TECHNICAL REPORT
R-91
May 2023

Estimating Joint Treatment Effects by Combining Multiple Experiments

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

Abstract

Estimating the effects of multi-dimensional treatments (i.e., joint treatment effects) is critical in many data-intensive domains, including genetics and drug evaluation. The main challenges for studying the joint treatment effects include the need for large sample sizes to explore different treatment combinations as well as potentially unsafe treatment interactions. In this paper, we develop machinery for estimating joint treatment effects by combining data from multiple experimental datasets. In particular, first, we develop new identification conditions for determining whether joint treatment effects can be expressed as a multi-distribution adjustment formula. Further, we develop estimators with statistically appealing properties such as consistency and robustness to model misspecification and slow convergence. Finally, we perform simulation studies that corroborate the effectiveness of the proposed methods.

1. Introduction

A large body of scientific research is concerned with estimating the effect of multi-dimensional treatments. For example, Genome-Wide Association Studies (GWAS) in computational biology application study the effect of multiple combinations of genes (Tam et al., 2019). As another example, estimating the multi-dimensional treatment effects is essential in the pharmaceutical industry because potential treatment-treatment interactions can lead to harmful effects to patients, potentially lethal in some situations. Consider two real-world scenarios in which understanding the treatment-treatment interaction is critical:

Example TTI (**Treatment-Treatment-Interaction** (Lee et al., 2019))**.** Many experimental studies have been conducted on the effects of antihypertensive drugs (X_1) on blood pressure (W) with baseline characteristics (C_1) (e.g., (Hansson et al., 1999)) and on the effects of anti-diabetic drugs (X_2) on cardiovascular disease (Y) with baseline characteristics (C_2) (e.g., (Ajjan & Grant, 2006)). Other studies reported that simultaneously taking both durgs was harmful to the population (Ferrannini & Cushman, 2012). This leaves open the question on how to evaluate the joint effect of antihypertensive and anti-diabetic medications from data coming from individual randomized studies. ■

Example MTI (**Multiple Treatments Interactions**)**.** Many experimental studies have been conducted on the effects of (1) taking an aspirin (X_1) on blood pressure (W_1) (e.g., (Hansson et al., 1998); (2) taking acetaminophen (X_2) on blood coagulation (W_2) (e.g., (Gazzard et al., 1974)); and (3) taking the ibuprofen (X_3) on the gastrointestinal disease (Y) (e.g., (Lesko & Mitchell, 1995)). Other more recent studies reported adverse drug reactions to taking ibuprofen with aspirins and acetaminophen (Moore et al., 2015). What are the causal effects of the combinations of such drugs? ■

Despite their critical importance, the analysis of multi-dimensional effects remain underrepresented compared to the vast literature on single-treatment experiments. This is primarily due to two major challenges: the requirement for large sample sizes to investigate all possible treatment combinations and the possibility of unsafe or unethical treatment interactions (Examples TTI and MTI). It is, therefore, of great importance to investigate the possibility of estimating joint treatment effects by combining data from multiple *marginal experiments*, which refer to experiments on a subset of treatments (e.g., a single treatment). In this paper, we present novel methods for estimating joint effects given data from multiple marginal experiments, as well as a qualitative description of the underlying causal system articulated in the form of a causal graph. Specifically,

1. We develop nonparametric identification criteria determining whether the joint treatment effects can be expressed through an adjustment formula using distributions from marginal experiments.

2. We construct estimators for the joint treatment effects using samples from marginal experiments and provide learning guarantees for the estimators.

3. We illustrate the empirical validity of the estimators through simulations, which corroborate the theory.

[1]Purdue University [2]Iowa State Univerity [3]Columbia University. Correspondence to: Yonghan Jung <jung222@purdue.edu>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

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

1.1. Related Work

Causal Effect Identification and Estimation. Recent advances in the literature of *generalized causal effect identification* (g-ID) lead to algorithmic solutions for determining the identification of a causal effect from a set of observational and experimental studies given a causal graph (Bareinboim & Pearl, 2012a; 2016; Lee et al., 2019; Lee & Bareinboim, 2020; Lee et al., 2020; Correa et al., 2021). In addition, recent progress has been made in developing estimators for any causal effects identifiable from observational data in a causal graph (Jung et al., 2020; 2021a;b; Bhattacharya et al., 2022; Jung et al., 2022). However, these estimators are not applicable to g-ID settings that involve multiple experimental distributions.

On a different thread, estimating causal effects from multiple experiments and observations has been investigated for some specific settings. For example, the problems of estimating a long-term effect of a single treatment by combining multiple short-term experimental studies into a surrogate variable have been recently studied (e.g., Bareinboim & Pearl (2012b); Athey et al. (2019; 2020); Imbens et al. (2022)). In epidemiology, estimators for causal effects in a target domain by combining multiple experiments in different source domains have been developed (e.g., Dahabreh et al. (2019); Colnet et al. (2020); Degtiar & Rose (2021); Shi et al. (2022)). However, these methods are not applicable when the goal is to estimate the joint treatment effects from multiple marginal experiments.

Treatment Combinations. The aforementioned examples are related to the analysis of treatment combinations, with aiming to attribute the joint treatment effects to either the effect of treatment combination or marginal treatment effects (e.g., VanderWeele & Knol (2014); Egami & Imai (2018); Parbhoo et al. (2021)). Existing literature commonly relies on the back-door criterion. Such assumptions, however, are not satisfied when latent confounders exist, as illustrated in Examples (TTI, MTI) and Figs. (1a, 2a).

Closer to our work is Saengkyongam & Silva (2020), which investigates the identifiability of joint effects for the additive models with Gaussian noises and continuous treatments by entangling observations and marginal experiments. However, this approach is inapplicable when the treatment variables are discrete, which is common in many applications. In contrast to these methods, we provide nonparametric identifiability criteria for the joint effects from marginal experiments based on a causal graph without imposing such constraints on the data-generating processes. Additionally, we develop estimators for joint treatments effect having statistically desirable properties.

2. Preliminaries

Notations. Each variable is represented with a capital letter (X) and its realized value with a small letter (x). We use bold letters ($\mathbf{X}$) to denote a random vector. Given an ordered set $\mathbf{X} = (X_1, \cdots, X_n)$ such that $X_i \prec X_j$ for $i < j$, we denote $\mathbf{X}^{(i)} = \{X_1, \cdots, X_i\}$. For a graph G over $\mathbf{V}$ and disjoint vectors $\mathbf{X}_1, \mathbf{X}_2 \subseteq \mathbf{V}$, we will use $G_{\overline{\mathbf{X}_1}\underline{\mathbf{X}_2}}$ as a subgraph of G in which all incoming edges to the node in $\mathbf{X}_1$ and all outgoing edges to the node in $\mathbf{X}_2$ are cut. For a discrete (e.g., binary) random vector $\mathbf{X}$ and its realized value $\mathbf{x} \in \mathfrak{D}_{\mathbf{X}}$ where $\mathfrak{D}_{\mathbf{X}}$ is the domain of $\mathbf{X}$, we use $\mathbb{1}_{\mathbf{x}}(\mathbf{X})$ to represent the indicator function such that $\mathbb{1}_{\mathbf{x}}(\mathbf{X}) = 1$ if $\mathbf{X} = \mathbf{x}$; $\mathbb{1}_{\mathbf{x}}(\mathbf{X}) = 0$ otherwise. For a random vector $\mathbf{X}$, we use $P(\mathbf{X})$ to denote its distribution and $p(\mathbf{x})$ as a corresponding density function at $\mathbf{X} = \mathbf{x}$. For a function f, $\mathbb{E}_P[f(\mathbf{X})] \coloneqq \int_{\mathfrak{S}_{\mathbf{X}}} f(\mathbf{x})p(\mathbf{x})\, d[\mathbf{x}]$ where $\mathfrak{S}_{\mathbf{X}}$ is the support for $\mathbf{X}$. For a sample set $D \coloneqq \{\mathbf{V}_{(i)}\}_{i=1}^{n}$ where $\mathbf{V}_{(i)}$ denotes the ith samples, we use $\mathbb{E}_D[f(\mathbf{V})] \coloneqq (1/n)\sum_{i=1}^{n} f(\mathbf{V}_{(i)})$. We use $\|f\|_P \coloneqq \sqrt{\mathbb{E}_P[(f(\mathbf{X}))^2]}$. If a function $\widehat{f}$ is a consistent estimator of f having a rate r_n, we will use $\widehat{f} - f = o_P(r_n)$. We will say $\hat{f}$ is L_2-consistent if $\|\hat{f} - f\|_P = o_P(1)$. We will use $\widehat{f} - f = O_P(1)$ if $\widehat{f} - f$ is bounded in probability. Also, $\widehat{f} - f$ is said to be bounded in probability at rate r_n if $\widehat{f} - f = O_P(r_n)$. Throughout the paper, we assume that samples D are independent.

Structural Causal Models. We use Structural Causal Models (SCMs) as our framework (Pearl, 2000; Bareinboim et al., 2022). An SCM $\mathcal{M}$ is a quadruple $\mathcal{M} = \langle \mathbf{U}, \mathbf{V}, P(\mathbf{U}), F\rangle$. $\mathbf{U}$ is a set of exogenous (latent) variables following a joint distribution $P(\mathbf{U})$. $\mathbf{V}$ is a set of endogenous (observable) variables whose values are determined by functions $F = \{f_{V_i}\}_{V_i \in \mathbf{V}}$ such that $V_i \leftarrow f_{V_i}(pa_i, u_i)$ where $PA_i \subseteq V$ and $U_i \subseteq U$. Each SCM $\mathcal{M}$ induces a distribution $P(\mathbf{V})$ and a causal graph $G = G(\mathcal{M})$ over $\mathbf{V}$ in which there exists a directed edge from every variable in PA_i to V_i and dashed-bidirected arrows encode common latent variables (e.g., see Fig. 1a). Performing an intervention fixing $\mathbf{X} = \mathbf{x}$ is represented through the do-operator, $do(\mathbf{X} = \mathbf{x})$, which encodes the operation of replacing the original equations of X (i.e., $f_X(pa_x, u_x)$) by the constant $x \in \mathfrak{D}_X$ for all $X \in \mathbf{X}$ and induces an interventional distribution $P(\mathbf{V}|do(\mathbf{x}))$. We will sometimes employ the shorthand notation $P_{\mathbf{x}}(\mathbf{y})$ to represent $P(\mathbf{y}|do(\mathbf{x}))$. We will use $P_{\text{rand}(\mathbf{X})}(\mathbf{Y}) \coloneqq \{P_{\mathbf{x}}(\mathbf{Y})\}_{\mathbf{x}\in\mathfrak{S}_{\mathbf{X}}}$. For a sample set $D \coloneqq \{\mathbf{V}_{(i)}\}_{i=1}^{n}$, D is said to follow $P_{\text{rand}(\mathbf{X})}(\mathbf{V})$ if each subsamples $D_{\mathbf{x}} \coloneqq \{\mathbf{V}_{(i)}\}_{\mathbf{V}_{(i)}\in D, \mathbf{X}_{(i)}=\mathbf{x}}$ follows $P_{\mathbf{x}}(\mathbf{V})$.

3. Combining Two Experiments

In this section, we address the challenge of estimating the combined effects by leveraging the results of two distinct experiments. In Section 3.1, we delve into the estimation of treatment-treatment interactions (TTI) based on the outcomes of two separate marginal experiments. Then, in Section 3.2, we extend our investigation to accommodate scenarios where the treatment effects in the source and target experiments may not align perfectly.

3.1. Treatment-Treatment Interaction

Our goal is to estimate joint effects by combining two *randomized controlled experiments*, as formally defined below.

Task TTI (Treatment-Treatment Interaction (TTI)). The task of estimating the treatment-treatment interactions (TTI) from two marginal experiments composes of

- **Input**: Two sets of samples, D_1 and D_2, which follow interventional distributions $P_{\text{rand}(X_1)}(C_1, X_1, W)$ and $P_{\text{rand}(X_2)}(C_1, W, X_1, X_2, Y)$, respectively. C_1 is a covariate for the experiment randomizing X_1 (i.e., $\text{rand}(X_1)$), and W, Y represents the outcomes of the experiments randomizing X_1 ($\text{rand}(X_1)$) and X_2 ($\text{rand}(X_2)$), respectively.
- **Query**: Estimation of $\mathbb{E}[Y|do(x_1, x_2)]$.

3.1.1. Adjustment Criterion for TTI (AC-TTI)

A sufficient graphical criterion for identifying the treatment-treatment interaction is the following:

Definition 1 (Adjustment criterion for Treatment-Treatment Interaction (AC-TTI)). A set $\{C_1, W\}$ is said to satisfy the *adjustment criterion for treatment-treatment interaction (AC-TTI)* w.r.t $\{(X_1, X_2), Y\}$ in G if

1. $(\{C_1, W\} \perp\!\!\!\perp X_2 | X_1)_{G_{\overline{X_1, X_2}}}$; there are no direct paths from X_2 to $\{C_1, W\}$ in $G_{\overline{X_1, X_2}}$; and
2. $(Y \perp\!\!\!\perp X_1 | C_1, W, X_2)_{G_{\underline{X_1}\overline{X_2}}}$; the back-door paths between X_1 and Y are blocked by $\{C_1, W\}$ in $G_{\overline{X_2}}$.

We make the following positivity assumption:

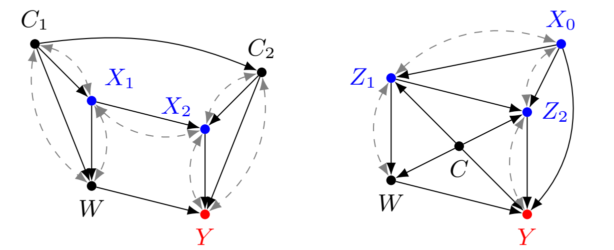
Assumption 1 (Positivity Assumption for AC-TTI). *$P_{x_1}(C_1, W), P_{x_2}(C_1, W), P_{x_2}(X_1|C_1, W)$ are strictly positive distributions for $\forall x_1, x_2 \in \mathfrak{D}_{X_1, X_2}$.*

Under AC-TTI and Assumption 1, the joint treatment effects $\mathbb{E}[Y|do(x_1, x_2)]$ is identifiable and given as follow:

Theorem 1 (Identification through AC-TTI). *Suppose AC-TTI in Def. 1 and Assumption 1 hold. Then, $\mathbb{E}[Y|do(x_1, x_2)]$ is identifiable from $P_{rand(X_1)}(C_1, W)$ and $P_{rand(X_2)}(C_1, W, X_1, Y)$ and the expression is:*

$$\mathbb{E}[Y|do(x_1, x_2)] = \mathbb{E}_{P_{x_1}}\left[\mathbb{E}_{P_{x_2}}[Y|C_1, W, x_1]\right]. \quad (1)$$

For example, in Fig. 1a, $\{C_1, W\}$ satisfies AC-TTI w.r.t. $\{(X_1, X_2), Y\}$. Therefore, with Assumption. 1, $\mathbb{E}[Y|do(x_1, x_2)]$ is identifiable from $P_{\text{rand}(X_1)}(C_1, W)$ and $P_{\text{rand}(X_2)}(C_1, W, X_1, Y)$ as in Eq. (1).



(a) Task TTI (b) Task gTTI

Figure 1: Example causal graphs for Section 3. Nodes representing the treatment and outcome are marked in blue and red respectively.

3.1.2. Estimators for AC-TTI

We define nuisance functional for estimating the AC-TTI functional in Eq. (1) as follow:

Definition 2 (Nuisance for AC-TTI). Nuisance functions for AC-TTI functional in Eq. (1) are defined as follow: For a fixed $x_1, x_2 \in \mathfrak{D}_{X_1, X_2}$ where x_1, x_2 are specified in Eq. (1), $\pi_0 := \pi_0(C_1, X_1, W) := \frac{P_{x_1}(W|C_1)}{P_{x_2}(W, X_1|C_1)}$. Also, $\mu_0 := \mu_0(C_1, X_1, W) := \mathbb{E}_{P_{x_2}}[Y|X_1, W, C_1]$. We will use $\pi := \pi(C_1, X_1, W) > 0$ and $\mu := \mu(C_1, X_1, W)$ to denote arbitrary[1] finite functions.

