, by Lemma S.5,

$$\phi_{\mathbf{S_X}\setminus\mathbf{X}} = \sum_{r=1}^{p}\sum_{\mathbf{x}}\left(\prod_{\substack{s=1\\s\neq r}}^{p}\mu_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{j_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{j_s}))}\right)\phi_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{j_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{j_s}))}$$

where

$$\begin{aligned}\phi_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{j_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{j_s}))} &= \frac{1}{\sum_{\mathbf{t}^2_{j_s}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^1_{j_s}}\phi_{Q[\mathbf{T}]} - \frac{\sum_{\mathbf{t}^1_{j_s}}\mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{j_s}}\mu_{Q[\mathbf{T}]}}\frac{1}{\sum_{\mathbf{t}^2_{j_s}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^2_{j_s}}\phi_{Q[\mathbf{T}]}\\ \mathcal{V}_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{j_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{j_s}))} &= \frac{1}{\sum_{\mathbf{t}^2_{j_s}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^1_{j_s}}\mathcal{V}_{Q[\mathbf{T}]} - \frac{\sum_{\mathbf{t}^1_{j_s}}\mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{j_s}}\mu_{Q[\mathbf{T}]}}\frac{1}{\sum_{\mathbf{t}^2_{j_s}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^2_{j_s}}\phi_{Q[\mathbf{T}]},\end{aligned}$$

and $\mu_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{j_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{j_s}))} \equiv \mathbb{E}_P[\mathcal{V}_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{j_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{j_s}))}]$. That is,

$$\begin{aligned}\phi_{\mathbf{S_X}\setminus\mathbf{X}} &= \sum_{r=1}^{p}\sum_{\mathbf{x}}\left(\underbrace{\prod_{\substack{s=1\\s\neq r}}^{p}\mu_{P_{\mathbf{v}\setminus\mathbf{t}}(\mathbf{b}_{j_s}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{j_s}))}}_{=c^r_{\mathbf{S_X}\setminus\mathbf{X}}}\right)\left(\frac{1}{\sum_{\mathbf{t}^2_{j_r}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^1_{j_r}}\phi_{Q[\mathbf{T}]} - \frac{\sum_{\mathbf{t}^1_{j_r}}\mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{j_r}}\mu_{Q[\mathbf{T}]}}\frac{1}{\sum_{\mathbf{t}^2_{j_r}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^2_{j_r}}\phi_{Q[\mathbf{T}]}\right)\\ &= \sum_{r=1}^{p}\sum_{\mathbf{x}}\frac{c^r_{\mathbf{S_X}\setminus\mathbf{X}}}{\sum_{\mathbf{t}^2_{j_r}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^2_{j_r}}\phi_{Q[\mathbf{T}]} - \sum_{r=1}^{p}\sum_{\mathbf{x}}\frac{\sum_{\mathbf{t}^1_{j_r}}\mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{j_r}}\mu_{Q[\mathbf{T}]}}\frac{c^r_{\mathbf{S_X}\setminus\mathbf{X}}}{\sum_{\mathbf{t}^2_{j_r}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^2_{j_r}}\phi_{Q[\mathbf{T}]}.\end{aligned}$$

Then,

$$\begin{aligned}\phi_{Q[\mathbf{T}\setminus\mathbf{X}]} &= \phi_{\mathbf{T}\setminus\mathbf{S_X}}\cdot\mathcal{Q}_{\mathbf{S_X}\setminus\mathbf{X}} + \phi_{\mathbf{S_X}\setminus\mathbf{X}}\cdot\mathcal{Q}_{\mathbf{T}\setminus\mathbf{S_X}},\\ &= \sum_{r=1}^{q}\frac{\mathcal{Q}_{\mathbf{S_X}\setminus\mathbf{X}}\cdot c^r_{\mathbf{T}\setminus\mathbf{S_X}}}{\sum_{\mathbf{t}^2_{i_r}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^1_{i_r}}\phi_{Q[\mathbf{T}]} - \sum_{r=1}^{q}\frac{\sum_{\mathbf{t}^1_{i_r}}\mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{i_r}}\mu_{Q[\mathbf{T}]}}\frac{c^r_{\mathbf{T}\setminus\mathbf{S_X}}\cdot\mathcal{Q}_{\mathbf{S_X}\setminus\mathbf{X}}}{\sum_{\mathbf{t}^2_{i_r}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^2_{i_r}}\phi_{Q[\mathbf{T}]}\\ &+ \sum_{\mathbf{x}}\sum_{r=1}^{p}\frac{c^r_{\mathbf{S_X}\setminus\mathbf{X}}\cdot\mathcal{Q}_{\mathbf{T}\setminus\mathbf{S_X}}}{\sum_{\mathbf{t}^2_{j_r}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^1_{j_r}}\phi_{Q[\mathbf{T}]} - \sum_{\mathbf{x}}\sum_{r=1}^{p}\frac{\sum_{\mathbf{t}^1_{j_r}}\mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{j_r}}\mu_{Q[\mathbf{T}]}}\frac{c^r_{\mathbf{S_X}\setminus\mathbf{X}}\cdot\mathcal{Q}_{\mathbf{T}\setminus\mathbf{S_X}}}{\sum_{\mathbf{t}^2_{j_r}}\mu_{Q[\mathbf{T}]}}\sum_{\mathbf{t}^2_{j_r}}\phi_{Q[\mathbf{T}]}\\ &= \sum_{a_1=1}^{2}\sum_{\mathbf{c}_{a_1}}\sum_{a_2=1}^{2}\sum_{r=1}^{m_{a_1}}d_{a_1a_2}\sum_{\mathbf{t}_{a_1a_2}}\phi_{Q[\mathbf{T}]},\end{aligned}$$