Now, we construct regression-based ('REG'), probability weighting ('PW') and double/debiased machine learning ('DML') (Chernozhukov et al., 2018) based estimators:

Definition 3 (AC-TTI estimators). Let D_1 and D_2 denote two separate samples following the distribution $P_{\text{rand}(X_1)}(C_1, W)$ and $P_{\text{rand}(X_2)}(C_1, W, X_1, Y)$, respectively. For fixed $x_1, x_2 \in \mathfrak{D}_{X_1, X_2}$, we define D_{x_1} and D_{x_2} as subsamples of D_1 and D_2 such that $X_1 = x_1$ and $X_2 = x_2$. Let μ and π denote the nuisances as defined in Definition 2. We now introduce the {REG, PW, DML} estimators for the AC-TTI-functional specified in Equation (1) as follows:

$$T^{reg} := \mathbb{E}_{D_{x_1}}[\mu(W, C_1, x_1))],$$
$$T^{pw} := \mathbb{E}_{D_{x_2}}[\pi(W, C_1, X_1)\mathbb{1}_{x_1}(X_1)Y],$$
$$T^{dml} := \mathbb{E}_{D_{x_1}}[\pi\mathbb{1}_{x_1}(X_1)\{Y - \mu\}] + \mathbb{E}_{D_1}[\mu(W, C_1, x_1))].$$

We assume that samples used for training the nuisance functions and evaluating the nuisances are independent:

Assumption 2 (Sample-splitting). *Samples for training nuisances and evaluating the estimators equipped with the trained nuisance are separate and independent*[2].

[1] Throughout the paper, μ, π may be understood as estimated nuisances for μ_0, π_0.

[2] This assumption is satisfied by applying cross-fitting algo-

We assume that nuisances can be estimated L_2 consistently. In practice, this assumption can be easily satisfied by employing flexible machine learning models.

Assumption 3 (**L_2 consistency of nuisances**). *Estimated nuisances are L_2 consistent; i.e., $\forall i \in \{1,2\}, \forall x_i \in \mathfrak{D}_{X_i}$,*

$$\|\mu(W, C_1, x_1) - \mu_0(W, C_1, x_1)\|_{P_{x_i}} = o_{P_{x_i}}(1),$$
$$\|\pi(W, C_1, X_1) - \pi_0(W, C_1, X_1)\|_{P_{x_2}} = o_{P_{x_2}}(1).$$

We also assume that the baseline covariates have the same distribution over all marginal experiments. Specifically,

Assumption 4 (**Shared Covariates**). *The distributions of the covariates C_1 are the same; i.e., For all $x_1, x_2 \in \mathfrak{D}_{X_1,X_2}$, $P_{x_1}(C_1) = P_{x_2}(C_1)$.*

Then, the errors for each estimator are given as follows:

Theorem 2 (**Error analysis for AC-TTI estimators**). *Under Assumptions (1,2,3,4) and AC-TTI in Def. 1, the error of the estimators in Def. 3, denoted $\epsilon^{est} := T^{est} - \mathbb{E}[Y|do(x_1, x_2)]$ for $est \in \{reg, pw, dml\}$ are:*

$$\epsilon^{reg} = R_1 + O_{P_{x_1}}(\|\mu - \mu_0\|),$$
$$\epsilon^{pw} = R_2 + O_{P_{x_2}}(\|\pi - \pi_0\|),$$
$$\epsilon^{dml} = R_1 + R_2 + O_{P_{x_2}}(\|\pi - \pi_0\| \, \|\mu - \mu_0\|),$$

where R_i is a random variable such that $\sqrt{n_i}R_i$ converges in distribution to the mean-zero normal random variable, where $n_i := |D_{x_i}|$ for $i \in \{1,2\}$.

We highlight that the DML estimator T^{dml} exhibits robustness property since ϵ^{dml} is bounded in probability at $n^{-1/2}$ rate (for $n = \min\{n_1, n_2\}$) whenever $\|\pi - \pi_0\|_{P_{x_2}} = O_{P_{x_2}}(n^{-1/4})$ and $\|\mu - \mu_0\|_{P_{x_2}} = O_{P_{x_2}}(n^{-1/4})$. Furthermore, the DML estimator displays the following doubly robustness property:

Corollary 2 (**Doubly robustness of the DML estimators** (Corollary of Thm. 2)). *Suppose Assumptions (1,2,3,4) and AC-TTI in Def. 1 hold. Suppose either $\pi = \pi_0$ or $\mu = \mu_0$. Then, T^{dml} is an unbiased estimator of $\mathbb{E}[Y|do(x_1, x_2)]$.*

3.2. Combining Two Arbitrary Experiments

In this section, we extend Task TTI to cases where the effect with two or more treatments (i.e., $|\mathbf{X}| \geq 2$) can be identified from two arbitrary experiments conducted on other variables, denoted as $\mathbf{Z}$. For example, let's consider a scenario extending Example TTI where we are interested in studying the effect of three factors: the antihypertensive drug (Z_1), the anti-diabetic drug (Z_2), and the individual's diet habits (X_0), on the occurrence of cardiovascular disease (as depicted in Fig. 1b) when we are given two marginal experiments randomizing Z_1 and Z_2 respectively. This extended task is referred to as *generalized treatment-treatment interactions* (gTTI) and is defined as follows:

Task gTTI (**Generalized TTI**). The task of *generalized TTI* composes of

- **Input**: Two samples sets D_1, D_2 following distributions $P_{\text{rand}(Z_1)}(\mathbf{V})$ and $P_{\text{rand}(Z_2)}(\mathbf{V})$, respectively.
- **Query**: Estimation of $\mathbb{E}[Y|do(\mathbf{x})]$.

We note that Task gTTI generalizes Task TTI in the sense that it does not require that $\mathbf{X}$ is identical to $\mathbf{Z}$.

3.2.1. Adjustment Criterion for gTTI (AC-gTTI)

A graphical criterion for identifying $\mathbb{E}[Y|do(\mathbf{x})]$ from two distributions $P_{\text{rand}(Z_1)}(\mathbf{V})$ and $P_{\text{rand}(Z_2)}(\mathbf{V})$ is the following:

Definition 4 (**Adjustment criterion for combining two experiments (AC-gTTI)**). A set of variables $\mathbf{A}$ is said to satisfy *adjustment criterion for generalized TTI (AC-gTTI)* w.r.t an ordered set $\mathbf{X}$ and Y in G if

1. $Z_1 \subseteq \mathbf{X}$ and $(\mathbf{A} \perp\!\!\!\perp \mathbf{X}\backslash Z_1 | Z_1)_{G_{\overline{\mathbf{X}}}}$; there are no direct paths from $\mathbf{X}\backslash Z_1$ to $\mathbf{A}$ in $G_{\overline{\mathbf{X}}}$; and
2. $Z_2 \subseteq \mathbf{X}$ and $(Y \perp\!\!\!\perp \mathbf{X}\backslash Z_2 | \mathbf{A}, Z_2)_{G_{\underline{\mathbf{X}\backslash Z_2}\overline{Z_2}}}$; the backdoor paths between $\mathbf{X}\backslash Z_2$ and Y are blocked by $\mathbf{A}$ in $G_{\overline{Z_2}}$.

We make the following positivity assumption:

Assumption 5 (**Positivity Assumption for AC-gTTI**). *$P_{z_1}(\mathbf{A}), P_{z_2}(\mathbf{A}), P_{z_2}(\mathbf{X}\backslash Z2|\mathbf{A})$ are strictly positive distributions for $\forall z_1, z_2 \in \mathfrak{D}_{Z_1,Z_2}$.*

Under AC-gTTI, the joint treatment effects $\mathbb{E}[Y|do(\mathbf{x})]$ is identifiable and given as follow:

Theorem 3 (**Identification through AC-gTTI**). *Suppose AC-gTTI in Def. 4 and Assumption 5 hold. Then, the query $\mathbb{E}[Y|do(\mathbf{x})]$ is identifiable from $P_{rand(Z_1)}(\mathbf{A})$ and $P_{rand(Z_2)}(\mathbf{A}, \mathbf{X}, Y)$ and given as follow:*

$$\mathbb{E}[Y|do(\mathbf{x})] = \mathbb{E}_{P_{z_1}}\left[\mathbb{E}_{P_{z_2}}[Y|\mathbf{A}, \mathbf{x}\backslash\mathbf{z}_2]\right]. \tag{2}$$

For example, in Fig. 1b, $\mathbf{A} := \{W, C\}$ satisfies AC-gTTI criterion w.r.t. $\{\mathbf{X} = (X_0, Z_1, Z_2), Y\}$. Therefore, under positivity, $\mathbb{E}[Y|do(\mathbf{x})]$ is expressible as in Eq. (2).

3.2.2. Estimators for AC-gTTI

We define nuisance functional for AC-gTTI functional in Eq. (2) as follow:

rithms (e.g., (Klaassen, 1987; Robins & Ritov, 1997; Zheng & van der Laan, 2011; Chernozhukov et al., 2018)), which split the samples and using one for training nuisances and another for evaluating the trained nuisances.

Definition 5 (Nuisances for AC-gTTI). Nuisance functions for estimating AC-gTTI functional in Eq. (2) are defined as follow: For a fixed $z_1, z_2 \in \mathfrak{D}_{Z_1,Z_2}$ where z_1, z_2 are specified in Eq. (2), $\pi_0 := \pi_0(\mathbf{A}, \mathbf{X}) := \frac{P_{z_1}(\mathbf{A})}{P_{z_2}(\mathbf{A}, \mathbf{X} \backslash Z_2)}$, and $\mu_0 := \mu_0(\mathbf{A}, \mathbf{X}) := \mathbb{E}_{P_{z_2}}[Y | \mathbf{X} \backslash Z_2, \mathbf{A}]$. We will use $\pi := \pi(\mathbf{A}, \mathbf{X}) > 0$ and $\mu := \mu(\mathbf{A}, \mathbf{X})$ to denote an arbitrary finite function.

Now, we construct regression-based ('REG'), probability weighting ('PW') and double/debiased machine learning ('DML') estimators:

Definition 6 (AC-gTTI estimators). Let D_1, D_2 denote two sample sets following distributions $P_{\text{rand}(Z_1)}(\mathbf{A})$ and $P_{\text{rand}(Z_2)}(\mathbf{A}, \mathbf{X}, Y)$, respectively. For a fixed $z_1, z_2 \in \mathfrak{D}_{Z_1,Z_2}$, we define D_{z_1} and D_{z_2} as subsamples of D_1 and D_2 such that $Z_1 = z_1$ and $Z_2 = z_2$. Let μ, π denote nuisances defined in Def. 5. Then, {REG, PW, DML} estimators for AC-gTTI functional defined as follow:

$$
\begin{aligned}
T^{reg} &:= \mathbb{E}_{D_{z_1}}\left[\mu(\mathbf{A}, \mathbf{x})\right], \\
T^{pw} &:= \mathbb{E}_{D_{z_2}}\left[\pi(\mathbf{A}, \mathbf{X}) \mathbb{1}_{\mathbf{x}}(\mathbf{X}) Y\right], \\
T^{dml} &:= \mathbb{E}_{D_{z_2}}\left[\pi \mathbb{1}_{\mathbf{x}}(\mathbf{X}) \{Y - \mu\}\right] + \mathbb{E}_{D_{z_1}}\left[\mu(\mathbf{A}, \mathbf{x})\right].
\end{aligned}
$$

We assume that nuisances can be estimated L_2 consistently.

Assumption 6 (L_2 consistency of nuisances). *Estimated nuisances are L_2 consistent; i.e., $\forall i \in \{1, 2\}, \forall z_i \in \mathfrak{D}_{Z_i}$,*

$$
\begin{aligned}
\|\mu(\mathbf{A}, \mathbf{x}) - \mu_0(\mathbf{A}, \mathbf{x})\|_{P_{z_i}} &= o_{P_{z_i}}(1), \\
\|\pi(\mathbf{A}, \mathbf{X}) - \pi_0(\mathbf{A}, \mathbf{X})\|_{P_{z_2}} &= o_{P_{z_2}}(1).
\end{aligned}
$$

Then, the error of each estimator are given as follows:

Theorem 4 (Error analysis for AC-gTTI estimators). *Under Assumptions (2,5,6) and AC-gTTI in Def. 4, the error of the estimators in Def. 6, denoted $\epsilon^{est} := T^{est} - \mathbb{E}[Y | do(\mathbf{x})]$ for est $\in$ {reg, pw, dml}, are:*

$$
\begin{aligned}
\epsilon^{reg} &= R_1 + O_{P_{z_1}}\left(\|\mu - \mu_0\|\right), \\
\epsilon^{pw} &= R_2 + O_{P_{z_2}}\left(\|\pi - \pi_0\|\right), \\
\epsilon^{dml} &= R_1 + R_2 + O_{P_{z_2}}\left(\|\pi - \pi_0\| \, \|\mu - \mu_0\|\right),
\end{aligned}
$$

where R_i is a random variable such that $\sqrt{n_i} R_i$ converges in distribution to the mean-zero normal random variable, where $n_i := |D_{z_i}|$.

We highlight that the DML estimator T^{dml} exhibits robustness property since ϵ^{dml} is bounded in probability at $n^{-1/2}$ (for $n = \min\{n_1, n_2\}$) rate whenever $\|\pi - \pi_0\|_{P_{z_2}} = O_{P_{z_2}}(n^{-1/4})$ and $\|\mu - \mu_0\|_{P_{z_2}} = O_{P_{z_2}}(n^{-1/4})$. Furthermore, the DML estimator displays the following doubly robustness property:

Corollary 4 (Doubly robustness of the DML estimators (Corollary of Thm. 4)**).** *Suppose Assumptions (2,5,6) and AC-gTTI in Def. 4 hold. Suppose either $\pi = \pi_0$ or $\mu = \mu_0$. Then, T^{dml} is an unbiased estimator of $\mathbb{E}[Y | do(\mathbf{x})]$.*

4. Combining Multiple (≥ 2) Experiments

In this section, we address the estimation of joint effects by leveraging multiple (more than two) experiments. Specifically, in Sec. 4.1, we focus on estimating multiple treatment interactions (MTI) using multiple marginal experiments. In Sec. 4.2, we extend this setting to estimate multiple treatment effects from multiple experiments in which the randomization was on each element in $\mathbf{Z}$ that are not necessarily matched with the cause of interest $\mathbf{X}$.

4.1. Multiple Treatment Interaction

We first introduce the formal version of the task.

Task MTI (Multiple-Treatment Interaction (MTI)). Estimating multiple treatment interaction (MTI) composes of

- **Input**: Multiple sets of samples $\{D_i\}_{i=1}^m$ drawn from a sequence of interventional distributions $\{P_{\text{rand}(X_i)}(\mathbf{C}^{(i)}, \mathbf{W}^{(i)}, \mathbf{X}^{(i)})\}_{i=1}^m$. $\{C_i, X_i, W_i\}$ for $i = 1, \cdots, m$ is the ith triplet corresponding to a covariate, a treatment, and an outcome.

- **Query**: Estimation of $\mathbb{E}[Y | do(\mathbf{x})]$ where $\mathbf{x} = \{x_i\}_{i=1}^m$ is a realization for an ordered set $\mathbf{X} := \{X_1, \cdots, X_m\}$, and $Y := W_m$.

4.1.1. Adjustment Criterion for MTI (AC-MTI)

A sufficient graphical criterion for identifying the multiple treatment interaction is the following:

Definition 7 (Adjustment criterion for Multiple Treatment Interaction (AC-MTI)). An ordered set $\{C_1, W_1, C_2, W_2, \cdots, C_{m-1}, W_{m-1}\}$ satisfies *adjustment criterion for multiple treatment interaction (AC-MTI)* w.r.t. $\{\mathbf{X}, Y\}$ for $\mathbf{X} = \{X_i\}_{i=1}^m$ in G if, for $i = 1, 2, \cdots, m$,

1. $\{X_j\}_{j>i}$ is non-ancestor of $\{\mathbf{X}^{(i)}, \mathbf{W}^{(i)}, \mathbf{C}^{(i)}\}$; and

2. $(Y \perp\!\!\!\perp X_i | \mathbf{C}^{(i-1)}, \mathbf{X}^{(i-1)}, \mathbf{W}^{(i)}, \mathbf{X}^{>i})_{G_{\underline{X_i}, \overline{\mathbf{X}^{>i}}}}$; the back-door paths between X_i and Y are blocked by $\mathbf{C}^{(i-1)}, \mathbf{X}^{(i-1)}, \mathbf{W}^{(i)}, \mathbf{X}^{>i}$ in the graph $G_{\overline{\mathbf{X}^{>i}}}$.

We make the following positivity assumption:

Assumption 7 (Positivity Assumption for AC-MTI). $\{P_{x_i}(W_i, C_i | \mathbf{W}^{(i-1)}, \mathbf{C}^{(i-1)}, \mathbf{X}^{(i-1)})\}_{i=1}^m$, $P_{x_{i+1}}(X_i | \mathbf{W}^{(i)}, \mathbf{C}^{(i)}, \mathbf{X}^{(i-1)})$ *for $i = 1, \cdots, m-1$ are strictly positive $\forall \mathbf{x} \in \mathfrak{D}_{\mathbf{X}}$.*

Under AC-MTI, the joint treatment effects $\mathbb{E}[Y|do(\mathbf{x})]$ is identifiable and given as follow:

Theorem 5 (Identification through AC-MTI). *Suppose AC-MTI in Def. 7 and Assumption 7 hold. Then, $\mathbb{E}[Y(\mathbf{x})]$ is identifiable from $\{P_{\text{rand}(X_i)}(\mathbf{C}^{(i)}, \mathbf{W}^{(i)}, \mathbf{X}^{(i-1)})\}_{i=1}^{m}$ as follow: Let $\mu_0^m \coloneqq \mathbb{E}_{P_{x_m}}\left[Y|\mathbf{W}^{(m-1)}, \mathbf{C}^{(m-1)}, \mathbf{X}^{(m-1)}\right]$, and for $i = m-1, \cdots, 2$,*

$$\mu_0^i \coloneqq \mathbb{E}_{P_{x_i}}\left[\overline{\mu}_0^{i+1}|\mathbf{W}^{(i-1)}, \mathbf{C}^{(i-1)}, \mathbf{X}^{(i-1)}\right],$$

where $\overline{\mu}_0^{i+1} \coloneqq \mu_0^{i+1}(\mathbf{W}^{(i)}, \mathbf{C}^{(i)}, x_i, \mathbf{X}^{(i-1)})$. Then,

$$\mathbb{E}[Y(\mathbf{x})] = \mathbb{E}_{P_{x_1}}\left[\mu_0^2(W_1, C_1, x_1)\right]. \tag{3}$$

For example, in Fig. 2a, $\{C_1, W_1, C_2, W_2\}$ satisfies AC-MTI w.r.t. $\{(X_1, X_2), Y\}$ in Def. 7. Therefore, with the positivity assumption in Assumption. 7, $\mathbb{E}[Y|do(\mathbf{x})]$ is identifiable from $\{P_{\text{rand}(X_i)}(\mathbf{C}^{(i)}, \mathbf{W}^{(i)}, \mathbf{X}^{(i-1)})\}_{i=1}^{m}$ as in Eq. (3).

4.1.2. Estimators for AC-MTI

We define nuisance functional for estimating the AC-MTI functional in Eq. (3) as follows:

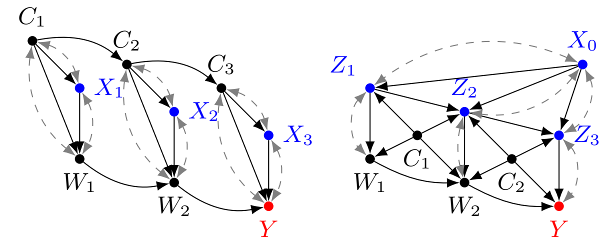
Definition 8 (Nuisances for AC-MTI). Nuisance functions for AC-MTI are defined as follows: For a fixed $\mathbf{x} \coloneqq \{x_1, \cdots, x_m\} \in \mathfrak{D}_{\mathbf{X}}$, let $\{\mu^i\}_{i=2}^m$ and $\{\overline{\mu}^i\}_{i=2}^m$ be the nuisances defined in Thm. 5. For $i = 1, \cdots, m-1$, $\pi_0^i \coloneqq \frac{P_{x_i}(W_i|C_i, \mathbf{C}^{(i-1)}, \mathbf{X}^{(i-1)}, \mathbf{W}^{(i-1)})}{P_{x_m}(W_i, X_i|C_i, \mathbf{C}^{(i-1)}, \mathbf{X}^{(i-1)}, \mathbf{W}^{(i-1)})}$, and $\pi_0^{(i)} \coloneqq \prod_{j=1}^{i} \pi_0^j(\mathbf{W}^{(j)}, \mathbf{C}^{(j)}, \mathbf{X}^{(j)})$. We will use $\pi^i(\mathbf{W}^{(i)}, \mathbf{C}^{(i)}, \mathbf{X}^{(i)}) > 0$ and $\mu^i(\mathbf{W}^{(i-1)}, \mathbf{C}^{(i-1)}, \mathbf{X}^{(i-1)})$ for any arbitrary[3] finite functions.

Now, we construct regression-based ('REG'), probability weighting ('PW') and double/debiased machine learning ('DML') estimators:

Definition 9 (AC-MTI estimators). Let D_i denote samples following $P_{\text{rand}(X_i)}(\mathbf{C}^{(i)}, \mathbf{W}^{(i)}, \mathbf{X}^{(i)})$ for $i = 1, 2, \cdots, m$. For a fixed $x_i \in \mathfrak{D}_{X_i}$, let D_{x_i} denote the subsamples of D_i such that $X_i = x_i$. Let $A_i \coloneqq \{W_i, C_i\}$ and $V_i \coloneqq \{A_i, X_i\}$. Let $\mu^{m+1} \coloneqq Y$. Let $\mathbb{1}_{\mathbf{x}}^{i-1} \coloneqq \mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)})$ for $i = 2, \cdots, m$. Then {REG, PW, DML} estimators are defined as follow:

$$T^{reg} \coloneqq \mathbb{E}_{D_{x_1}}\left[\mu^2(W_1, C_1, x_1))\right],$$
$$T^{pw} \coloneqq \mathbb{E}_{D_{x_m}}\left[\pi^{(m-1)}\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right],$$
$$T^{dml} \coloneqq \sum_{i=2}^{m} \mathbb{E}_{D_{x_i}}\left[\pi^{(i-1)}\mathbb{1}_{\mathbf{x}}^{i-1}\{\overline{\mu}^{i+1} - \mu^i\}\right] + \mathbb{E}_{D_{x_1}}\left[\overline{\mu}^2\right].$$

We assume that nuisances can be estimated L_2 consistently.



(a) Task MTI (b) Task gMTI

Figure 2: Example causal graphs for Section 4.

Assumption 8 (L_2 consistency of nuisances). *Estimated nuisances are L_2-consistent; Specifically,*

$$\|\mu^{i+1} - \mu_0^{i+1}\|_{P_{x_i}} = o_{P_{x_i}}(1),\ \forall i \in \{1, 2, \cdots, m-1\}$$
$$\|\mu^i - \mu_0^i\|_{P_{x_i}} = o_{P_{x_i}}(1),\ \forall i \in \{2, \cdots, m\}$$
$$\|\pi^i - \pi^i\|_{P_{x_{i+1}}} = o_{P_{x_{i+1}}}(1),\ \forall i \in \{1, \cdots, m-1\}.$$

We assume that treatments and outcomes will have the same distribution over all marginal experiments:

Assumption 9 (Shared Covariates). *For any fixed $i, j \in \{1, 2, \cdots, m-1\}$ s.t. $j > i$ and any fixed $x_i, x_j \in \mathfrak{D}_{X_i, X_j}$, the baseline covariates C_i's distribution satisfies the following: $P_{x_i}(C_i|\mathbf{C}^{(i-1)}, \mathbf{X}^{(j-1)}, \mathbf{W}^{(j-1)}) = P_{x_j}(C_i|\mathbf{C}^{(i-1)}, \mathbf{X}^{(i-1)}, \mathbf{W}^{(i-1)})$.*

Theorem 6 (Error analysis of AC-MTI estimators). *Under Assumptions (2,7,8,9) and AC-MTI in Def. 7, the error of the estimators in Def. 9, denoted $\epsilon^{est} \coloneqq T^{est} - \mathbb{E}[Y|do(\mathbf{x})]$ for est $\in$ {reg, pw, dml}, are:*

$$\epsilon^{reg} = R_1 + O_{P_{x_1}}\left(\|\mu^1 - \mu_0^1\|\right),$$
$$\epsilon^{pw} = R_m + O_{P_{x_m}}(\|\pi^{(m-1)} - \pi_0^{(m-1)}\|),$$
$$\epsilon^{dml} = \sum_{i=1}^{m} R_i + \sum_{i=2}^{m} O_{P_{x_i}}\left(\|\mu^i - \mu_0^i\|\|\pi^{i-1} - \pi_0^{i-1}\|\right),$$

where R_i is a random variable such that $\sqrt{n_i}R_i$ converges in distribution to the mean-zero normal random variable, where $n_i \coloneqq |D_{x_i}|$ for $i \in \{1, \cdots, m\}$.

We highlight that the DML estimator T^{dml} in Def. 9 exhibits robustness property since the error ϵ^{dml} is bounded in probability at a rate $O_{P_{x_i}}(n^{-1/2})$ (for $n = \min\{n_1, n_2, \cdots, n_m\}$) rate whenever $\|\mu^i - \mu_0^i\|_{P_{x_i}} = O_{P_{x_i}}(n^{-1/4})$ and $\|\pi^{i-1} - \pi_0^{i-1}\|_{P_{x_i}} = O_{P_{x_i}}(n^{-1/4})$. Furthermore, the DML estimator displays the following multiply robustness property:

Corollary 6 (Multiply robustness of the DML estimators (Corollary of Thm. 6)**).** *Suppose Asumptions (2,7,8,9) and AC-MTI in Def. 7 hold. For $i = 2, \cdots, m-1$, suppose either $\pi^{i-1} = \pi_0^{i-1}$ or $\mu^i = \mu_0^i$. Then, T^{dml} in Def. 9 is an unbiased estimator of $\mathbb{E}[Y|do(\mathbf{x})]$.*

[3] Throughout the paper, $\mu^i, \overline{\mu}^i, \pi^i$ may be understood as estimated nuisances for $\mu_0^i, \overline{\mu}_0^i, \pi_0^i$.

4.2. Combining Multiple Arbitrary Experiments

In this section, we generalize Task MTI to the case where multiple treatment effects $\mathbb{E}[Y|do(\mathbf{x})]$ can be identified from arbitrary sets of experiments. We label this task as '*generalized multiple-teratment-interaction (gMTI)*':

Task gMTI (Generalized MTI). The task of *generaliazed MTI* composes of

- **Input**: Multiple sets of samples $\{D_i\}_{i=1}^m$ following distributions $\{P_{\text{rand}(Z_i)}(\mathbf{V})\}_{i=1}^m$.
- **Query**: Estimation of $\mathbb{E}[Y|do(\mathbf{x})]$.

We note that Task gMTI is a generalization of Task MTI since $\mathbf{X}$ is not necessarily identical to $\mathbf{Z}$.

4.2.1. Adjustment Criterion for gMTI

A graphical criterion for identifying the effect $\mathbb{E}[Y|do(\mathbf{x})]$ is the following:

Definition 10 (Adjustment criterion for gMTI (AC-gMTI)). Let $\mathbf{Z} := \{Z_1, \cdots, Z_m\} \subseteq \mathbf{X}$ denote the subset of treatments. Let $\{\ell_i\}_{i=1}^m \subseteq \{1, 2, \cdots, |\mathbf{X}|\}$ denote the index of $\mathbf{Z}$; i.e., $\mathbf{Z} = \{X_{\ell_1}, \cdots, X_{\ell_m}\}$. Let $\overline{X}_1 := \{X_j\}_{j \leq \ell_1}$, $\overline{X}_{m+1} := \{X_j\}_{j > \ell_m}$, and $\overline{X}_i := \{X_j\}_{\ell_{i-1} < j \leq \ell_i}$ for $i = 2, 3, \cdots, m$. An ordered set $\mathbf{A} := \{\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_m\}$ satisfies *adjustment criterion for combining multiple experiments (AC-gMTI)* w.r.t. $\{\mathbf{X}, Y\}$ in G if, for $i = 1, 2, \cdots, m-1$,

1. $(\mathbf{A}_i \perp\!\!\!\perp \overline{\mathbf{X}}^{>i-1} \backslash Z_i | \overline{\mathbf{X}}^{(i-1)}, \mathbf{A}^{(i-1)}, Z_i)_{G_{\overline{\overline{\mathbf{X}}^{>i-1}}}}$;
2. $(Y \perp\!\!\!\perp \overline{X}_i | \mathbf{A}^{(i)}, \overline{\mathbf{X}}^{(i-1)}, \overline{\mathbf{X}}^{>i})_{G_{\underline{\overline{X}_i}, \overline{\overline{\mathbf{X}}^{>i}}}}$; and
3. $(Y \perp\!\!\!\perp \overline{\mathbf{X}}^{\geq m} \backslash Z_m | \mathbf{A}^{(m-1)}, \overline{\mathbf{X}}^{(m-1)}, Z_m)_{G_{\overline{Z_m}, \underline{\overline{\mathbf{X}}^{\geq m} \backslash Z_m}}}$.

We make the following positivity assumption:

Assumption 10 (Positivity Assumption for AC-gMTI). $P_{z_m}(\overline{X}_m \backslash Z_m, \overline{X}_{m+1} | \mathbf{A}^{(m-1)}, \overline{\mathbf{X}}^{(m-1)})$ *and* $\{P_{z_i}(\mathbf{A}_i | \mathbf{A}^{(i-1)}, \overline{\mathbf{X}}^{(i-1)}), P_{z_{i+1}}(\mathbf{A}_i | \mathbf{A}^{(i-1)}, \overline{\mathbf{X}}^{(i-1)})\}_{i=1}^{m-1}$, $\{P^{i+1}(\overline{X}_i | \mathbf{A}^{(i)}, \overline{\mathbf{X}}^{(i-1)})\}_{i=1}^{m-1}$ *are strictly positive distributions* $\forall i \in \{1, \cdots, m\}, \forall z_i \in \mathfrak{D}_{Z_i}$.

Under AC-gMTI in Def. 10, $\mathbb{E}[Y|do(\mathbf{x})]$ is identifiable and given as follow:

Theorem 7 (Identification through AC-gMTI). *Suppose AC-gMTI in Def. 10 and Assumption 10 hold. Then,* $\mathbb{E}[Y|do(\mathbf{x})]$ *is identifiable from* $\{P_{\text{rand}(Z_i)}(\mathbf{A}^{(i)}, \overline{\mathbf{X}}^{(i)})\}_{i=1}^m$ *and given as follow:*

$$\mu_0^m := \mathbb{E}_{P_{z_m}}\left[Y | \mathbf{A}^{(m-1)}, \mathbf{X} \backslash Z_m\right]$$
$$\overline{\mu}_0^m := \mathbb{E}_{P_{z_m}}\left[Y | \mathbf{A}^{(m-1)}, \overline{\mathbf{x}}_{m-1:m+1}, \overline{\mathbf{X}}^{(m-2)}\right]$$
$$\mu_0^{m-1} := \mathbb{E}_{P_{z_{m-1}}}\left[\overline{\mu}_0^m \middle| \mathbf{A}^{(m-2)}, \overline{\mathbf{X}}^{(m-2)}\right],$$

where $\overline{X}_{m-1:m+1} := \{\overline{X}_{m-1}, \overline{X}_m, \overline{X}_{m+1}\}$. *For* $i = m-2, \cdots, 2$,

$$\mu_0^i := \mathbb{E}_{P_{z_i}}\left[\mu^{i+1}(\mathbf{A}^{(i)}, \overline{x}_i, \overline{\mathbf{X}}^{(i-1)}) \middle| \mathbf{A}^{(i-1)}, \overline{\mathbf{X}}^{(i-1)}\right],$$

and $\overline{\mu}_0^{i+1} := \mu_0^{i+1}(\mathbf{A}^{(i)}, \overline{x}_i, \overline{\mathbf{X}}^{(i-1)})$. *Then,*

$$\mathbb{E}[Y(\mathbf{x})] = \mathbb{E}_{P_{z_1}}\left[\overline{\mu}_0^2\right]. \tag{4}$$

For example, in Fig. 2b, $\{\mathbf{A}_1, \mathbf{A}_2\}$ where $\mathbf{A}_1 := \{C_1, W_1\}$ and $\mathbf{A}_2 := \{C_2, W_2\}$ satisfies AC-gMTI criterion in Def. 10 w.r.t. $\{\mathbf{X}, Y\}$ where $\mathbf{X} := \{X_0, Z_1, Z_2, Z_3\}$. Therefore, with the positivity in Assumption 10, $\mathbb{E}[Y|do(\mathbf{x})]$ is identifiable from $\{P_{z_i}(\mathbf{A}^{(i)}, \overline{\mathbf{X}}^{(i)})\}_{i=1}^m$ as in Eq. (4).

4.2.2. Estimators for AC-gMTI

We define nuisance functional for estimating the AC-gMTI functional in Eq. (4) as follow:

Definition 11 (Nuisances for AC-gMTI). Nuisance functions for AC-gMTI are defined as follows: For a fixed $\mathbf{z} := \{z_1, \cdots, z_m\} \in \mathfrak{D}_{\mathbf{Z}}$, let $\{\mu_0^i\}_{i=2}^m$ be the nuisances defined in Thm. 7. For $i = 1, \cdots, m-2$, $\pi_0^i := \frac{P_{z_i}(A_i | \mathbf{A}^{(i-1)}, \overline{\mathbf{X}}^{(i-1)})}{P_{z_m}(A_i, \overline{X}_i | \mathbf{A}^{(i-1)}, \overline{\mathbf{X}}^{(i-1)})}$, and $\pi_0^{(i)} := \prod_{j=1}^i \pi_0^j(\mathbf{A}^{(j)}, \overline{\mathbf{X}}^{(j)})$. Also, $\pi_0^{m-1} := \frac{P_{z_{m-1}}(A_{m-1} | \mathbf{A}^{(m-2)}, \overline{\mathbf{X}}^{(m-2)})}{P_{z_m}(A_{m-1}, \overline{X}_{m-1:m+1} | \mathbf{A}^{(m-2)}, \overline{\mathbf{X}}^{(m-2)})}$, and $pi_0^{(m-1)} := \pi_0^{(m-2)} \times \pi_0^{m-1}$, where $\overline{X}_{m-1:m+1} := \{\overline{X}_{m-1}, \overline{X}_m, \overline{X}_{m+1}\}$. For all $i = 1, 2, \cdots, m-1$, we will use $\pi^i(\mathbf{W}^{(i)}, \mathbf{C}^{(i)}, \mathbf{X}^{(i)}) > 0$ and μ^i and $\overline{\mu}^i$ to denote arbitrary finite functions.