where $\mathbf{C}_1 = \emptyset$ and $\mathbf{C}_2 = \mathbf{X}$; $m_1 = q$ and $m_2 = p$; $d_{11} = \frac{\mathcal{Q}_{\mathbf{S_X}\setminus\mathbf{X}}\cdot c^r_{\mathbf{T}\setminus\mathbf{S_X}}}{\sum_{\mathbf{t}^2_{i_r}}\mu_{Q[\mathbf{T}]}}$, $d_{12} = -\frac{\sum_{\mathbf{t}^1_{i_r}}\mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{i_r}}\mu_{Q[\mathbf{T}]}}\frac{c^r_{\mathbf{T}\setminus\mathbf{S_X}}\cdot\mathcal{Q}_{\mathbf{S_X}\setminus\mathbf{X}}}{\sum_{\mathbf{t}^2_{i_r}}\mu_{Q[\mathbf{T}]}}$, $d_{21} = \frac{c^r_{\mathbf{S_X}\setminus\mathbf{X}}\cdot\mathcal{Q}_{\mathbf{T}\setminus\mathbf{S_X}}}{\sum_{\mathbf{t}^2_{j_r}}\mu_{Q[\mathbf{T}]}}$, and $d_{22} = \frac{\sum_{\mathbf{t}^1_{j_r}}\mu_{Q[\mathbf{T}]}}{\sum_{\mathbf{t}^2_{j_r}}\mu_{Q[\mathbf{T}]}}\frac{c^r_{\mathbf{S_X}\setminus\mathbf{X}}\cdot\mathcal{Q}_{\mathbf{T}\setminus\mathbf{S_X}}}{\sum_{\mathbf{t}^2_{j_r}}\mu_{Q[\mathbf{T}]}}$; $t_{11} = \mathbf{t}^1_{i_r}$, $t_{12} = \mathbf{t}^2_{i_r}$, $t_{21} = \mathbf{t}^1_{j_r}$, and $t_{22} = \mathbf{t}^2_{j_r}$.

Note we assume $\phi_{Q[\mathbf{T}]} = \sum_{j=1}^{\ell} f_j \sum_{\mathbf{j}_j} \phi_j(\boldsymbol{\omega}_j, \boldsymbol{\theta}_j)$. Then,

$$\begin{aligned}\phi_{Q[\mathbf{T}\setminus\mathbf{X}]} &= \sum_{a_1=1}^{2}\sum_{\mathbf{c}_{a_1}}\sum_{a_2=1}^{2}\sum_{r=1}^{m_{a_1}} d_{a_1a_2}\sum_{\mathbf{t}_{a_1a_2}}\sum_{j=1}^{\ell} f_j \sum_{\mathbf{h}_j}\phi_j(\boldsymbol{\omega}_j,\boldsymbol{\theta}_j)\\ &= \sum_{a_1=1}^{2}\sum_{a_2=1}^{2}\sum_{r=1}^{m_{a_1}}\sum_{j=1}^{\ell} d_{a_1a_2} f_j \sum_{\mathbf{c}_{a_1}}\sum_{\mathbf{t}_{a_1a_2}}\sum_{\mathbf{h}_j}\phi_j(\boldsymbol{\omega}_j,\boldsymbol{\theta}_j).\end{aligned}$$

This implies that $\phi_{Q[\mathbf{T}\setminus\mathbf{X}]}$ is in the form of a linear combination of $\phi_j(\boldsymbol{\omega}_j, \boldsymbol{\theta}_j)$.

If line 15 (or Lemma 5) is invoked, then the IF is represented as follows:

$$\begin{aligned}\phi_{Q[\mathbf{T}]} &= \frac{Q[\mathcal{R}_{\mathbf{T}\setminus\mathcal{R}_\mathbf{A}}]}{Q[\mathcal{R}_\mathbf{A}\cap\mathcal{R}_{\mathbf{T}\setminus\mathcal{R}_\mathbf{A}}]}\phi_{Q[\mathcal{R}_\mathbf{A}]} + \frac{Q[\mathcal{R}_\mathbf{A}]}{Q[\mathcal{R}_\mathbf{A}\cap\mathcal{R}_{\mathbf{T}\setminus\mathcal{R}_\mathbf{A}}]}\phi_{Q[\mathcal{R}_{\mathbf{T}\setminus\mathcal{R}_\mathbf{A}}]} - \frac{Q[\mathcal{R}_\mathbf{A}]\cdot Q[\mathcal{R}_{\mathbf{T}\setminus\mathcal{R}_\mathbf{A}}]}{Q[\mathcal{R}_\mathbf{A}\cap\mathcal{R}_{\mathbf{T}\setminus\mathcal{R}_\mathbf{A}}]}\frac{1}{Q[\mathcal{R}_\mathbf{A}\cap\mathcal{R}_{\mathbf{T}\setminus\mathcal{R}_\mathbf{A}}]}\phi_{Q[\mathcal{R}_\mathbf{A}\cap\mathcal{R}_{\mathbf{T}\setminus\mathcal{R}_\mathbf{A}}]},\\ &= \sum_{b_1=1}^{3} c_{b_1}\phi_{Q[\mathbf{T}_{b_1}]},\end{aligned}$$

where $c_1 = \frac{Q[\mathcal{R}_{\mathbf{T}\setminus\mathcal{R}_\mathbf{A}}]}{Q[\mathcal{R}_\mathbf{A}\cap\mathcal{R}_{\mathbf{T}\setminus\mathcal{R}_\mathbf{A}}]}$, $c_2 = \frac{Q[\mathcal{R}_\mathbf{A}]}{Q[\mathcal{R}_\mathbf{A}\cap\mathcal{R}_{\mathbf{T}\setminus\mathcal{R}_\mathbf{A}}]}$ and $c_3 = -\frac{Q[\mathcal{R}_\mathbf{A}]\cdot Q[\mathcal{R}_{\mathbf{T}\setminus\mathcal{R}_\mathbf{A}}]}{Q[\mathcal{R}_\mathbf{A}\cap\mathcal{R}_{\mathbf{T}\setminus\mathcal{R}_\mathbf{A}}]}\frac{1}{Q[\mathcal{R}_\mathbf{A}\cap\mathcal{R}_{\mathbf{T}\setminus\mathcal{R}_\mathbf{A}}]}$; and $\mathbf{T}_1 = \mathcal{R}_\mathbf{A}$, $\mathbf{T}_2 = \mathcal{R}_{\mathbf{T}\setminus\mathcal{R}_\mathbf{A}}$, and $\mathbf{T}_3 = \mathcal{R}_\mathbf{A}\cap\mathcal{R}_{\mathbf{T}\setminus\mathcal{R}_\mathbf{A}}$. Therefore $\phi_{Q[\mathbf{T}]}$ will be a linear combination of IFs for CE-1 whenever $\phi_{\mathbf{T}_{b_1}}$ are written as a linear combination of IFs for CE-1 (i.e., $\sum_{j=1}^{\ell} f_j \sum_{\mathbf{j}_j}\phi_j(\boldsymbol{\omega}_j,\boldsymbol{\theta}_j)$).