Now, we construct regression-based ('REG'), probability weighting ('PW') and double/debiased machine learning ('DML') estimators:

Definition 12 (AC-gMTI estimators). Let D_i denote samples following $P_{\text{rand}(Z_i)}(\mathbf{V})$ for $i = 1, 2, \cdots, m$. For a fixed $z_i \in \mathfrak{D}_{Z_i}$, let D_{z_i} denote the subsamples of D_i such that $Z_i = z_i$. Let $\mu^{m+1} := Y$. Let $\mathbb{1}_{\mathbf{x}}^{i-1} := \mathbb{1}_{\overline{\mathbf{x}}^{(i-1)}}(\overline{\mathbf{X}}^{(i-1)})$. Then {REG, PW, DML} estimators are:

$$T^{reg} := \mathbb{E}_{D_{z_1}}\left[\mu^2(A_1, \overline{x}_1))\right],$$
$$T^{pw} := \mathbb{E}_{D_{z_m}}\left[\pi^{(m-1)}(\mathbf{A}^{(m-1)}, \mathbf{X}) \mathbb{1}_{\mathbf{x}}(\mathbf{X}) Y\right],$$
$$T^{dml} := \sum_{i=2}^m \mathbb{E}_{D_{z_i}}\left[\pi^{(i-1)} \mathbb{1}_{\mathbf{x}}^{i-1} \{\overline{\mu}^{i+1} - \mu^i\}\right] + \mathbb{E}_{D_{z_1}}\left[\overline{\mu}^2\right].$$

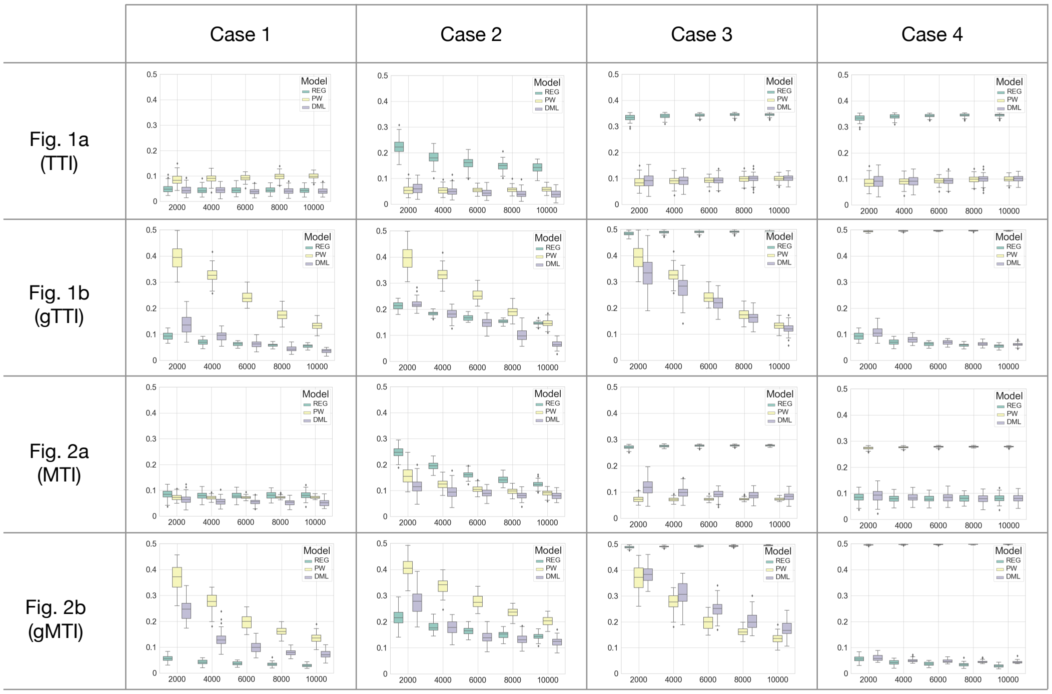


Figure 3: AAE Plots for Figs. (1a, 1b, 2a, 2b) for Cases {1,2,3,4} depicted in the Experimental Setup section. The x-axis and y-axis are the number of samples and AAE, respectively. Plots can be zoomed in.

We assume that nuisances can be estimated L_2 consistently.

Assumption 11 (**L_2 consistency of nuisances**). *Estimated nuisances $\{\mu^i\}_{i=2}^m$ and $\{\pi^i\}_{i=1}^{m-1}$ are L_2 consistent; Specifically,*

$$\|\mu^{i+1} - \mu_0^{i+1}\|_{P_{z_i}} = o_{P_{z_i}}(1), \ \forall i \in \{1, 2, \cdots, m-1\}$$
$$\|\mu^{i} - \mu_0^{i}\|_{P_{z_i}} = o_{P_{z_i}}(1), \ \forall i \in \{2, \cdots, m\}$$
$$\|\pi^{i} - \pi^{i}\|_{P_{z_{i+1}}} = o_{P_{z_{i+1}}}(1), \ \forall i \in \{1, \cdots, m-1\}.$$

Theorem 8 (**Error analysis of the AC-gMTI estimators**). *Under Assumptions (2,10,11) and AC-gMTI in Def. 10, the error of the estimators in Def. 12, denoted $\epsilon^{est} := T^{est} - \mathbb{E}[Y|do(\mathbf{x})]$ for est $\in \{reg, pw, dml\}$, are:*

$$\epsilon^{reg} = R_1 + O_{P_{z_1}}(\|\mu^1 - \mu_0^1\|),$$
$$\epsilon^{pw} = R_m + O_{P_{z_m}}(\|\pi^{(m-1)} - \pi_0^{(m-1)}\|),$$
$$\epsilon^{dml} = \sum_{i=1}^{m} R_i + \sum_{i=2}^{m} O_{P_{z_i}}(\|\mu^i - \mu_0^i\|\|\pi^{i-1} - \pi_0^{i-1}\|),$$

where R_i is a variable such that $\sqrt{n_i}R_i$ converges in distribution to the mean-zero normal random variable, where $n_i := |D_i|$ for $i \in \{1, \cdots, m\}$.

We highlight that the DML estimator T^{dml} in Def. 12 exhibits robustness property since the error ϵ^{dml} is bounded in probability at a rate $O_{P_{z_i}}(n^{-1/2})$ (for $n = \min\{n_1, n_2, \cdots, n_m\}$) rate whenever $\|\mu^i - \mu_0^i\|_{P_{z_i}} = O_{P_{z_i}}(n^{-1/4})$ and $\|\pi^{i-1} - \pi_0^{i-1}\|_{P_{z_i}} = O_{P_{z_i}}(n^{-1/4})$. Furthermore, the DML estimator displays the following multiply robustness property:

Corollary 8 (**Multiply robustness of the DML estimators** (Corollary of Thm. 8)). *Suppose Assumptions (2,10,11) and AC-gMTI in Def. 10 hold. For $i = 2, \cdots, m-1$, suppose either $\pi^{i-1} = \pi_0^{i-1}$ or $\mu^i = \mu_0^i$. Then, T^{dml} in Def. 12 is an unbiased estimator of $\mathbb{E}[Y|do(\mathbf{x})]$.*

5. Experiments

In this section, we demonstrated the proposed estimators in Defs. (3,6) for combining two experiments presented in Sec. 3, and Defs. (9, 12) for combining multiple experiments in Sec. 4. Details of the experiments and a different simulation example are provided in Appendice E.

Accuracy Measure. For each Tasks (TTI, gTTI, MTI, gMTI) and given n samples $D_1, \cdots, D_m$, we will use $T^{\text{est}}(\mathbf{x})$ for est $\in \{\text{reg}, \text{pw}, \text{dml}\}$ be the $\{\text{REG}, \text{PW}, \text{DML}\}$ estimator for the joint treatment effects $\mathbb{E}[Y|do(\mathbf{x})]$. For each est $\in \{\text{reg}, \text{pw}, \text{dml}\}$, we assess the quality of the estimator by computing the *average absolute error* $\text{AAE}^{\text{est}} := \frac{1}{|\mathcal{X}|}\sum_{\mathbf{x}\in\mathcal{X}} |T^{\text{est}}(\mathbf{x}) - \mathbb{E}[Y|do(\mathbf{x})]|$ where $\mathcal{X}$ denote the set of all possible values for $\mathbf{x}$ and $|\mathcal{X}|$ its cardinality. Nuisance functions are estimated using gradient boosting models called XGBoost (Chen & Guestrin, 2016). We ran 100 simulations for each $N = \{2000, 4000, 6000, 8000, 10000\}$. We label the box-plot for these AAEs as 'AAE-plot'.

Experimental Setup. We measure the AAE^{est} for each four scenarios: **(Case 1)** there were no noises nor misspecification in estimating nuisances; **(Case 2)**; The 'converging noise' ϵ, decaying at a $N^{-\alpha}$ rate (i.e., $\epsilon \sim \text{Normal}(N^{-\alpha}, N^{-2\alpha})$) for $\alpha = 1/4$, was added in estimating nuisances; **(Case 3)** Nuisances $\{\mu^i\}_{i=2}^{m}$ are wrongly estimated; **(Case 4)** $\{\pi^i\}_{i=1}^{m-1}$ are wrongly estimated. Case 2 is a scenario to highlight fast convergence property of the DML estimator implied in Thms. (2,4,6,8) where T^{dml} converges faster ($n^{-1/2}$-rate) when other estimators $T^{\text{reg}}, T^{\text{pw}}$ converge at $n^{-1/4}$-rate. Cases {3,4} are scenarios highlighting doubly robustness property of the DML estimator formalized in Corolaries (2,4,6,8).

Experimental Results. The AAE plots for all cases are presented in Fig. 3. All {DML, REG, IPW} estimators converges in Case 1 as the sample size grows. In Case 2 where the estimated nuisances are controlled to converge at $n^{-1/4}$ rate, the DML estimators T^{dml} outperform the other two estimators by achieving a fast convergence. This result corroborates the robustness property in Thms. (2,4,6,8). In Cases (3,4) where the estimated nuisances for $\{\mu^i\}_{i=2}^{m}$ or $\{\pi^i\}_{i=1}^{m-1}$ are wrongly specified, the DML estimator T^{dml} converges while other estimators fail to converge. This result corroborates the doubly robustness property in Coros. (2,4,6,8).

Project STAR Dataset. To provide empirical evidence in a real-world setting, beyond the analysis of the synthetic dataset, we applied the proposed estimators to Project STAR dataset (Krueger & Whitmore, 2001; Schanzenbach, 2006). Project STAR is an experimental study investigating teacher/student ratios' impact on academic achievement for kindergarten through third-grade students. More detailed information on the analysis and comprehensive results are provided in Appendix D. In summary, the proposed DML estimator T^{dml} exhibits fast convergence and doubly robustness properties in real-world scenarios.

6. Conclusions

We introduced a set of identification conditions for estimating joint causal effects by combining multiple marginal experiments (Thms. (1,3,5,7)), developed corresponding estimators (Defs. (3,6,9,12)), and analyzed their statistical properties (Thms. (2,4,6,8) and Coros. (2,4,6,8)) for various Task (TTI,gTTI,MTI,gMTI). Our experimental results demonstrate that the proposed estimators are consistent estimators of the joint effect $\mathbb{E}[Y|do(\mathbf{x})]$. Additionally, the proposed DML estimators were found to be robust against model misspecification and slow convergence rate in learning nuisances. We hope this work can help data scientists to estimate joint treatment effects from multiple experiments in a more principled and efficient manner.

Acknowledgement

We thank Iván Díaz and the reviewers for their valuable feedback and assistance in improving this paper. Elias Bareinboim and Yonghan Jung were supported in part by funding from the Alfred P. Sloan Foundation, NSF, ONR, AFOSR, DoE, Amazon, and JP Morgan. Jin Tian was partially supported by NSF grant IIS-2231797.

References

Ajjan, R. A. and Grant, P. J. Cardiovascular disease prevention in patients with type 2 diabetes: The role of oral anti-diabetic agents. *Diabetes and Vascular Disease Research*, 3(3):147–158, 2006.

Athey, S., Chetty, R., Imbens, G. W., and Kang, H. The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely. Technical report, National Bureau of Economic Research, 2019.

Athey, S., Chetty, R., and Imbens, G. Combining experimental and observational data to estimate treatment effects on long term outcomes. *arXiv preprint arXiv:2006.09676*, 2020.

Bareinboim, E. and Pearl, J. Causal inference by surrogate experiments: z-identifiability. In *In Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence*, pp. 113–120. AUAI Press, 2012a.

Bareinboim, E. and Pearl, J. Transportability of causal effects: Completeness results. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, pp. 698–704, 2012b.

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

Bareinboim, E., Correa, J. D., Ibeling, D., and Icard, T. On pearl's hierarchy and the foundations of causal inference. In *Probabilistic and causal inference: the works of judea pearl*, pp. 507–556. 2022.

Bhattacharya, R., Nabi, R., and Shpitser, I. Semiparametric inference for causal effects in graphical models with hidden variables. *Journal of Machine Learning Research*, 23:1–76, 2022.

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.

Colnet, B., Mayer, I., Chen, G., Dieng, A., Li, R., Varoquaux, G., Vert, J.-P., Josse, J., and Yang, S. Causal inference methods for combining randomized trials and observational studies: a review. *arXiv preprint arXiv:2011.08047*, 2020.

Correa, J., Lee, S., and Bareinboim, E. Nested counterfactual identification from arbitrary surrogate experiments. *Advances in Neural Information Processing Systems*, 34, 2021.

Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, 96(1):187–199, 2009.

Dahabreh, I. J., Robertson, S. E., Tchetgen, E. J., Stuart, E. A., and Hernán, M. A. Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals. *Biometrics*, 75(2):685–694, 2019.

Degtiar, I. and Rose, S. A review of generalizability and transportability. *arXiv preprint arXiv:2102.11904*, 2021.

Egami, N. and Imai, K. Causal interaction in factorial experiments: Application to conjoint analysis. *Journal of the American Statistical Association*, 2018.

Ferrannini, E. and Cushman, W. C. Diabetes and hypertension: the bad companions. *The Lancet*, 380(9841): 601–610, 2012.

Gazzard, B., Clark, R., Borirakchanyavat, V., and Williams, R. A controlled trial of heparin therapy in the coagulation defect of paracetamol-induced hepatic necrosis. *Gut*, 15 (2):89–93, 1974.

Gentzel, A. M., Pruthi, P., and Jensen, D. How and why to use experimental data to evaluate methods for observational causal inference. In *International Conference on Machine Learning*, pp. 3660–3671. PMLR, 2021.

Hansson, L., Zanchetti, A., Carruthers, S. G., Dahlöf, B., Elmfeldt, D., Julius, S., Ménard, J., Rahn, K. H., Wedel, H., Westerling, S., et al. Effects of intensive blood-pressure lowering and low-dose aspirin in patients with hypertension: principal results of the hypertension optimal treatment (hot) randomised trial. *The Lancet*, 351 (9118):1755–1762, 1998.

Hansson, L., Lindholm, L. H., Ekbom, T., Dahlöf, B., Lanke, J., Scherstén, B., Wester, P., Hedner, T., de Faire, U., Group, S.-H.-. S., et al. Randomised trial of old and new antihypertensive drugs in elderly patients: cardiovascular mortality and morbidity the swedish trial in old patients with hypertension-2 study. *The Lancet*, 354(9192):1751–1756, 1999.

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

Imbens, G., Kallus, N., Mao, X., and Wang, Y. Long-term causal inference under persistent confounding via data combination. *arXiv preprint arXiv:2202.07234*, 2022.

Jung, Y., Tian, J., and Bareinboim, E. Learning causal effects via weighted empirical risk minimization. *Advances in Neural Information Processing Systems*, 33, 2020.

Jung, Y., Tian, J., and Bareinboim, E. Estimating identifiable causal effects on markov equivalence class through double machine learning. In *Proceedings of the 38th International Conference on Machine Learning*, 2021a.

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

Jung, Y., Kasiviswanathan, S., Tian, J., Janzing, D., Blöbaum, P., and Bareinboim, E. On measuring causal contributions via do-interventions. In *International Conference on Machine Learning*, pp. 10476–10501. PMLR, 2022.

Kennedy, E. H., Balakrishnan, S., G'Sell, M., et al. Sharp instruments for classifying compliers and generalizing causal effects. *Annals of Statistics*, 48(4):2008–2030, 2020.