We conclude that $\phi_{Q[\mathbf{D}]}$ is always a linear combination of IFs for CE-1. This completes the proof. □

Lemma S.8. *The IF given in Lemma 1 for CE-1 is a Neyman orthogonal score with respect to $\eta = (\boldsymbol{\omega}, \boldsymbol{\theta})$ defined in Lemma 1.*

Proof. We recall that an IF for CE-1 is given as follows, by Eq. (B.8):

$$\phi(\mathbf{V}; \eta = \{\boldsymbol{\omega}, \boldsymbol{\theta}\}, \psi) = \theta_{0,1} - \psi + \sum_{\substack{k=0\\ \mathbf{C}_k\neq\emptyset}}^{m} \omega_k\left(\theta_{k,1} - \theta_{k,2}\right).$$

Let $\{i_1, i_2, \cdots, i_p\} = \{k \in \{1, \cdots, m\} | \mathbf{C}_k \neq \emptyset\}$. Then, we can rewrite the IF as

$$\phi(\mathbf{V}; \eta = \{\boldsymbol{\omega}, \boldsymbol{\theta}\}, \psi) = \theta_{0,1} - \psi + \sum_{r=1}^{p} \omega_{i_r}\left(\theta_{i_r,1} - \theta_{i_r,2}\right).$$

We rewrite the set of nuisances as $\boldsymbol{\omega} = \{\omega_{i_r}\}_{r=1}^{p}$ and $\boldsymbol{\theta} = \{\theta_{0,1}\} \cup \{(\theta_{i_r,1}, \theta_{i_r,2})\}_{r=1}^{p}$. For any nuisance $\eta_s \in \boldsymbol{\omega}$ or $\eta_s \in \boldsymbol{\theta}$, we will use η_s^* with an asterisk $(*)$ mark to denote the true nuisance. Let $\boldsymbol{\omega}^* = \{\omega_{i_r}^*\}_{r=1}^{p}$, $\boldsymbol{\theta}^* = \{(\theta_{i_r,1}^*, \theta_{i_r,2}^*)\}_{r=1}^{p}$, and $\eta^* = (\boldsymbol{\omega}^*, \boldsymbol{\theta}^*)$ be a set of true nuisances.

We recall that ϕ is a Neyman orthogonal score if it satisfies (1) $\mathbb{E}_P[\phi(\mathbf{V}; \psi, \eta^*)] = 0$ and (2) $\frac{\partial}{\partial\eta}\mathbb{E}_P[\phi(\mathbf{V}; \psi, \eta)]|_{\eta=\eta^*} = 0$. To check the first condition, we first note that

$$\mathbb{E}_P[\theta_{0,1}^*] - \psi = \mathbb{E}_{P(\mathbf{B}_{0_{\max}})}[\mathbb{E}_{P_\pi}[I_\mathbf{y}(\mathbf{Y})|\mathbf{B}_{0_{\max}}]] - \psi = \mathbb{E}_{P_\pi}[I_\mathbf{y}(\mathbf{Y})] - \psi = \psi - \psi = 0,$$

where the third equality holds since $\mathbf{B}_{0_{\max}} \not\subseteq \mathbf{X}$, and $P(\mathbf{B}_{0_{\max}}) = P_\pi(\mathbf{B}_{0_{\max}})$. Also, for any $r \in \{1, 2, \cdots, p\}$,

$$\begin{aligned}\mathbb{E}_P[\omega_{i_r}^*\left(\theta_{i_r,1}^* - \theta_{i_r,2}^*\right)] &= \mathbb{E}_{P(\text{pre}_\mathbf{T}(\mathbf{B}_{i_r,\min}))}\left[\mathbb{E}_P\left\{\omega_{i_r}^*\left(\theta_{i_r,1}^* - \theta_{i_r,2}^*\right)|\text{pre}_\mathbf{T}(\mathbf{B}_{i_r,\min})\right\}\right]\\ &= \mathbb{E}_{P(\text{pre}_\mathbf{T}(\mathbf{B}_{i_r,\min}))}\left[\omega_{i_r}^*\underbrace{\mathbb{E}_P\left\{\theta_{i_r,1}^*|\text{pre}_\mathbf{T}(\mathbf{B}_{i_r,\min})\right\}}_{=\theta_{i_r,2}^*} - \theta_{i_r,2}^*\right]\\ &= 0,\end{aligned}$$

where the first equality holds by the law of total expectation, and the second equality holds since $\omega^*_{i_r}$ and $\theta^*_{i_r,2}$ are functions of $\text{pre}_{\mathbf{T}}(\mathbf{B}_{i_r,\min})$ and $\mathbb{E}_P\left\{\theta^*_{i_r,1}|\text{pre}_{\mathbf{T}}(\mathbf{B}_{i_r,\min})\right\} = \theta^*_{i_r,2}$.

We now check the second condition. Let i_r be any fixed index. Let $\eta^*_{\omega_{i_r}} \equiv \{\omega_{i_r}\} \cup \{\eta^* \backslash \omega^*_{i_r}\}$, which is a nuisance set constructed by replacing the true nuisance $\omega^*_{i_r}$ to an arbitrary nuisance ω_{i_r} from η^*. Also, let $\theta_{i_{r-1}}$ denote $\theta_{i_r,2}$ and $\theta_{i_{r-1},1}$ (note $\theta_{i_{r-1},1}$ and $\theta_{i_r,2}$ are the same nuisance, since they compose of the same set of conditional probabilities). Let $\eta^*_{\theta_{i_{r-1}}} \equiv \{\theta_{i_{r-1}}\} \cup \{\eta^* \backslash \theta^*_{i_{r-1}}\}$. Then, ϕ is a Neyman orthogonal score if the derivative of the expectation of ϕ w.r.t. nuisances evaluated at the true nuisance is zero (Chernozhukov et al., 2018, Def.2.1).

First,

$$\frac{\partial}{\partial \omega_{i_r}} \mathbb{E}_P\left[\phi(\mathbf{V}; \eta^*_{\omega_{i_r}}, \psi)\right] |_{\omega_{i_r}=\omega^*_{i_r}} = \frac{\partial}{\partial \omega_{i_r}} \sum_{j=r}^{m} \left(\mathbb{E}_P\left[\omega_{i_j}\left(\theta^*_{i_j,1} - \theta^*_{i_j,2}\right)\right]\right) |_{\omega_{i_r}=\omega^*_{i_r}} = 0,$$

where the last equality holds since $\mathbb{E}_P\left[\omega_{i_j}\left(\theta^*_{i_j,1} - \theta^*_{i_j,2}\right)\right] = \mathbb{E}_P\left[\omega_{i_j}\left(\mathbb{E}_P\left\{\theta^*_{i_j,1}|\text{pre}_{\mathbf{T}}(\mathbf{B}_{i_j,\min})\right\} - \theta^*_{i_j,2}\right)\right] = \mathbb{E}_P\left[\omega_{i_j}\left(\theta^*_{i_j,2} - \theta^*_{i_j,2}\right)\right] = 0.$

Second,

$$\frac{\partial}{\partial \theta_{i_{r-1}}} \mathbb{E}_P\left[\phi(\mathbf{V}; \eta^*_{\theta_{i_{r-1}}}, \psi)\right] |_{\theta_{i_r}=\theta^*_{i_{r-1}}} = \frac{\partial}{\partial \theta_{i_{r-1}}} \mathbb{E}_P\left[\omega^*_{i_{r-1}}\theta_{i_{r-1},1} - \omega^*_{i_r}\theta_{i_r,2}\right] = \mathbb{E}_P[\omega^*_{i_{r-1}} - \omega^*_{i_r}] = 0.$$

The second equality holds since $\theta_{i_{r-1},1}, \theta_{i_r,2}$ are the same nuisance composing the same set of conditional probabilities. The third equality holds by the law of total expectation. The first equality holds since

$$\begin{aligned}
&\frac{\partial}{\partial \theta_{i_{r-1}}} \mathbb{E}_P\left[\phi(\mathbf{V}; \eta, \psi)\right] \\
&= \frac{\partial}{\partial \theta_{i_{r-1}}} \mathbb{E}_P\left[\sum_{s=1}^{p} \omega_{i_s}\left(\theta_{i_s,1} - \theta_{i_s,2}\right)\right] \\
&= \frac{\partial}{\partial \theta_{i_{r-1}}} \mathbb{E}_P\left[\sum_{s=1}^{p} \omega_{i_s}\theta_{i_s,1} - \sum_{s=1}^{p} \omega_{i_s}\theta_{i_s,2}\right] \\
&= \underbrace{\frac{\partial}{\partial \theta_{i_{r-1}}} \mathbb{E}_P\left[\sum_{\substack{s=1 \\ \theta_{i_{s-1},1} \neq \theta_{i_{r-1}}}}^{p} \omega_{i_s}\theta_{i_s,1} - \sum_{\substack{s=1 \\ \theta_{i_s,2} \neq \theta_{i_{r-1}}}}^{p} \omega_{i_s}\theta_{i_s,2}\right]}_{=0} + \frac{\partial}{\partial \theta_{i_r}} \mathbb{E}_P\left[\sum_{\substack{s=1 \\ \theta_{i_{s-1},1} = \theta_{i_{r-1}}}}^{p} \omega_{i_s}\theta_{i_s,1} - \sum_{\substack{s=1 \\ \theta_{i_s,2} = \theta_{i_{r-1}}}}^{p} \omega_{i_s}\theta_{i_s,2}\right] \\
&= \frac{\partial}{\partial \theta_{i_r}} \mathbb{E}_P\left[\omega_{i_{r-1}}\theta_{i_{r-1},1} - \omega_{i_r}\theta_{i_r,2}\right].
\end{aligned}$$

Therefore, ϕ is a Neyman orthogonal score with respect to η. □

Proposition B.1 (Restated Prop. 3). *Let $P_{\mathbf{x}}(\mathbf{y})$ be identified as $P_{\mathbf{x}}(\mathbf{y}) = \psi \equiv \Psi(P)$. Then, the IF $\phi_{P_{\mathbf{x}}(\mathbf{y})} = \mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})} - \mathbb{E}_P[\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})}]$, where $\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})}$ is derived by Algo. 1 IFP, is a Neyman orthogonal score for ψ.*

Proof. Recall that the set of nuisances for $\phi_{P_{\mathbf{x}}(\mathbf{y})}$ is $\eta = \{\boldsymbol{\omega}_j, \boldsymbol{\theta}_j\}_{j=1}^{\ell}$, where $\boldsymbol{\omega}_j = \{\omega_{j,k}\}_{k=1}^{m_j}$ and $\boldsymbol{\theta}_j = \{\theta_{j,0,1}\} \cup \{\theta_{j,k,1}, \theta_{j,k,2}\}_{k=1}^{m_j}$ are nuisances as specified in Lemma 6. By Lemma S.7, $\phi_{P_{\mathbf{x}}(\mathbf{y})}$ is a linear combination of IFs of CE-1 in the form of

$$\phi_{P_{\mathbf{x}}(\mathbf{y})} = \sum_{\mathbf{d} \backslash \mathbf{y}} \sum_{j=1}^{\ell} f_j \sum_{\mathbf{h}_j} \phi_j(\boldsymbol{\omega}_j, \boldsymbol{\theta}_j).$$

For any fixed $j \in \{1, 2, \cdots, \ell\}$ and $k \in \{1, 2, \cdots, m_j\}$, let $\eta_{j,k} \in \{\omega_{j,k}, (\theta_{j,k-1,1}, \theta_{j,k,2})\}$. Let η^* denote a set of true nuisances. Let $\eta^*_{\eta_{j,k}} \equiv \{\eta_{j,k}\} \cup \{\eta^* \backslash \eta^*_{j,k}\}$ constructed by replacing $\eta^*_{j,k}$ in η^* with an arbitrary nuisance $\eta_{j,k}$.