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

Krueger, A. B. and Whitmore, D. M. The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project star. *The Economic Journal*, 111(468):1–28, 2001.

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

Lee, S., Correa, J. D., and Bareinboim, E. General identifiability with arbitrary surrogate experiments. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence*. AUAI Press, 2019.

Lee, S., Correa, J., and Bareinboim, E. General transportability–synthesizing observations and experiments from heterogeneous domains. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 10210–10217, 2020.

Lesko, S. M. and Mitchell, A. A. An assessment of the safety of pediatric ibuprofen: a practitioner-based randomized clinical trial. *Jama*, 273(12):929–933, 1995.

Louizos, C., Shalit, U., Mooij, J. M., Sontag, D., Zemel, R., and Welling, M. Causal effect inference with deep latent-variable models. *Advances in neural information processing systems*, 30, 2017.

Moore, N., Pollack, C., and Butkerait, P. Adverse drug reactions and drug–drug interactions with over-the-counter nsaids. *Therapeutics and clinical risk management*, 11: 1061, 2015.

Parbhoo, S., Bauer, S., and Schwab, P. Ncore: Neural counterfactual representation learning for combinations of treatments. *arXiv preprint arXiv:2103.11175*, 2021.

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

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.

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.

Saengkyongam, S. and Silva, R. Learning joint nonlinear effects from single-variable interventions in the presence of hidden confounders. In *Conference on Uncertainty in Artificial Intelligence*, pp. 300–309. PMLR, 2020.

Schanzenbach, D. W. What have researchers learned from project star? *Brookings papers on education policy*, (9): 205–228, 2006.

Shi, X., Pan, Z., and Miao, W. Data integration in causal inference. *Wiley Interdisciplinary Reviews: Computational Statistics*, pp. e1581, 2022.

Stock, J. H., Watson, M. W., et al. *Introduction to econometrics*, volume 104. Addison Wesley Boston, 2003.

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, 20(8):467–484, 2019.

VanderWeele, T. J. and Knol, M. J. A tutorial on interaction. *Epidemiologic methods*, 3(1):33–72, 2014.

Zhang, J. and Bareinboim, E. Near-optimal reinforcement learning in dynamic treatment regimes. *Advances in Neural Information Processing Systems*, 32, 2019.

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

Supplement to "Estimating Joint Treatment Effects by Combining Multiple Experiments"

Contents

A. Preliminaries

In this section, we present the preliminary concepts and notation used in this paper. Let $\mathbf{W}$ and $\mathbf{X}$ be sets of variables that are subsets of $\mathbf{V}$, which are induced from the structural causal model (SCM) $\mathcal{M}$. Given a realization $\mathbf{x}$ of $\mathbf{X}$, we denote $\mathbf{W}(\mathbf{x})$ as the counterfactual $\mathbf{W}$ under the hypothetical scenario where $\mathbf{X}$ takes the value $\mathbf{x}$. In other words, $\mathbf{W}(\mathbf{x})$ represents a random vector generated by the submodel $\mathcal{M}_{\mathbf{x}}$.

A.1. The Axioms of Structural Counterfactuals

Definition A.1 (The Axioms of Structural Counterfactuals (Pearl, 2000, Chapter 7.3.1)). For any three sets of endogenous variables $\mathbf{X}, \mathbf{Y}, \mathbf{W}$ in a causal model and $\mathbf{x}, \mathbf{w} \in \mathfrak{D}_{\mathbf{X},\mathbf{W}}$, the following holds:

- **Composition**: $\mathbf{W}(\mathbf{x}) = \mathbf{w} \implies \mathbf{Y}(\mathbf{x}, \mathbf{w}) = \mathbf{Y}(\mathbf{x})$.
- **Effectiveness**: $\mathbf{X}(\mathbf{w}, \mathbf{x}) = \mathbf{x}$.
- **Reversibility**: $\mathbf{Y}(\mathbf{x}, \mathbf{w}) = \mathbf{y}$ and $\mathbf{W}(\mathbf{x}, \mathbf{y}) = \mathbf{w} \implies \mathbf{Y}(\mathbf{x}) = \mathbf{y}$.

Theorem A.1 (Soundness and Completeness of the Axioms (Pearl, 2000, Theorems {7.3.3, 7.3.6})). *The Axioms of structural counterfactuals in Def. A.1 are sound and complete for all causal models.*

Remark 1 ((Pearl, 2000, page 230)). *In the recursive (acyclic) system, Reversibility is followed from Composition. Therefore, Composition and Effectiveness are sound and complete.*

Definition A.2 (Potential Response, Counterfactuals (Pearl, 2000, Def. 7.1.4)). Let $(\mathbf{X}, \mathbf{Y}) \subseteq \mathbf{V}$ generated by the SCM $\mathcal{M}$. The counterfactual of $\mathbf{Y}$ at $\mathbf{x}$, denoted $\mathbf{Y}(\mathbf{x})$, is the variable $\mathbf{Y}$ induced by the submodel $\mathcal{M}_{\mathbf{x}}$.

In this section, we will use $P_{\text{rand}(\mathbf{X})}(\mathbf{V}) \coloneqq \{P(\mathbf{V}(\mathbf{x}))\}_{\mathbf{x} \in \mathfrak{D}_{\mathbf{X}}}$ to denote a collection of counterfactual distributions $P(\mathbf{V}(\mathbf{x}))$ over all possible realizations $\mathbf{x} \in \mathfrak{D}_{\mathbf{X}}$. We will denote the density of P as p.

B. Identification based on Potential Outcome Framework

In this section, we introduce more results on identifying causal effects. For $\mathbf{x} \in \mathfrak{D}_{\mathbf{X}}$, we use $(\mathbf{W}\backslash\mathbf{X})(\mathbf{x})$ to denote the counterfactual of $\mathbf{W}\backslash\mathbf{X}$ at $\mathbf{x}$. We use $(\mathbf{w}\backslash\mathbf{x})(\mathbf{x})$ to denote its realization.

B.1. Treatment-Treatment Interaction based on Potential Outcome Framework

We present a sufficient identification criterion for estimating treatment-treatment interactions using potential outcome frameworks based on two marginal experiments.

Definition B.1 (Adjustment criterion for treatment-treatment-interaction – Potential Outcome (AC-TTI-PO)). A set of variables $\{C_1, W\}$ is said to satisfy the adjustment criterion for treatment-treatment interaction (AC-TTI) w.r.t. discrete treatments (X_1, X_2) and the outcome Y from two sets of distributions $P_{\text{rand}(X_1)}(C_1, X_1, W)$ and $P_{\text{rand}(X_2)}(C_1, X_1, W, X_2, Y)$ if

1. $W(x_1, x_2) = W(x_1)$, $C_1(x_1, x_2) = C_1(x_1)$; i.e., the outcome W and the covariate C_1 is invariant of the second intervention $X_2 = x_2$.
2. $Y(x_1, x_2) \perp\!\!\!\perp X_1(x_2) | W_1(x_1, x_2), C_1(x_1, x_2)$; i.e., the first intervention $X_1 = x_1$ is non-informative to the joint experimental outcome $Y(x_1, x_2)$ given covariates $C_1(x_1, x_2)$ and the first outcome $W(x_1, x_2)$.

The treatment-treatment interaction can be identified as follow:

Theorem B.1 (Identification through AC2-TTI-PO). *Suppose the condition AC-TTI-PO in Def. B.1 holds. For any fixed $x_1, x_2 \in \mathfrak{D}_{X_1,X_2}$, define $P^1 \coloneqq P_{x_1} \in P_{rand(X_1)}$ and $P^2 \coloneqq P_{x_2} \in P_{rand(X_2)}$. Assume the following positivity condition* *holds for* $\forall x_1, x_2, w, c_1 \in \mathfrak{D}_{X_1,X_2,W,C_1}$*:*

$$\frac{p^2(w, c_1)}{p^2(w, c_1)} p^2(X_1 = x_1 | w, c_1) > 0. \tag{B.1}$$

Then, the query $\mathbb{E}[Y(x_1, x_2)]$ *is identifiable from two distributions* P^1, P^2 *and given as follow:*

$$\mathbb{E}[Y(x_1, x_2)] = \mathbb{E}_{P^1}[\mathbb{E}_{P^2}[Y | W, C_1, X_1 = x_1]]. \tag{B.2}$$

Proof of Theorem B.1.

$$\begin{aligned}
\mathbb{E}[Y(x_1, x_2)] &= \mathbb{E}[\mathbb{E}[Y(x_1, x_2) | W(x_1, x_2), C_1(x_1, x_2)]] \\
&\stackrel{1}{=} \mathbb{E}[\mathbb{E}[Y(x_1, x_2) | W(x_1, x_2), C_1(x_1, x_2), X_1(x_2) = x_1]] \\
&\stackrel{2}{=} \mathbb{E}[\mathbb{E}[Y(x_1, x_2) | W(x_1), C_1(x_1), X_1(x_2) = x_1]] \\
&\stackrel{3}{=} \mathbb{E}[\mathbb{E}[Y(x_2) | W(x_1), C_1(x_1), X_1(x_2) = x_1]] \\
&\stackrel{4}{=} \mathbb{E}_{P^1}[\mathbb{E}[Y(x_2) | W, C_1, X_1(x_2) = x_1]],
\end{aligned}$$

where

- $\stackrel{1}{=}$ holds by the given condition that $Y(x_1, x_2) \perp\!\!\!\perp X_1(x_2) | W(x_1, x_2), C_1(x_1, x_2)$ and the positivity $P(X_1(x_2) = x_1 | W(x_1, x_2), C_1(x_1, x_2)) > 0$ for any x_1, x_2. To witness, it suffices to show that $p_{X_1(x_2)|W(x_1,x_2),C_1(x_1,x_2)}(x_1 | w, c_1) > 0$.

$$\begin{aligned}
&p_{X_1(x_2)|W(x_1,x_2),C_1(x_1,x_2)}(x_1 | w, c_1) \\
&= \frac{p_{X_1(x_2),W(x_1,x_2),C_1(x_1,x_2)}(x_1, w, c_1)}{p_{W(x_1,x_2),C_1(x_1,x_2)}(w, c_1)} \\
&\stackrel{1a}{=} \frac{p_{X_1(x_2),W(x_1,x_2),C_1(x_1,x_2)}(x_1, w, c_1)}{p_{W(x_1),C_1(x_1)}(w, c_1)} \\
&\stackrel{1b}{=} \frac{p_{X_1(x_2),W(x_2),C_1(x_2)}(x_1, w, c_1)}{p_{W(x_1),C_1(x_1)}(w, c_1)} \\
&= \frac{p_{X_1(x_2),W(x_2),C_1(x_2)}(x_1, w, c_1)}{p_{W(x_1),C_1(x_1)}(w, c_1)} \frac{p_{W(x_2),C_1(x_2)}(w, c_1)}{p_{W(x_2),C_1(x_2)}(w, c_1)} \\
&= \frac{p_{W(x_2),C_1(x_2)}(w, c_1)}{p_{W(x_1),C_1(x_1)}(w, c_1)} p_{X_1(x_2)|W(x_2),C_1(x_2)}(x_1 | w, c_1) \\
&= \frac{p^2_{W,C_1}(w, c_1)}{p^1_{W,C_1}(w, c_1)} p^2_{X_1|W,C_1}(x_1 | w, c_1) \\
&\stackrel{1c}{>} 0,
\end{aligned}$$

  where

  - $\stackrel{1a}{=}$ holds by the first condition of the AC-TTI-PO in Def. B.1, stating that $W(x_1, x_2) = W(x_1)$ and $C_1(x_1, x_2) = C_1(x_1)$.
  - $\stackrel{1b}{=}$ holds by the Composition axiom in Def. A.1. Specifically, $X_1(x_2) = x_1$ implies $W(x_1, x_2) = W(x_2)$ and $C_1(x_1, x_2) = C_1(x_2)$.
  - $\stackrel{1c}{>}$ holds by the given assumption.

- $\stackrel{2}{=}$ holds since $W(x_1, x_2) = W(x_1)$ and $C_1(x_1, x_2) = C_1(x_1)$ by the first condition of the AC-TTI-PO in Def. B.1.

- $\overset{3}{=}$ holds by the Composition axiom in Def. A.1. Specifically, $X_1(x_2) = x_1$ implies $Y(x_1, x_2) = Y(x_2)$.

- $\overset{4}{=}$ by the definition of P^1.

We note that $\mathbb{E}\left[Y(x_2)|W(x_1), C_1(x_1), X_1(x_2) = x_1\right]$ is estimable from $P^2(Y|W, C_1, X_1)$ since

$$\begin{aligned} P^2(Y|W, C_1, X_1 = x_1) &\overset{5}{=} P(Y(x_2)|W(x_2), C_1(x_2), X_1(x_2) = x_1) \\ &\overset{6}{=} P(Y(x_2)|W(x_1, x_2), C_1(x_1, x_2), X_1(x_2) = x_1) \\ &\overset{7}{=} P(Y(x_2)|W(x_1), C_1(x_1), X_1(x_2) = x_1), \end{aligned}$$

where

- $\overset{5}{=}$ holds by the definition of P^2.

- $\overset{6}{=}$ holds since $W(x_1, x_2) = W(x_2)$ and $C_1(x_1, x_2) = C_1(x_2)$ when $X_1(x_2) = x_1$ by Composition axiom in Def. A.1.

- $\overset{7}{=}$ holds by the first condition of the AC-TTI-PO in Def. B.1, stating that $W(x_1, x_2) = W(x_1)$ and $C_1(x_1, x_2) = C_1(x_1)$.

Therefore,

$$\mathbb{E}_{P^1}\left[\mathbb{E}\left[Y(x_2)|W, C_1, X_1(x_2) = x_1\right]\right] = \mathbb{E}_{P^1}\left[\mathbb{E}_{P^2}\left[Y|W, C_1, X_1 = x_1\right]\right].$$

□

B.2. Combining Two Experiments based on Potential Outcome Framework

We provide an adjustment criterion based on potential outcome frameworks for combining two experiments as follow:

Definition B.2 (Adjustment criterion for combining two experiments – Potential Outcome (AC2-PO)). A set of variables $\mathbf{A}$ is said to satisfy the adjustment criterion (AC2) w.r.t discrete treatments $\mathbf{X}$ and the outcome Y from two sets of distributions $P_{\text{rand}(Z_1)}(\mathbf{A})$ and $P_{\text{rand}(Z_2)}(\mathbf{A}, \mathbf{X}, Y)$ if

1. $\mathbf{A}(\mathbf{x}) = \mathbf{A}(z_1)$; i.e., $Z_1 \subseteq \mathbf{X}$ and $\mathbf{X} \backslash Z_1$ is causally irrelevant to $\mathbf{A}$ given Z_1;
2. $Z_2 \subseteq \mathbf{X}$ and $Y(\mathbf{x}) \perp\!\!\!\perp (\mathbf{X} \backslash Z_2)(z_2)|\mathbf{A}(\mathbf{x})$.

Under Def. B.2, the causal effect is identified as follows:

Theorem B.2 (Identification through AC2-PO). *Suppose the condition AC2-PO in Def. B.2 holds. Let*

$$\begin{aligned} P^1(\mathbf{A}) &\coloneqq P(\mathbf{A}(z_1)) \\ P^2(Y, \mathbf{X} \backslash Z_2, \mathbf{A}) &\coloneqq P(Y(z_2), (\mathbf{X} \backslash Z_2)(z_2), \mathbf{A}(z_2)), \end{aligned}$$

and p^1, p^2 are densities for distributions P^1, P^2. Assume the following positivity condition:

$$\frac{p^2(\mathbf{a})}{p^1(\mathbf{a})} p^2(\mathbf{X} \backslash Z_2 = \mathbf{x} \backslash z_2 | \mathbf{a}),\ \forall \mathbf{x}, \mathbf{a} \in \mathcal{X} \times \mathcal{A}. \tag{B.3}$$

Then, the query $\mathbb{E}\left[Y(\mathbf{x})\right]$ is identifiable from two distributions P^1, P^2 and given as follow:

$$\mathbb{E}\left[Y(\mathbf{x})\right] = \mathbb{E}_{P^1}\left[\mathbb{E}_{P^2}\left[Y|\mathbf{A}, \mathbf{x} \backslash z_2\right]\right]. \tag{B.4}$$

Proof of Theorem B.2.

$$
\begin{aligned}
\mathbb{E}\left[Y(\mathbf{x})\right] &= \mathbb{E}\left[\mathbb{E}\left[Y(\mathbf{x})|\mathbf{A}(\mathbf{x})\right]\right] \\
&\overset{1}{=} \mathbb{E}\left[\mathbb{E}\left[Y(\mathbf{x})|\mathbf{A}(\mathbf{x}), (\mathbf{X}\backslash Z_2)(z_2) = \mathbf{x}\backslash z_2\right]\right] \\
&\overset{2}{=} \mathbb{E}\left[\mathbb{E}\left[Y(\mathbf{x})|\mathbf{A}(z_1), (\mathbf{X}\backslash Z_2)(z_2) = \mathbf{x}\backslash z_2\right]\right] \\
&\overset{3}{=} \mathbb{E}\left[\mathbb{E}\left[Y(z_2)|\mathbf{A}(z_1), (\mathbf{X}\backslash Z_2)(z_2) = \mathbf{x}\backslash z_2\right]\right] \\
&\overset{4}{=} \mathbb{E}_{P_1}\left[\mathbb{E}\left[Y(z_2)|\mathbf{A}, (\mathbf{X}\backslash Z_2)(z_2) = \mathbf{x}\backslash z_2\right]\right],
\end{aligned}
$$

where

- $\overset{1}{=}$ holds since $Y(\mathbf{x}) \perp\!\!\!\perp (\mathbf{X}\backslash Z_2)(z_2)|\mathbf{A}(\mathbf{x})$, and the positivity $P((\mathbf{X}\backslash Z_2)(z_2)|\mathbf{A}(\mathbf{x})) > 0$ for any $\mathbf{x}$. To witness, it suffices to show that $p_{(\mathbf{X}\backslash Z_2)(z_2)|\mathbf{A}(\mathbf{x})}(\mathbf{x}\backslash z_2|\mathbf{a}) > 0$.

$$
\begin{aligned}
p_{(\mathbf{X}\backslash Z_2)(z_2)|\mathbf{A}(\mathbf{x})}(\mathbf{x}\backslash z_2|\mathbf{a}) &= \frac{p_{\mathbf{X}\backslash Z_2,\mathbf{A}(\mathbf{x})}(\mathbf{x}\backslash z_2, \mathbf{a})}{p_{\mathbf{A}(\mathbf{x})}(\mathbf{a})} \\
&\overset{1a}{=} \frac{p_{(\mathbf{X}\backslash Z_2)(z_2),\mathbf{A}(\mathbf{x})}(\mathbf{x}\backslash z_2, \mathbf{a})}{p_{\mathbf{A}(z_1)}(\mathbf{a})} \\
&\overset{1b}{=} \frac{p_{(\mathbf{X}\backslash Z_2)(z_2),\mathbf{A}(z_2)}(\mathbf{x}\backslash z_2, \mathbf{a})}{p_{\mathbf{A}(z_1)}(\mathbf{a})} \\
&= \frac{p_{(\mathbf{X}\backslash Z_2)(z_2),\mathbf{A}(z_2)}(\mathbf{x}\backslash z_2, \mathbf{a})}{p_{\mathbf{A}(z_1)}(\mathbf{a})} \frac{p_{\mathbf{A}(z_2)}(\mathbf{a})}{p_{\mathbf{A}(z_2)}(\mathbf{a})} \\
&= \frac{p_{\mathbf{A}(z_2)}(\mathbf{a})}{p_{\mathbf{A}(z_1)}(\mathbf{a})} p_{(\mathbf{X}\backslash Z_2)(z_2)|\mathbf{A}(z_2)}(\mathbf{x}\backslash z_2|\mathbf{a}) \\
&= \frac{p^2(\mathbf{a})}{p^1(\mathbf{a})} p^2(\mathbf{x}\backslash z_2|\mathbf{a}) \\
&\overset{1c}{>} 0,
\end{aligned}
$$

  where

  - $\overset{1a}{=}$ holds by the first condition of the AC2-PO in Def. B.2, stating that $\mathbf{A}(\mathbf{x}) = \mathbf{A}(z_1)$.
  - $\overset{1b}{=}$ holds by the Composition axiom in Def. A.1. Specifically, $(\mathbf{X}\backslash Z_2)(z_2) = \mathbf{x}\backslash z_2$ implies $\mathbf{A}(\mathbf{x}) = \mathbf{A}(z_2)$.
  - $\overset{1c}{>}$ holds by the given assumption.

- $\overset{2}{=}$ holds since $\mathbf{A}(\mathbf{x}) = \mathbf{A}(z_2)$.
- $\overset{3}{=}$ holds by the Composition axiom in Def. A.1. Specifically, $(\mathbf{X}\backslash Z_2)(z_2) = \mathbf{x}\backslash \mathbf{z}_2$ implies $Y(\mathbf{x}) = Y(z_2)$.
- $\overset{4}{=}$ holds by the definition of P^1.

We note that $\mathbb{E}\left[Y(z_2)|\mathbf{A}(z_1), (\mathbf{X}\backslash Z_2)(z_2) = \mathbf{x}\backslash z_2\right]$ is estimable from $P_2(Y|\mathbf{A}, \mathbf{X}\backslash Z_2)$ since

$$
\begin{aligned}
P_2(Y|\mathbf{A}, \mathbf{X}\backslash Z_2 = \mathbf{x}\backslash z_2) &\overset{5}{=} P(Y(z_2)|\mathbf{A}(z_2), (\mathbf{X}\backslash Z_2)(z_2) = \mathbf{x}\backslash z_2) \\
&\overset{6}{=} P(Y(z_2)|\mathbf{A}(\mathbf{x}), (\mathbf{X}\backslash Z_2)(z_2) = \mathbf{x}\backslash z_2) \\
&\overset{7}{=} P(Y(z_2)|\mathbf{A}(z_1), (\mathbf{X}\backslash Z_2)(z_2) = \mathbf{x}\backslash z_2),
\end{aligned}
$$

where

- $\overset{5}{=}$ holds by the definition.

- $\stackrel{6}{=}$ holds since $\mathbf{A}(\mathbf{x})$ when $(\mathbf{X}\backslash Z_2)(z_2) = \mathbf{x}\backslash z_2$ by Composition axiom in Def. A.1.
- $\stackrel{7}{=}$ holds since $\mathbf{A}(z_1) = \mathbf{A}(\mathbf{x})$.

Therefore,

$$\mathbb{E}_{P^1}\left[\mathbb{E}\left[Y(x_2)|\mathbf{A}, (\mathbf{X}\backslash Z_2)(z_2) = \mathbf{x}\backslash z_2\right]\right] = \mathbb{E}_{P^1}\left[\mathbb{E}_{P^2}\left[Y|\mathbf{A}, \mathbf{x}\backslash z_2\right]\right].$$

□

B.3. Multiple Treatment Interaction based on Potential Outcome Framework

We first provide a sufficient identification criterion based on potential outcome frameworks for estimating multiple treatment interaction from multiple marginal experiments.

Definition B.3 (Adjustment Criterion for MTI – Potential Outcome (AC-MTI-PO)). A set of variables $\{\mathbf{C}, \mathbf{W}\}$ is said to satisfy the adjustment criterion for multiple treatment interaction (AC-MTI-PO) w.r.t. discrete treatments $\mathbf{x} = \{x_i\}_{i=1}^m$ and the outcome Y from multiple distributions $\{P_{x_i}(\mathbf{V})\}_{i=1}^m$ if

1. $W_i(\mathbf{x}) = W_i(\mathbf{x}^{(i)})$, $C_i(\mathbf{x}) = C_i(\mathbf{x}^{(i)})$ for $i = 1, 2, \cdots, m-1$; i.e., the ith joint outcome $W_i(\mathbf{x})$ and the covariate $C_i(\mathbf{x})$ are invariant to the next interventions $X_{i+1}, \cdots, X_m$.
2. For all $i = 1, 2, \cdots, m-1$, $X_i = X_i(\mathbf{x}^{(i+1:k)})$, $\forall k \in \{i+1, \cdots, m\}$; i.e., X_i is invariant to any intervention $X_k = x_k$ for $k > i$.
3. $Y(\mathbf{x}) \perp\!\!\!\perp X_i | \mathbf{C}^{(i)}(\mathbf{x}), \mathbf{X}^{(i-1)}, \mathbf{W}^{(i)}(\mathbf{x})$ for $i = 1, 2, \cdots, m-1$; i.e., the ith intervention $X_i = x_i$ is non-informative to the joint outcome $Y(\mathbf{x})$ given the ith outcome $W_i(\mathbf{x})$, covariate $C_i(\mathbf{x})$ and previous observations $\mathbf{X}^{(i-1)}, \mathbf{W}^{(i-1)}, \mathbf{C}^{(i-1)}$.

Under Def. B.3, the causal effect $\mathbb{E}\left[Y(\mathbf{x})|\mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{X}^{(i-1)}\right]$ for $i = 1, 2, \cdots, m-1$ can be expressed in a recursive form as follow:

Lemma B.1. *Suppose the condition AC-MTI-PO in Def. B.3 holds. Let $A_i := \{W_i, C_i\}$. Assume the following positivity condition holds: For $i = 1, 2, \cdots, m-1$*

$$p_{X_i|\mathbf{A}^{(i)}(\mathbf{x}), \mathbf{X}^{(i-1)}}(x_i|\mathbf{a}^{(i)}, \mathbf{x}^{(i-1)}) > 0, \ \forall \mathbf{a}, \mathbf{x} \in \mathfrak{D}_{\mathbf{A}, \mathbf{X}}. \tag{B.5}$$

Then,

$$\mathbb{E}\left[Y(\mathbf{x})\middle|\mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{X}^{(i-1)}\right] = \mathbb{E}\left[\mathbb{E}\left[Y(\mathbf{x})|\mathbf{A}^{(i)}(\mathbf{x}), X_i = x_i, \mathbf{X}^{(i-1)}\right]|\mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{X}^{(i-1)}\right],$$

where $\mathbf{A}_0 := \emptyset$ and $X_0 := \emptyset$.

Proof of Lemma B.1.

$$\begin{aligned}\mathbb{E}\left[Y(\mathbf{x})\middle|\mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{X}^{(i-1)}\right] &\stackrel{1}{=} \mathbb{E}\left[\mathbb{E}\left[Y(\mathbf{x})\middle|\mathbf{A}^{(i)}(\mathbf{x}), \mathbf{X}^{(i-1)}\right]\middle|\mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{X}^{(i-1)}\right] \\ &\stackrel{2}{=} \mathbb{E}\left[\mathbb{E}\left[Y(\mathbf{x})\middle|\mathbf{A}^{(i)}(\mathbf{x}), X_i = x_i, \mathbf{X}^{(i-1)}\right]\middle|\mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{X}^{(i-1)}\right],\end{aligned}$$

where

- $\stackrel{1}{=}$ holds by marginalizing over $\mathbf{A}_i(\mathbf{x})$.

- $\stackrel{2}{=}$ holds by the condition in the AC-MTI-PO; i.e., $Y(\mathbf{x}) \perp\!\!\!\perp X_i | \mathbf{A}^{(i)}(\mathbf{x}), \mathbf{X}^{(i-1)}$ for $i = 1, 2, \cdots, m-1$, and the given positivity condition in Eq. (B.5).

$\square$

Corollary B.1 (Corollary of Lemma B.1). *Suppose the condition AC-MTI-PO in Def. B.3 holds. Let $A_i := \{W_i, C_i\}$. Assume the positivity condition given in Eq. (B.5). Let*

$$\nu_0^i(\mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{X}^{(i-1)}) := \mathbb{E}\left[Y(\mathbf{x}) | \mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{X}^{(i-1)}\right], \text{ for } i = m, m-1, \cdots, 2.$$

Then,

$$\nu_0^i(\mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{X}^{(i-1)}) = \mathbb{E}\left[\nu_0^{i+1}(\mathbf{A}^{(i)}(\mathbf{x}), x_i, \mathbf{X}^{(i-1)}) \middle| \mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{X}^{(i-1)}\right] \text{ for } i = m-1, \cdots, 2,$$

where $\mathbf{A}_0 := \emptyset$ and $X_0 := \emptyset$. Furthermore,

$$\mathbb{E}\left[Y(\mathbf{x})\right] = \mathbb{E}\left[\nu_0^2(\mathbf{A}^{(1)}(\mathbf{x}), x_1)\right].$$

Proof of Corollary B.1. We first note that the equations

$$\nu_0^i(\mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{X}^{(i-1)}) := \mathbb{E}\left[Y(\mathbf{x}) | \mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{X}^{(i-1)}\right], \text{ for } i = m, m-1, \cdots, 2,$$

is immediately followed by Lemma B.1. Therefore, it only suffices to show the following:

$$\mathbb{E}\left[Y(\mathbf{x})\right] = \mathbb{E}\left[\nu_0^2(\mathbf{A}^{(1)}(\mathbf{x}), x_1)\right].$$

To witness,

$$\begin{aligned}\mathbb{E}\left[Y(\mathbf{x})\right] &= \mathbb{E}\left[\mathbb{E}\left[Y(\mathbf{x}) | \mathbf{A}^{(1)}(\mathbf{x})\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[Y(\mathbf{x}) | \mathbf{A}^{(1)}(\mathbf{x}), X_1 = x_1\right]\right] \\ &= \mathbb{E}\left[\nu_0^2(\mathbf{A}^{(1)}(\mathbf{x}), x_1)\right],\end{aligned}$$

where the second equation holds by the condition in the AC-MTI-PO; i.e., $Y(\mathbf{x}) \perp\!\!\!\perp X_i | \mathbf{A}^{(i)}(\mathbf{x}), \mathbf{X}^{(i-1)}$ for $i = 1, 2, \cdots, m-1$, and the given positivity condition in Eq. (B.5). $\square$

Lemma B.2. *Suppose the condition AC-MTI-PO in Def. B.3 holds. Let $A_i := \{W_i, C_i\}$. Assume the positivity condition given in Eq. (B.5). For $i = m, \cdots, 1$, and*

$$\nu_0^i(\mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{X}^{(i-1)}) := \mathbb{E}\left[Y(\mathbf{x}) | \mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{X}^{(i-1)}\right].$$

For $i = 1, 2, \cdots, m$, let P^i denote a distribution defined as follow:

$$P^i(\mathbf{W}^{(i)}, \mathbf{C}^{(i)}, \mathbf{X}^{(i)}) := P(\mathbf{W}^{(i)}(x_i), \mathbf{C}^{(i)}(x_i), \mathbf{X}^{(i)}(x_i)),$$

and $p^1, \cdots, p^{m-1}$ are densities for distributions $P^1, \cdots, P^m$. Let

$$\mu_0^m(\mathbf{A}^{(m-1)}, \mathbf{X}^{(m-1)}) := \mathbb{E}_{P^m}\left[Y | \mathbf{A}^{(m-1)}, \mathbf{X}^{(m-1)}\right],$$

and for $i = m-1, \cdots, 1$*,*

$$\mu_0^i(\mathbf{A}^{(i-1)}, \mathbf{X}^{(i-1)}) \coloneqq \mathbb{E}_{P^i}\left[\mu_0^{i+1}(\mathbf{A}^{(i)}, x_i, \mathbf{X}^{(i-1)}) \Big| \mathbf{A}^{(i-1)}, \mathbf{X}^{(i-1)}\right].$$

Then, for $i = m, \cdots, 1$*,*

$$\mu_0^i(\mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{x}^{(i-1)}) = \nu_0^i(\mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{x}^{(i-1)}).$$

Proof of Lemma B.2. We first show that

$$\mu_0^m(\mathbf{A}^{(m-1)}, \mathbf{x}^{(m-1)}) = \nu_0^m(\mathbf{A}^{(m-1)}(\mathbf{x}), \mathbf{x}^{(m-1)}).$$

To witness,

$$\begin{aligned}
\nu_0^m(\mathbf{A}^{(m-1)}(\mathbf{x}), \mathbf{x}^{(m-1)}) &\coloneqq \mathbb{E}\left[Y(\mathbf{x}) \big| \mathbf{A}^{(m-1)}(\mathbf{x}), \mathbf{X}^{(m-1)} = \mathbf{x}^{(m-1)}\right] \\
&\overset{1}{=} \mathbb{E}\left[Y(\mathbf{x}) \big| \mathbf{A}^{(m-1)}(\mathbf{x}), \mathbf{X}^{(m-1)}(x_m) = \mathbf{x}^{(m-1)}\right] \\
&\overset{2}{=} \mathbb{E}\left[Y(x_m) | \mathbf{A}^{(m-1)}(\mathbf{x}), \mathbf{X}^{(m-1)}(x_m) = \mathbf{x}^{(m-1)}\right] \\
&\overset{3}{=} \mathbb{E}\left[Y(x_m) | \mathbf{A}^{(m-1)}(x_m), \mathbf{X}^{(m-1)}(x_m) = \mathbf{x}^{(m-1)}\right] \\
&= \mathbb{E}_{P^m}\left[Y | \mathbf{A}^{(m-1)}, \mathbf{X}^{(m-1)} = \mathbf{x}^{(m-1)}\right] \\
&\eqqcolon \mu_0^m(\mathbf{A}^{(m-1)}, \mathbf{x}^{(m-1)}),
\end{aligned}$$

where

- $\overset{1}{=}$ holds since $\mathbf{X}^{(m-1)}(x_m) = \mathbf{X}^{(m-1)}$ by the condition of the AC-MTI-PO in Def. B.3 stating that treatment variables are invariant to the next interventions.