We recall that ϕ is a Neyman orthogonal score if it satisfies (1) $\mathbb{E}_P[\phi(\mathbf{V};\psi,\eta^*)]=0$ and (2) $\frac{\partial}{\partial\eta}\mathbb{E}_P[\phi(\mathbf{V};\psi,\eta)]|_{\eta=\eta^*}=0$. To witness $\phi_{P_{\mathbf{x}}(\mathbf{y})}$ is a Neyman orthogonal score, we first check whether $\mathbb{E}_P[\phi_{P_{\mathbf{x}}(\mathbf{y})}(\mathbf{V};\eta^*,\psi)]=0$ holds. This holds, since $\mathbb{E}_P[\phi_j(\boldsymbol{\omega}_j^*,\boldsymbol{\theta}_j^*)]=0$ by Lemma S.8. Then, we will check whether $\frac{\partial}{\partial\eta_{j,k}}\mathbb{E}_P[\phi_{P_{\mathbf{x}}(\mathbf{y})}(\mathbf{V};\eta^*_{\eta_{j,k}},\psi)]|_{\eta_{j,k}=\eta^*_{j,k}}=0$ holds. Consider the following:

$$\begin{aligned}\frac{\partial}{\partial\eta_{j,k}}\mathbb{E}_P[\phi_{P_{\mathbf{x}}(\mathbf{y})}(\mathbf{V};\eta^*_{\eta_{j,k}},\psi)]|_{\eta_{j,k}=\eta^*_{j,k}} &= \frac{\partial}{\partial\eta_{j,k}}\sum_{\mathbf{d}\setminus\mathbf{y}}\sum_{\substack{j=1\\ \eta_{j,k}\in\{\boldsymbol{\omega}_j,\boldsymbol{\theta}_j\}}}^{\ell} f_j\sum_{\mathbf{h}_j}\mathbb{E}_P[\phi_j(\boldsymbol{\omega}_j,\boldsymbol{\theta}_j)]\\ &=\sum_{\mathbf{d}\setminus\mathbf{y}}\sum_{\substack{j=1\\ \eta_{j,k}\in\{\boldsymbol{\omega}_j,\boldsymbol{\theta}_j\}}}^{\ell} f_j\sum_{\mathbf{h}_j}\frac{\partial}{\partial\omega_{j,k}}\mathbb{E}_P[\phi_j(\boldsymbol{\omega}_j,\boldsymbol{\theta}_j)]|_{\eta_{j,k}=\eta^*_{j,k}}=0,\end{aligned}$$

where the last equality holds by Lemma S.8. This implies that $\phi_{P_{\mathbf{x}}(\mathbf{y})}$ is a Neyman orthogonal score. □

Lemma S.9 (Sufficient condition for consistency). *Let T_N be an estimator for ψ. If T_N is asymptotically unbiased (i.e., $\lim_{N\to\infty}\mathbb{E}_P[T_N]-\psi=0$) and $\lim_{N\to\infty}Var(T_N)=0$, then T_N is consistent for ψ; i.e., $T_N-\psi=o_P(1)$.*

Proof. For any $\epsilon>0$, by Markov inequality,

$$P(|T_N-\psi|>\epsilon)\le\frac{\mathbb{E}_P[(T_N-\psi)^2]}{\epsilon^2}.$$

Let $\mu_{T_N}\equiv\mathbb{E}_P[T_N]$. By the bias-variance decomposition, we note

$$\mathbb{E}_P[(T_N-\psi)^2]=(\mu_{T_N}-\psi)^2+\mathbb{E}_P[(T_N-\mu_{T_N})^2],$$

and therefore,

$$\begin{aligned}P(|T_N-\psi|>\epsilon)&\le\frac{1}{\epsilon^2}\left((\mu_{T_N}-\psi)^2+\mathbb{E}_P[(T_N-\mu_{T_N})^2]\right)\\&=\frac{1}{\epsilon^2}\left((\mu_{T_N}-\psi)^2+Var(T_N)\right).\end{aligned}$$

By the given condition that $\mu_{T_N}-\psi\to0$, and $\lim_{N\to\infty}Var(T_N)=0$, we have

$$\begin{aligned}\lim_{N\to\infty}P(|T_N-\psi|>\epsilon)&\le\lim_{N\to\infty}\frac{1}{\epsilon^2}\left((\mu_{T_N}-\psi)^2+Var(T_N)\right)\\&=0,\end{aligned}$$

implying that $\lim_{N\to\infty}P(|T_N-\psi|>\epsilon)=0$. □

Lemma S.10. *Suppose $P_{\mathbf{x}}(\mathbf{y})$ is identified as CE-1. Let T_N denote the DML-IDP estimator (Def. 5) of $P_{\mathbf{x}}(\mathbf{y})$ constructed based on the UIF of $P_{\mathbf{x}}(\mathbf{y})$ in Lemma 1 with nuisances $\boldsymbol{\omega}=\{\omega_k\}_{k=1}^m$ and $\boldsymbol{\theta}=\{\theta_{0,1}\}\cup\{\theta_{k,1},\theta_{k,2}\}_{k=1}^m$. Then, T_N is consistent if, for every k, either estimates $\widehat{\omega}_k$ or $(\widehat{\theta}_{k-1,1},\widehat{\theta}_{k,2})$ converge to the true nuisances at rate $o_P(1)$.*

Proof. Let $\mathcal{V}$ denote the UIF. Let $\widehat{\mathcal{V}}\equiv\mathcal{V}(\mathbf{V};\widehat{\eta})$. To show that T_N is a consistent estimator for ψ (i.e., $T_N-\psi=o_P(1)$), it suffices to show that $\mathbb{E}_P\left[\widehat{\mathcal{V}}-\mathcal{V}\right]=o_P(1)$ by Lemma S.9, since $Var(T_N)=\frac{1}{N}Var(\widehat{\mathcal{V}})\to0$ as $N\to\infty$ by the strict boundedness assumption.