- $\overset{2}{=}$ holds since

$$\mathbf{X}^{(m-1)}(x_m) = \mathbf{x}^{(m-1)} \implies Y(\mathbf{x}) = Y(\mathbf{x}^{(m-1)}, x_m) = Y(x_m),$$

  by Composition axiom in Axiom A.1.

- $\overset{3}{=}$ holds since

$$\mathbf{X}^{(m-1)}(x_m) = \mathbf{x}^{(m-1)} \implies \mathbf{A}^{(m-1)}(\mathbf{x}) = \mathbf{A}^{(m-1)}(\mathbf{x}^{(m-1)}, x_m) = \mathbf{A}^{(m-1)}(x_m),$$

  by Composition axiom in Axiom A.1.

We now make an induction hypothesis as follow: For any given $i \in \{2, \cdots, m-1\}$ suppose the following holds:

$$\mu_0^{i+1}(\mathbf{A}^{(i)}, \mathbf{x}^{(i)}) = \nu_0^{i+1}(\mathbf{A}^{(i)}(\mathbf{x}), \mathbf{x}^{(i)}).$$

Then,

$$
\begin{aligned}
\nu_0^i(\mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{x}^{(i-1)}) &\overset{4}{=} \mathbb{E}\left[\nu_0^{i+1}(\mathbf{A}^{(i)}(\mathbf{x}), x_i, \mathbf{X}^{(i-1)}) \middle| \mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{X}^{(i-1)} = \mathbf{x}^{(i-1)}\right] \\
&= \mathbb{E}\left[\nu_0^{i+1}(\mathbf{A}^{(i)}(\mathbf{x}), \mathbf{x}^{(i)}) \middle| \mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{x}^{(i-1)}\right] \\
&\overset{5}{=} \mathbb{E}\left[\mu_0^{i+1}(\mathbf{A}^{(i)}, \mathbf{x}^{(i)}) \middle| \mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{x}^{(i-1)}\right] \\
&\overset{6}{=} \mathbb{E}\left[\mu_0^{i+1}(\mathbf{A}^{(i)}(x_i), \mathbf{x}^{(i)}) \middle| \mathbf{A}^{(i-1)}(x_i), \mathbf{x}^{(i-1)}(x_i)\right] \\
&= \mathbb{E}_{P^i}\left[\mu_0^{i+1}(\mathbf{A}^{(i)}, \mathbf{x}^{(i)}) | \mathbf{A}^{(i-1)}, \mathbf{x}^{(i-1)}\right] \\
&= \mu_0^i(\mathbf{A}^{(i-1)}, \mathbf{x}^{(i-1)}),
\end{aligned}
$$

where

- $\overset{4}{=}$ implies from Corollary B.1.
- $\overset{5}{=}$ because of the induction hypothesis.
- $\overset{6}{=}$ holds because $\mathbf{X}^{(i-1)} = \mathbf{X}^{(i-1)}(x_i)$ by the Composition Axiom in Def. A.1, and

  $$\mathbf{X}^{(i-1)}(x_i) = \mathbf{x}^{(i-1)} \implies \mathbf{A}^{(i-1)}(\mathbf{x}) = \mathbf{A}^{(i-1)}(\mathbf{x}^{(i-1)}) = \mathbf{A}^{(i-1)}(\mathbf{X}^{(i-1)}(x_i) = \mathbf{x}^{(i-1)}, x_i) = \mathbf{A}^{(i-1)}(x_i)$$

  by applying the Composition Axiom in Def. A.1.

This proves that, for $i = m, m-1, \cdots, 1$,

$$\mu_0^i(\mathbf{A}^{(i-1)}, \mathbf{x}^{(i-1)}) = \nu_0^i(\mathbf{A}^{(i-1)}(\mathbf{x}), \mathbf{x}^{(i-1)}).$$

□

Theorem B.3 (Identification through AC-MTI-PO). *Suppose the condition AC-MTI-PO in Def. B.3 holds. For $i = 1, 2, \cdots, m$, let P^i denote a distribution defined as follow:*

$$P^i(\mathbf{W}^{(i)}, \mathbf{C}^{(i)}, \mathbf{X}^{(i)}) \coloneqq P(\mathbf{W}^{(i)}(x_i), \mathbf{C}^{(i)}(x_i)), \mathbf{X}^{(i)}(x_i)),$$

and $p^1, \cdots, p^{m-1}$ are densities for distributions $P^1, \cdots, P^m$. Assume the following positivity condition holds: For all $i = 1, 2, \cdots, m-1$

$$\frac{p^{i+1}(w_i, c_i|\mathbf{w}^{(i-1)}, \mathbf{c}^{(i-1)}, \mathbf{x}^{(i-1)})}{p^i(w_i, c_i|\mathbf{w}^{(i-1)}, \mathbf{c}^{(i-1)}, \mathbf{x}^{(i-1)})} p^{i+1}(X_i = x_i|\mathbf{w}^{(i)}, \mathbf{c}^{(i)}, \mathbf{x}^{(i-1)}) > 0;\ \forall \mathbf{w}, \mathbf{c}, \mathbf{x} \in \mathfrak{D}_{\mathbf{W},\mathbf{C},\mathbf{X}}. \tag{B.6}$$

Then, the query $\mathbb{E}[Y(\mathbf{x})]$ is identifiable from distributions $P^1, \cdots, P^m$, and given as follow: Let

$$\mu_0^m(\mathbf{A}^{(m-1)}, \mathbf{X}^{(m-1)}) \coloneqq \mathbb{E}_{P^m}\left[Y|\mathbf{A}^{(m-1)}, \mathbf{X}^{(m-1)}\right],$$

and for $i = m-1, m-2, \cdots, 2$,

$$\mu_0^i(\mathbf{A}^{(i-1)}, \mathbf{X}^{(i-1)}) \coloneqq \mathbb{E}_{P^i}\left[\mu_0^{i+1}(\mathbf{A}^{(i)}, x_i, \mathbf{X}^{(i-1)})|\mathbf{A}^{(i-1)}, \mathbf{X}^{(i-1)}\right].$$

Then,

$$\mathbb{E}[Y(\mathbf{x})] = \mathbb{E}_{P^1}\left[\mu_0^2(\mathbf{A}^{(1)}, \mathbf{x}^{(1)})\right].$$

Proof of Theorem B.3. Suppose the positivity condition in Eq. (B.5) is equivalent to the condition in Eq. (B.6). Then, Theorem B.3 is implied by Lemmas (B.1, B.2) and Corollary B.1.

The equivalence between Eq. (B.5) and Eq. (B.6) are the following. We will use $A_i \coloneqq \{W_i, C_i\}$. Then,

$$
\begin{aligned}
& p_{X_i|\mathbf{A}^{(i)}(\mathbf{x}),\mathbf{X}^{(i-1)}}(x_i|\mathbf{a}^{(i)},\mathbf{x}^{(i-1)}) \\
&= \frac{p_{X_i,A_i(\mathbf{x})|\mathbf{A}^{(i-1)}(\mathbf{x}),\mathbf{X}^{(i-1)}}(x_i,a_i|\mathbf{a}^{(i-1)},\mathbf{x}^{(i-1)})}{p_{A_i(\mathbf{x})|\mathbf{A}^{(i-1)}(\mathbf{x}),\mathbf{X}^{(i-1)}}(a_i|\mathbf{a}^{(i-1)},\mathbf{x}^{(i-1)})} \\
&= \frac{p_{X_i(x_{i+1}),A_i(x_{i+1})|\mathbf{A}^{(i-1)}(x_{i+1}),\mathbf{X}^{(i-1)}(x_{i+1})}(x_i,a_i|\mathbf{a}^{(i-1)},\mathbf{x}^{(i-1)})}{p_{A_i(x_i)|\mathbf{A}^{(i-1)}(x_i),\mathbf{X}^{(i-1)}(x_i)}(a_i|\mathbf{a}^{(i-1)},\mathbf{x}^{(i-1)})} \\
&= \frac{p_{X_i(x_{i+1}),A_i(x_{i+1})|\mathbf{A}^{(i-1)}(x_{i+1}),\mathbf{X}^{(i-1)}(x_{i+1})}(x_i,a_i|\mathbf{a}^{(i-1)},\mathbf{x}^{(i-1)})}{p_{A_i(x_i)|\mathbf{A}^{(i-1)}(x_i),\mathbf{X}^{(i-1)}(x_i)}(a_i|\mathbf{a}^{(i-1)},\mathbf{x}^{(i-1)})} \frac{p_{A_i(x_{i+1})|\mathbf{A}^{(i-1)}(x_{i+1}),\mathbf{X}^{(i-1)}(x_{i+1})}(a_i|\mathbf{a}^{(i-1)},\mathbf{x}^{(i-1)})}{p_{A_i(x_{i+1})|\mathbf{A}^{(i-1)}(x_{i+1}),\mathbf{X}^{(i-1)}(x_{i+1})}(a_i|\mathbf{a}^{(i-1)},\mathbf{x}^{(i-1)})} \\
&= \frac{p_{A_i(x_{i+1})|\mathbf{A}^{(i-1)}(x_{i+1}),\mathbf{X}^{(i-1)}(x_{i+1})}(a_i|\mathbf{a}^{(i-1)},\mathbf{x}^{(i-1)})}{p_{A_i(x_i)|\mathbf{A}^{(i-1)}(x_i),\mathbf{X}^{(i-1)}(x_i)}(a_i|\mathbf{a}^{(i-1)},\mathbf{x}^{(i-1)})} p_{X_i(x_{i+1})|\mathbf{A}^{(i)}(x_{i+1}),\mathbf{X}^{(i-1)}(x_{i+1})}(x_i|\mathbf{a}^{(i)},\mathbf{x}^{(i-1)}) \\
&= \frac{p^{i+1}(w_i,c_i|\mathbf{w}^{(i-1)},\mathbf{c}^{(i-1)},\mathbf{x}^{(i-1)})}{p^{i}(w_i,c_i|t\mathbf{w}^{(i-1)},\mathbf{c}^{(i-1)},\mathbf{x}^{(i-1)})} p^{i+1}(X_i = x_i|\mathbf{w}^{(i)},\mathbf{c}^{(i)},\mathbf{x}^{(i-1)}).
\end{aligned}
$$

□

B.4. Combining Multiple Experiments based on Potential Outcome Framework

We provide an adjustment criterion based on potential outcome frameworks for combining two experiment as follow:

Definition B.4 (Adjustment criterion for combining multiple experiments – Potential Outcome (AC-gMTI-PO)). Let $\mathbf{X} \coloneqq \{X_1, \cdots, X_{m_x}\}$ and Y denote an ordered treatments and outcome variables. Let $\mathbf{Z} \coloneqq \{Z_1, \cdots, Z_m\} \subseteq \mathbf{X}$ denote the subset of treatments. Let $\{\ell_i\}_{i=1}^{m} \subseteq \{1, 2, \cdots, m_x\}$ denote the index of $\mathbf{Z}$; i.e., $\mathbf{Z} = \{X_{\ell_1}, \cdots, X_{\ell_m}\}$. Let $\overline{X}_1 \coloneqq \{X_j\}_{j \leq \ell_1}$, $\overline{X}_{m+1} \coloneqq \{X_j\}_{j > \ell_m}$, and $\overline{X}_i \coloneqq \{X_j\}_{\ell_{i-1} < j \leq \ell_i}$ for $i = 2, 3, \cdots, m$. A set of topologically ordered variables $\mathbf{A} \coloneqq \{\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_m\}$ is said to satisfy the adjustment criterion for combining multiple experiments (AC-gMTI-PO) w.r.t. treatments $\mathbf{x} = \{x_i\}_{i=1}^{m_x}$ and the outcome Y from multiple distributions $\{P_{z_i}(\mathbf{A}^{(i)}, \overline{\mathbf{X}}^{(i)})\}_{i=1}^{m}$ if

1. $\mathbf{A}_i(\mathbf{x}) = \mathbf{A}_i(\mathbf{z}^{(i)})$ for $i = 1, 2, \cdots, m-1$ and $\overline{X}_i(z_j) = \overline{X}_i$ for all $i.j \in \{1, 2, \cdots, m\}$ where $i \leq j$.
2. $Y(\mathbf{x}) \perp\!\!\!\perp \overline{X}_i | \mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{X}}^{(i-1)}$ for $i = 1, 2, \cdots, m-1$
3. $Y(\mathbf{x}) \perp\!\!\!\perp \{\overline{X}_m \backslash Z_m, \overline{X}_{m+1}(z_m)\} | \mathbf{A}^{(m-1)}(\mathbf{x}), \overline{\mathbf{X}}^{(m-1)}$.

We first note that the causal effect $\mathbb{E}\left[Y(\mathbf{x})\right]$ can be represented in a recursive form as follow:

Lemma B.3. *Suppose the condition AC-gMTI-PO in Def. B.4 holds. Assume the following positivity condition holds: For* $\forall \overline{x}_m, \overline{x}_{m+1}, \mathbf{a}^{(m-1)}, \overline{\mathbf{x}}^{(m-1)} \in \mathfrak{D}_{\overline{X}_m, \overline{X}_{m+1}, \mathbf{A}^{(m-1)}, \overline{\mathbf{X}}^{(m-1)}}$,

$$p_{\overline{X}_m \backslash Z_m, \overline{X}_{m+1}(z_m) | \mathbf{A}^{(m-1)}(\mathbf{x}), \overline{\mathbf{X}}^{(m-1)}}(\overline{x}_m \backslash z_m, \overline{x}_{m+1} | \mathbf{a}^{(m-1)}, \overline{\mathbf{x}}^{(m-1)}) > 0, \tag{B.7}$$

and for $i = 1, 2, \cdots, m-1$

$$p_{\overline{X}_i | \mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{X}}^{(i-1)}}(\overline{x}_i | \mathbf{a}^{(i)}, \overline{\mathbf{x}}^{(i-1)}) > 0, \; \forall \mathbf{a}^{(i)}, \mathbf{x}^{(i)} \in \mathcal{A}^{(i)} \times \mathcal{X}^{(i)}. \tag{B.8}$$

Let

$$
\begin{aligned}
\nu_0^m(\mathbf{A}^{(m-1)}(\mathbf{x}), \mathbf{x} \backslash z_m) &\coloneqq \mathbb{E}\left[Y(\mathbf{x}) | \mathbf{A}^{(m-1)}(\mathbf{x}), (\mathbf{X} \backslash Z_m)(z_m) = \mathbf{x} \backslash z_m\right] \\
\nu_0^{m-1}(\mathbf{A}^{(m-2)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-2)}) &\coloneqq \mathbb{E}\left[\nu_0^m(\mathbf{A}^{(m-1)}(\mathbf{x}), \mathbf{x} \backslash z_m) \middle| \mathbf{A}^{(m-2)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-2)}\right],
\end{aligned}
$$

and for $i = m-2, \cdots, 2$,

$$\nu_0^i(\mathbf{A}^{(i-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(i-1)}) := \mathbb{E}\left[\nu_0^{i+1}(\mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{x}}^{(i)}) \middle| \mathbf{A}^{(i-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(i-1)}\right].$$

Then,

$$\mathbb{E}\left[Y(\mathbf{x})\right] = \mathbb{E}\left[\nu^2(A_1(\mathbf{x}), \overline{x}_1)\right].$$

Proof of Lemma B.3. Let

$$\eta_0^{i+1}(\mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{x}}^{(i)}) := \mathbb{E}\left[Y(\mathbf{x}) | \mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{x}}^{(i)}\right], \quad \text{for } i = 1, 2, \cdots, m-1.$$

Then, the causal effect can be written as

$$\mathbb{E}\left[Y(\mathbf{x})\right] = \mathbb{E}\left[\eta_0^2(A_1(\mathbf{x}), \overline{x}_1)\right],$$

since

$$\mathbb{E}\left[Y(\mathbf{x})\right] = \mathbb{E}\left[\mathbb{E}\left[Y(\mathbf{x})|A_1(\mathbf{x})\right]\right] = \mathbb{E}\left[\mathbb{E}\left[Y(\mathbf{x})|\overline{x}_1, A_1(\mathbf{x})\right]\right] = \mathbb{E}\left[\eta^2(A_1(\mathbf{x}), \overline{x}_1)\right],$$

where the second equation holds since $Y(\mathbf{x}) \perp\!\!\!\perp \overline{X}_i | \mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{X}}^{(i-1)}$ for $i = 1, 2, \cdots, m-1$ by Def. B.4 and the positivity condition in Eq. (B.8).