Let $\mathbf{X}_k=\mathbf{B}_{j_k}$, i.e., $\mathbf{X}=\{\mathbf{X}_1\prec\cdots\prec\mathbf{X}_m\}=\{\mathbf{B}_{j_1}\prec\cdots\prec\mathbf{B}_{j_m}\}$. Then, we rewrite $\omega_k\equiv\prod_{r=1}^k\frac{I_{\mathbf{x}_r}(\mathbf{X}_r)}{P(\mathbf{X}_r|\text{pre}_{\mathbf{T}}(\mathbf{X}_r))}$. Let $\omega_j^k\equiv\prod_{r=j}^k\frac{I_{\mathbf{x}_r}(\mathbf{X}_r)}{P(\mathbf{X}_r|\text{pre}_{\mathbf{T}}(\mathbf{X}_r))}$. Let $w_r\equiv\frac{I_{\mathbf{x}_r}(\mathbf{X}_r)}{P(\mathbf{X}_r|\text{pre}_{\mathbf{T}}(\mathbf{X}_r))}$. Let

$$Q_j\equiv\theta_{j-1,1}+\sum_{k=j}^m\omega_j^k\left(\theta_{k,1}-\theta_{k,2}\right).$$

Then, $Q_1=\mathcal{V}$.

Since

$$\begin{aligned}
Q_j &= \theta_{j-1,1} + w_j\left(\theta_{j,1} - \theta_{j,2}\right) + w_j w_{j+1}\left(\theta_{j+1,1} - \theta_{j+1,2}\right) + \cdots \\
Q_{j+1} &= \theta_{j,1} + w_{j+1}\left(\theta_{j+1,1} - \theta_{j+1,2}\right) + w_{j+1} w_{j+2}\left(\theta_{j+2,1} - \theta_{j+2,2}\right) + \cdots \\
w_j Q_{j+1} &= w_j\theta_{j,1} + w_j w_{j+1}\left(\theta_{j+1,1} - \theta_{j+1,2}\right) + w_j w_{j+1} w_{j+2}\left(\theta_{j+2,1} - \theta_{j+2,2}\right) + \cdots,
\end{aligned}$$

we can rewrite Q_j as

$$\begin{aligned}
Q_j &= w_j\left(Q_{j+1} - \theta_{j,1}\right) + w_j\left(\theta_{j,1} - \theta_{j,2}\right) + \theta_{j-1,1} \\
&= w_j\left(Q_{j+1} - \theta_{j,2}\right) + \theta_{j-1,1}.
\end{aligned}$$

Let $\widehat{Q}_j$ denote an estimated Q_j with $\widehat{\eta}$. Then,

$$\begin{aligned}
&\mathbb{E}_P\left[\widehat{Q}_j - \theta_{j,2}\right] \\
&= \mathbb{E}_P\left[\widehat{w}_j\left(\widehat{Q}_{j+1} - \widehat{\theta}_{j,2}\right) + \widehat{\theta}_{j-1,1} - \theta_{j,2}\right] \\
&= \mathbb{E}_P\left[\widehat{w}_j\left(\widehat{Q}_{j+1} - \theta_{j+1,2}\right) + \widehat{w}_j\left(\theta_{j+1,2} - \widehat{\theta}_{j,2}\right) + \widehat{\theta}_{j-1,1} - \theta_{j,2}\right] \\
&= \mathbb{E}_P\left[\widehat{w}_j\left(\widehat{Q}_{j+1} - \theta_{j+1,2}\right) + \widehat{w}_j\left(\theta_{j,1} - \widehat{\theta}_{j,2}\right) + \left(\theta_{j,2} - \widehat{\theta}_{j,2}\right)\right] \\
&= \mathbb{E}_P\left[\widehat{w}_j\left(\widehat{Q}_{j+1} - \theta_{j+1,2}\right)\right] + \mathbb{E}_P\left[\widehat{w}_j\left(\theta_{j,2} - \widehat{\theta}_{j,2}\right) + \left(\theta_{j,2} - \widehat{\theta}_{j,2}\right)\right] \\
&= \mathbb{E}_P\left[\widehat{w}_j\left(\widehat{Q}_{j+1} - \theta_{j+1,2}\right)\right] + \mathbb{E}_P\left[\widehat{w}_j\left(\theta_{j,2} - \widehat{\theta}_{j,2}\right) + \left(\theta_{j,2} - \widehat{\theta}_{j,2}\right)\right] \\
&= \mathbb{E}_P\left[\widehat{w}_j\left(\widehat{Q}_{j+1} - \theta_{j+1,2}\right)\right] + o_P\left(\left\|\widehat{w}_j - w_j\right\|_2 \cdot \left\|\theta_{j,2} - \widehat{\theta}_{j,2}\right\|_2\right).
\end{aligned}$$

Note that $\mathbb{E}_P[\theta_{1,2}] = \mathbb{E}_P[\theta_{0,1}] = \psi = \mathbb{E}_P[\mathcal{V}]$. Then, this implies that

$$\mathbb{E}_P\left[\widehat{\mathcal{V}} - \mathcal{V}\right] = \mathbb{E}_P\left[\widehat{Q}_1 - \theta_{1,2}\right] = \sum_{k=1}^{m} o_P\left(\left\|\widehat{w}_j - w_j\right\|_2 \cdot \left\|\theta_{j,2} - \widehat{\theta}_{j,2}\right\|_2\right). \tag{B.15}$$

Under the strict boundedness assumption and the given condition (i.e., either estimates $\widehat{\omega}_j$ or $(\widehat{\theta}_{j-1,1}, \widehat{\theta}_{j,2})$ converge to the true nuisances at rate $o_P(1)$), Eq. (B.15) $= \sum_{k=1}^{m} o_P\left(\left\|\widehat{w}_j - w_j\right\|_2 \cdot \left\|\theta_{j,2} - \widehat{\theta}_{j,2}\right\|_2\right) = o_P(1)$. That is, $\mathbb{E}_P\left[\widehat{\mathcal{V}} - \mathcal{V}\right] = o_P(1)$.

□

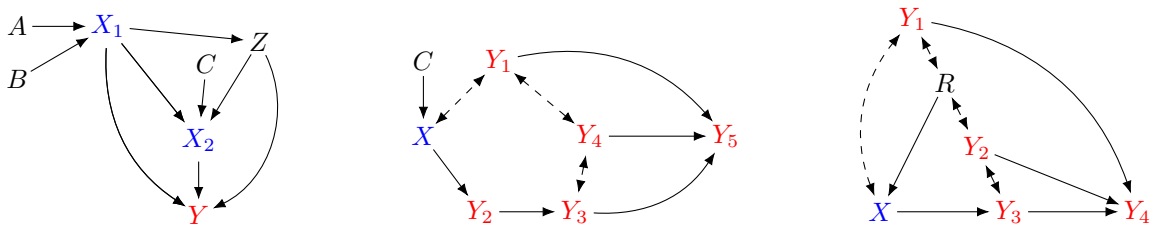
Theorem B.2 (**Restated Thm. 2**). *Let T_N be the DML-IDP estimator of $P_{\mathbf{x}}(\mathbf{y})$ defined in Def. 5 constructed based on the UIF $\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})}(\mathbf{V}; \eta = \{\boldsymbol{\omega}_j, \boldsymbol{\theta}_j\}_{j=1}^{\ell})$ where $\boldsymbol{\omega}_j = \{\omega_{j,k}\}_{k=1}^{m_j}$ and $\boldsymbol{\theta}_j = \{\theta_{j,0,1}\} \cup \{\theta_{j,k,1}, \theta_{j,k,2}\}_{k=1}^{m_j}$ are nuisances as specified in Lemma 6. Suppose T_N is bounded from above by some constant $C \in \mathbb{R}$; i.e., $T_N < C < \infty$. Then,*

*1. **Debiasedness**: T_N is $\sqrt{N}$-consistent and asymptotically normal if estimates for all nuisances converge to the true nuisances at least at rate $o_P(N^{-1/4})$.*

*2. **Doubly robustness**: T_N is consistent if, for every $j = 1, \cdots, \ell$ and $k = 1, \cdots, m_j$, either estimates $\widehat{\omega}_{j,k}$ or $(\widehat{\theta}_{j,k-1,1}, \widehat{\theta}_{j,k,2})$ converge to the true nuisances at rate $o_P(1)$.*

Proof. We note that $\phi_{P_{\mathbf{x}}(\mathbf{y})}(\mathbf{V}; \eta, \psi) = \mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})}(\mathbf{V}; \eta) - \psi$, where $\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})}$ is derived from IFP in Algo. 1, is a Neyman orthogonal score for ψ with nuisances η by Prop. 3. We also note that T_N is a DML estimator satisfying the definition in (Chernozhukov et al., 2018, Def.3.1) since T_N satisfies $\sum_{p\in\{0,1\}} \frac{2}{N} \sum_{\mathbf{V}_{(i)} \in \mathcal{D}_p} \phi_{P_{\mathbf{x}}(\mathbf{y})}\left(\mathbf{V}_{(i)}; \widehat{\eta}, T_N\right) = 0$. Then, the **Debiasedness** property follows by (Chernozhukov et al., 2018, Thm.3.1).

The **Doubly robustness** property comes from that $\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})}$ is an arithmetic function of UIFs $\mathcal{V}_j(\boldsymbol{\omega}_j, \boldsymbol{\theta}_j)$ for CE-1 (by Lemma 6) that are given by Lemma 1. Specifically, an arithmetic function is a continuous function of $\mathcal{V}_j(\boldsymbol{\omega}_j, \boldsymbol{\theta}_j)$ by the strict



(a) A DAG for Fig. 2a (b) A DAG for Fig. 2b (c) A DAG for Fig. 1

Figure C.4: DAGs for the SCMs used in the experiments corresponding to the PAGs in Figs. 2a, 2b, and 1.

boundedness assumption. By Lemma S.10, each estimates of $\mathcal{V}_j(\boldsymbol{\omega}_j, \boldsymbol{\theta}_j)$ achieves consistency under the given condition. Then, by the continuous mapping theorem, the estimate of $\mathcal{V}_{P_{\mathbf{x}}(\mathbf{y})}$ achieves consistency. This implies that T_N is a consistent estimator of ψ.

□

C. Details of Experiments

C.1. Evaluating nuisances

We estimate $\omega_k \in \boldsymbol{\omega}$ in the UIF for CE-1 (in Lemma 1) by estimating the conditional probabilities composing ω_k and plugging those into the functional ω_k (i.e., estimating $\widehat{P}(\mathbf{b}_{j_r}|\text{pre}_{\mathbf{T}}(\mathbf{b}_{j_r}))$ for $r = 1, \cdots, k$ and plugging those into the functional of ω_k). Conditional probabilities are estimated using a gradient boosting model XGBoost (Chen & Guestrin, 2016).

For $(\theta_{k,1}, \theta_{k,2}) \in \boldsymbol{\theta}$, we use backward-iterated regression in the literature (Bang & Robins, 2005; Van der Laan & Rose, 2011; Molina et al., 2017; Rotnitzky et al., 2017). For $k = m, m-1, \cdots, 1$, given $\widehat{\theta}_{k,1}$ (where $\widehat{\theta}_{m,1} = \{I_{\mathbf{y}}(\mathbf{Y}_{(i)})\}_{i=1}^{N}$),

1. Estimate $\widehat{\theta}_{k,2}$ by regressing $\widehat{\theta}_{k,1}$ onto $\text{pre}_{\mathbf{T}}(\mathbf{B}_{k_{\min}})$; i.e., $\widehat{\theta}_{k,2} = f_{\widehat{\theta}_{k,1}}(\text{pre}_{\mathbf{T}}(\mathbf{B}_{k_{\min}}))$, where $f_a(\mathbf{b})$ is a regression estimate regressing a onto $\mathbf{b}$ (e.g., neural networks, gradient boosting, etc). In the experiments, we employed a gradient boosting model XGBoost (Chen & Guestrin, 2016)); then,

2. Estimate $\widehat{\theta}_{k-1,1}$ by evaluating $f_{\widehat{\theta}_{k,1}}((\text{pre}_{\mathbf{T}}(\mathbf{B}_{k_{\min}})\backslash\{\mathbf{B}_{j_k}\}, \mathbf{b}_{j_k}))$, a regression estimate evaluated at covariates $\text{pre}_{\mathbf{T}}(\mathbf{B}_{k_{\min}})$ where $\mathbf{B}_{j_k}$ is fixed to $\mathbf{b}_{j_k}$.

C.2. Structural Causal Models used in the experiments

In generating synthetic data for the simulation, we specify a Structural Causal Model (SCM) for each PAG (which is concealed for the sake of the tested algorithms). The directed acyclic graphs (DAGs) (with bidirected edges encoding latent variables) corresponding to the SCMs are shown in Fig. C.4. The DAGs in Fig. C.4 correspond to the PAGs in Fig. 1 and 2 which represent the Markov equivalence class (MEC) of the corresponding DAGs.

The following notations are used. Let $N(\mu, \sigma^2)$ denote a random variable following Normal distribution with the mean and the variance equals to μ, and σ^2, respectively. For a continuous random variable A, let $\mathcal{C}(A)$ denote a mapping assigning a discrete value corresponding to the value of A. For any random variable D, let $\mathcal{B}(D) \equiv \text{Bernoulli}(\text{logit}^{-1}(D))$, where $\text{logit}^{-1}(\cdot)$ is an inverse-logit function, and $\text{Bernoulli}(p)$ for $p \in (0, 1)$ denotes a Bernoulli random variable.

Fig. 2a. The SCM corresponding to Fig. 2a is the following:

$$\begin{aligned}
f_A &= \mathcal{C}(N(2,1)) \\
f_B &= \mathcal{C}(N(-1,2)) \\
f_C &= \mathcal{C}(N(0,0.5)) \\
f_{X_1}(A,B) &= \mathcal{B}\left(N(0,1)+A-B-1\right) \\
f_Z(X_1) &= \mathcal{C}(N(2X_1-1,0.5)) \\
f_{X_2}(X_1,C,Z) &= \mathcal{B}\left(N(0,1)-Z+C\right) \\
f_Y(X_1,X_2,Z) &= \mathcal{B}\left(X_1\cdot Z - X_2\cdot Z + N(0,1)\right).
\end{aligned}$$

Fig. 2b. The SCM corresponding to Fig. 2b is the following:

$$\begin{aligned}
f_{U_{X,Y_1}} &= N(2,1) \\
f_{U_{Y_1,Y_4}} &= N(-2,2) \\
f_{U_{Y_3,Y_4}} &= N(-1,0.5) \\
f_C &= \mathcal{C}(N(1,2)) \\
f_X(C,U_{X,Y_1}) &= \mathcal{B}\left(N(0,1)+C-2U_{X,Y_1}+2\right) \\
f_{Y_1}(U_{X,Y_1},U_{Y_1,Y_4}) &= \mathcal{C}\left(N\left(N(0,1)-2U_{X,Y_1}+U_{Y_1,Y_4},3\right)\right) \\
f_{Y_2}\left(X\right) &= \mathcal{C}\left(N(2X-1,3)\right) \\
f_{Y_3}\left(Z,U_{Y_3,Y_4}\right) &= \mathcal{C}\left(N(Z\cdot U_{Y_3,Y_4},3)\right) \\
f_{Y_4}\left(U_{Y_1,Y_4},U_{Y_3,Y_4}\right) &= \mathcal{C}\left(N(2U_{Y_1,Y_4}-U_{Y_3,Y_4},3)\right) \\
f_Y\left(Y_1,Y_3,Y_4\right) &= \mathcal{B}\left(-Y_1Y_4+Y_3Y_4+Y_1-2\right).
\end{aligned}$$

Fig. 1. The SCM corresponding to Fig. 1 is the following:

$$\begin{aligned}
f_{U_{X,Y_1}} &= N(2,1) \\
f_{U_{Y_1,R}} &= N(-1,2) \\
f_{U_{R,Y_2}} &= N(-2,2) \\
f_{U_{Y_2,Y_3}} &= N(1,2) \\
f_{Y_1}(U_{X,Y_1},U_{Y_1,R}) &= \mathcal{C}(N(U_{X,Y_1}-U_{Y_1,R},2)) \\
f_R(U_{Y_1,R},U_{R,Y_2}) &= \mathcal{C}(N(N(0,1)-2U_{Y_1,R}+U_{R,Y_2},3)) \\
f_X(R,U_{X,Y_1}) &= \mathcal{B}\left(N\left(N(0,1)+R-2U_{X,Y_1}+2,3\right)\right) \\
f_{Y_2}\left(U_{R,Y_2},U_{Y_2,Y_3}\right) &= \mathcal{C}\left(N\left(N(0,1)-U_{Y_2,Y_3}+3U_{R,Y_2},3\right)\right) \\
f_{Y_3}\left(X,U_{Y_2,Y_3}\right) &= \mathcal{C}\left(N((2X-1)U_{Y_2,Y_3},3)\right) \\
f_{Y_4}\left(Y_1,Y_2,Y_3\right) &= \mathcal{B}\left(-Y_1Y_3+Y_1Y_2+Y_1-2\right).
\end{aligned}$$

C.3. Computing ground truth

We establish ground-truth $\mu(\mathbf{x}) \equiv P_{\mathbf{x}}(\mathbf{y})$ by generating a data set from the submodel $M_{\mathbf{x}}$ of the given SCM M. That is, we replace f_{X_i} for $X_i \in \mathbf{X}$ to the constant $f_{X_i} = x_i$, generate the dataset, and compute the ground-truth by $\mu(\mathbf{x}) \equiv \frac{N_{\mathbf{X}=\mathbf{x},\mathbf{Y}=\mathbf{y}}}{N_{\mathbf{X}=\mathbf{x}}}$, where N is the number of data generated from $M_{\mathbf{x}}$, and $N_{\mathbf{X}=\mathbf{x}} \equiv \sum_{i=1}^{N} I_{\mathbf{x}}(\mathbf{X}_{(i)})$, and $N_{\mathbf{X}=\mathbf{x},\mathbf{Y}=\mathbf{y}} \equiv \sum_{i=1}^{N} I_{\mathbf{x}}(\mathbf{X}_{(i)}) I_{\mathbf{y}}(\mathbf{Y}_{(i)})$.

C.4. Code

Code can be found in https://github.com/yonghanjung/ICML21-DMLIDP.