We will show the following:

$$\begin{aligned} \eta_0^m(\mathbf{A}^{(m-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-1)}) &= \nu_0^m(\mathbf{A}^{(m-1)}(\mathbf{x}), \mathbf{x} \backslash z_m) \\ \eta_0^{i+1}(\mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{x}}^{(i)}) &= \nu_0^{i+1}(\mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{x}}^{(i)}) \ \text{ for } i = m-2, \cdots, 1. \end{aligned}$$

First equation can be witnessed by

$$\begin{aligned} \eta_0^m(\mathbf{A}^{(m-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-1)}) &= \mathbb{E}\left[Y(\mathbf{x}) | \mathbf{A}^{(m-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-1)}\right] \\ &\overset{2}{=} \mathbb{E}\left[Y(\mathbf{x}) | \mathbf{A}^{(m-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-1)}, \{\overline{X}_m \backslash Z_m = \overline{x}_m \backslash z_m, \overline{X}_{m+1}(z_m) = \overline{x}_{m+1}\}\right] \\ &\overset{3}{=} \mathbb{E}\left[Y(\mathbf{x}) | \mathbf{A}^{(m-1)}(\mathbf{x}), (\mathbf{X} \backslash Z_m)(z_m) = \mathbf{x} \backslash z_m\right], \\ &= \nu_0^m(\mathbf{A}^{(m-1)}(\mathbf{x}), \mathbf{x} \backslash z_m), \end{aligned}$$

where

- $\overset{2}{=}$ holds by $Y(\mathbf{x}) \perp\!\!\!\perp \{\overline{X}_m \backslash Z_m, \overline{X}_{m+1}(z_m)\} | \mathbf{A}^{(m-1)}(\mathbf{x}), \overline{\mathbf{X}}^{(m-1)}$ in Def. B.4 and the positivity condition in Eq. (B.7).

- $\overset{3}{=}$ holds since $\overline{X}_i(z_j) = \overline{X}_i$ for all i, j, as given in Def. B.4.

The second equation can be witnessed as follow:

$$\begin{aligned}
\eta_0^{m-1}(\mathbf{A}^{(m-2)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-2)}) &:= \mathbb{E}\left[Y(\mathbf{x})\middle|\mathbf{A}^{(m-2)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-2)}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[Y(\mathbf{x})|\mathbf{A}^{(m-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-2)}\right]\middle|\mathbf{A}^{(m-2)}(\mathbf{x}), \overline{\mathbf{X}}^{(m-2)}\right] \\
&\overset{4}{=} \mathbb{E}\left[\mathbb{E}\left[Y(\mathbf{x})|\mathbf{A}^{(m-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-1)}\right]\middle|\mathbf{A}^{(m-2)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-2)}\right] \\
&= \mathbb{E}\left[\eta_0^{m}(\mathbf{A}^{(m-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-1)})\middle|\mathbf{A}^{(m-2)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-2)}\right] \\
&\overset{5}{=} \mathbb{E}\left[\nu^{m}(\mathbf{A}^{(m-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-1)})\middle|\mathbf{A}^{(m-2)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-2)}\right] \\
&= \nu_0^{m-1}(\mathbf{A}^{(m-2)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-2)}),
\end{aligned}$$

where

- $\overset{4}{=}$ holds since $Y(\mathbf{x}) \perp\!\!\!\perp \overline{X}_i|\mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{X}}^{(i-1)}$ for $i = 1, 2, \cdots, m-1$ by Def. B.4 and the positivity condition in Eq. (B.8).
- $\overset{5}{=}$ holds since $\eta^m(\mathbf{A}^{(m-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-1)}) = \nu_0^m(\mathbf{A}^{(m-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-1)})$.

Now, we make an induction hypothesis as follow: For any $i+1 \in \{m-1\cdots, 3\}$, suppose the following holds:

$$\eta_0^{i+1}(\mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{x}}^{(i)}) = \nu_0^{i+1}(\mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{x}}^{(i)}).$$

Then,

$$\begin{aligned}
\eta_0^{i}(\mathbf{A}^{(i-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(i-1)}) &:= \mathbb{E}\left[Y(\mathbf{x})|\mathbf{A}^{(i-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(i-1)}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[Y(\mathbf{x})|\mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{x}}^{(i-1)}\right]\middle|\mathbf{A}^{(i-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(i-1)}\right] \\
&\overset{7}{=} \mathbb{E}\left[\mathbb{E}\left[Y(\mathbf{x})|\mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{x}}^{(i)}\right]\middle|\mathbf{A}^{(i-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(i-1)}\right] \\
&= \mathbb{E}\left[\eta_0^{i+1}(\mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{x}}^{(i)})\middle|\mathbf{A}^{(i-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(i-1)}\right] \\
&\overset{8}{=} \mathbb{E}\left[\nu_0^{i+1}(\mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{x}}^{(i)})\middle|\mathbf{A}^{(i-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(i-1)}\right] \\
&= \nu_0^{i}(\mathbf{A}^{(i-1)}(\mathbf{x}), \overline{\mathbf{X}}^{(i-1)}, \overline{x}_i \backslash z_i),
\end{aligned}$$

where

- $\overset{7}{=}$ holds since $Y(\mathbf{x}) \perp\!\!\!\perp \overline{X}_i|\mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{X}}^{(i-1)}$ for $i = 1, 2, \cdots, m-1$ by Def. B.4 and the positivity condition in Eq. (B.8).
- $\overset{8}{=}$ by induction hypothesis.

Therefore, $\eta^i = \nu^i$ for all $i = 1, 2, \cdots, m$. This completes the proof. □

Theorem B.4 (Identification through AC-gMTI-PO). *Suppose the condition AC-gMTI-PO in Def. B.4 holds. For $i = 1, 2, \cdots, m$, let P^i denote a distribution defined as follow:*

$$\begin{aligned}
P^i(\mathbf{A}^{(i)}, \overline{\mathbf{X}}^{(i)} \backslash Z_i) &:= P(\mathbf{A}^{(i)}(z_i), (\overline{\mathbf{X}}^{(i)} \backslash Z_i)(z_i)), \;\; \textit{for } i = 1, 2, \cdots, m-1 \\
P^m(\mathbf{A}^{(m)}, \mathbf{X} \backslash Z_m, Y) &:= P(\mathbf{A}^{(m)}(z_m), (\mathbf{X} \backslash Z)(z_m), Y(z_m)).
\end{aligned}$$

Assume the following positivity condition holds: For $\forall \overline{x}_m, \overline{x}_{m+1}, \mathbf{a}^{(m-1)}, \overline{\mathbf{x}}^{(m-1)} \in \mathfrak{D}_{\overline{X}_m, \overline{X}_{m+1}, \mathbf{A}^{(m-1)}, \overline{\mathbf{X}}^{(m-1)}}$*,*

$$p^m_{\overline{X}_m \setminus Z_m, \overline{X}_{m+1} | \mathbf{A}^{(m-1)}, \overline{\mathbf{X}}^{(m-1)}}(\overline{x}_m \setminus z_m, \overline{x}_{m+1} | \mathbf{a}^{(m-1)}, \overline{\mathbf{x}}^{(m-1)}) > 0, \tag{B.9}$$

and for $i = 1, 2, \cdots, m-1$*,*

$$\frac{p^{i+1}(a_i | \mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)})}{p^{i}(a_i | \mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)})} p^{i+1}(\overline{x}_i | \mathbf{a}^{(i)}, \overline{\mathbf{x}}^{(i-1)}) > 0, \ \forall \mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i)} \in \overline{\mathcal{X}}^{(i)} \times \mathcal{A}^{(i-1)}. \tag{B.10}$$

Then, the query $\mathbb{E}\left[Y(\mathbf{x})\right]$ *is identifiable from distributions* $P^1, \cdots, P^m$*, and given as follow: Let*

$$\mu_0^m(\mathbf{A}^{(m-1)}, \mathbf{x} \setminus z_m) := \mathbb{E}_{P^m}\left[Y | \mathbf{A}^{(m-1)}, \mathbf{x} \setminus z_m\right]$$
$$\mu_0^{m-1}(\mathbf{A}^{(m-2)}, \overline{\mathbf{x}}^{(m-2)}) := \mathbb{E}_{P^{m-1}}\left[\mu_0^m(\mathbf{A}^{(m-1)}, \mathbf{x} \setminus z_m) \middle| \mathbf{A}^{(m-2)}, \overline{\mathbf{x}}^{(m-2)}\right],$$

and for $i = m-2, \cdots, 2$

$$\mu_0^i(\mathbf{A}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)}) := \mathbb{E}_{P^i}\left[\mu^{i+1}(\mathbf{A}^{(i)}, \overline{\mathbf{x}}^{(i)}) \middle| \mathbf{A}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)}\right].$$

Then,

$$\mathbb{E}\left[Y(\mathbf{x})\right] = \mathbb{E}_{P^1}\left[\mu^2(A_1, \overline{x}_1)\right].$$

Proof of Theorem B.4. We first show that the positivity conditions in Eqs (B.7, B.8) match with Eqs. (B.9, B.10).

Eq. (B.7) = Eq. (B.9) holds since

$$\begin{aligned}
\text{Eq. (B.7)} &= p_{(\overline{X}_m \setminus Z_m(z_m), \overline{X}_{m+1}(z_m) | \mathbf{A}^{(m-1)}(\mathbf{x}), \overline{\mathbf{X}}^{(m-1)}}(\overline{x}_m \setminus z_m, \overline{x}_{m+1} | \mathbf{a}^{(m-1)}, \overline{\mathbf{x}}^{(m-1)}) \\
&= p_{(\overline{X}_m \setminus Z_m(z_m), \overline{X}_{m+1}(z_m) | \mathbf{A}^{(m-1)}(\overline{\mathbf{x}}^{(m-1)}), \overline{\mathbf{X}}^{(m-1)}}(\overline{x}_m \setminus z_m, \overline{x}_{m+1} | \mathbf{a}^{(m-1)}, \overline{\mathbf{x}}^{(m-1)}) \\
&= p_{(\overline{X}_m \setminus Z_m(z_m), \overline{X}_{m+1}(z_m) | \mathbf{A}^{(m-1)}, \overline{\mathbf{X}}^{(m-1)}}(\overline{x}_m \setminus z_m, \overline{x}_{m+1} | \mathbf{a}^{(m-1)}, \overline{\mathbf{x}}^{(m-1)}) \\
&= p_{(\overline{X}_m \setminus Z_m(z_m), \overline{X}_{m+1}(z_m) | \mathbf{A}^{(m-1)}(z_m), \overline{\mathbf{X}}^{(m-1)}(z_m)}(\overline{x}_m \setminus z_m, \overline{x}_{m+1} | \mathbf{a}^{(m-1)}, \overline{\mathbf{x}}^{(m-1)}) \\
&= \text{Eq. (B.9)}.
\end{aligned}$$

Eq. (B.8) = Eq. (B.10) holds since

$$
\begin{aligned}
&\text{Eq. (B.8)}\\
&= \frac{p_{\overline{X}_i, A_i(\mathbf{x})|\mathbf{A}^{(i-1)}(\mathbf{x}),\overline{\mathbf{X}}^{(i-1)}}(\overline{x}_i, a_i|\mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)})}{p_{A_i(\mathbf{x})|\mathbf{A}^{(i-1)}(\mathbf{x}),\overline{\mathbf{X}}^{(i-1)}}(a_i|\mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)})}\\
&= \frac{p_{\overline{X}_i, A_i(\mathbf{z}^{(i)})|\mathbf{A}^{(i-1)}(\mathbf{z}^{(i-1)}),\overline{\mathbf{X}}^{(i-1)}}(\overline{x}_i, a_i|\mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)})}{p_{A_i(\mathbf{z}^{(i)})|\mathbf{A}^{(i-1)}(\mathbf{z}^{(i-1)}),\overline{\mathbf{X}}^{(i-1)}}(a_i|\mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)})}\\
&= \frac{p_{\overline{X}_i, A_i|\mathbf{A}^{(i-1)},\overline{\mathbf{X}}^{(i-1)}}(\overline{x}_i, a_i|\mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)})}{p_{A_i(\mathbf{z}^{(i)})|\mathbf{A}^{(i-1)},\overline{\mathbf{X}}^{(i-1)}}(a_i|\mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)})}\\
&= \frac{p_{\overline{X}_i(z_{i+1}), A_i(z_{i+1})|\mathbf{A}^{(i-1)}(z_{i+1}),\overline{\mathbf{X}}^{(i-1)}(z_{i+1})}(\overline{x}_i, a_i|\mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)})}{p_{A_i(z_i)|\mathbf{A}^{(i-1)}(z_i),\overline{\mathbf{X}}^{(i-1)}(z_i),}(a_i|\mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)})}\\
&= \frac{p_{\overline{X}_i(z_{i+1}), A_i(z_{i+1})|\mathbf{A}^{(i-1)}(z_{i+1}),\overline{\mathbf{X}}^{(i-1)}(z_{i+1})}(\overline{x}_i, a_i|\mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)})}{p_{A_i(z_i)|\mathbf{A}^{(i-1)}(z_i),\overline{\mathbf{X}}^{(i-1)}(z_i),}(a_i|\mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)})} \times \frac{p_{A_i(z_{i+1})|\mathbf{A}^{(i-1)}(z_{i+1}),\overline{\mathbf{X}}^{(i-1)}(z_{i+1})}(a_i|\mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)})}{p_{A_i(z_{i+1})|\mathbf{A}^{(i-1)}(z_{i+1}),\overline{\mathbf{X}}^{(i-1)}(z_{i+1})}(a_i|\mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)})}\\
&= \frac{p_{A_i(z_{i+1})|\mathbf{A}^{(i-1)}(z_{i+1}),\overline{\mathbf{X}}^{(i-1)}(z_{i+1})}(a_i|\mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)})}{p_{A_i(z_i)|\mathbf{A}^{(i-1)}(z_i),\overline{\mathbf{X}}^{(i-1)}(z_i)}(a_i|\mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)})} p_{\overline{X}_i(z_{i+1})|\mathbf{A}^{(i)}(z_{i+1}),\overline{\mathbf{X}}^{(i-1)}(z_{i+1})}(\overline{x}_i|\mathbf{a}^{(i)}, \overline{\mathbf{x}}^{(i-1)})\\
&= \frac{p^{i+1}(a_i|\mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)})}{p^{i}(a_i|\mathbf{a}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)})} p^{i+1}(\overline{x}_i|\mathbf{a}^{(i)}, \overline{\mathbf{x}}^{(i-1)}) =: \text{Eq. (B.10)}.
\end{aligned}
$$

Now, it suffices to show that

$$
\begin{aligned}
\nu_0^m(\mathbf{A}^{(m-1)}(\mathbf{x}), \mathbf{x}\backslash z_m) &= \mu_0^m(\mathbf{A}^{(m-1)}(\mathbf{x}), \mathbf{x}\backslash z_m)\\
\nu_0^{i+1}(\mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{x}}^{(i)}) &= \mu_0^{i+1}(\mathbf{A}^{(i)}, \overline{\mathbf{x}}^{(i)}), \quad \text{for } i = m-2, \cdots, 1,
\end{aligned}
$$

where ν^i for $i = m, \cdots, 2$ are defined in Lemma B.3.

First,

$$
\begin{aligned}
\nu_0^m(\mathbf{A}^{(m-1)}(\mathbf{x}), \mathbf{x}\backslash z_m) &:= \mathbb{E}\left[Y(\mathbf{x})|\mathbf{A}^{(m-1)}(\mathbf{x}), (\mathbf{X}\backslash Z_m)(z_m) = \mathbf{x}\backslash z_m\right]\\
&= \mathbb{E}\left[Y(z_m)|\mathbf{A}^{(m-1)}(z_m), (\mathbf{X}\backslash Z_m)(z_m) = \mathbf{x}\backslash z_m\right]\\
&= \mathbb{E}_{P^m}\left[Y|\mathbf{A}^{(m-1)}, \mathbf{x}\backslash z_m\right]\\
&=: \mu^m(\mathbf{A}^{(m-1)}, \mathbf{x}\backslash z_m).
\end{aligned}
$$

Also,

$$
\begin{aligned}
\nu_0^{m-1}(\mathbf{A}^{(m-2)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-2)}) &:= \mathbb{E}\left[\nu_0^m(\mathbf{A}^{(m-1)}(\mathbf{x}), \mathbf{x}\backslash z_m)\middle|\mathbf{A}^{(m-2)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-2)}\right],\\
&= \mathbb{E}\left[\mu_0^m(\mathbf{A}^{(m-1)}, \mathbf{x}\backslash z_m)\middle|\mathbf{A}^{(m-2)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-2)}\right]\\
&= \mathbb{E}\left[\mu_0^m(\mathbf{A}^{(m-1)}(\mathbf{z}^{(m-1)}), \mathbf{x}\backslash z_m)\middle|\mathbf{A}^{(m-2)}(\mathbf{x}), \overline{\mathbf{x}}^{(m-2)}\right]\\
&= \mathbb{E}\left[\mu_0^m(\mathbf{A}^{(m-1)}(z_{m-1}), \mathbf{x}\backslash z_m)\middle|\mathbf{A}^{(m-2)}(z_{m-1}), \overline{\mathbf{X}}^{(m-2)}(z_{m-1}) = \overline{\mathbf{x}}^{(m-1)}\right]\\
&= \mathbb{E}_{P^{m-1}}\left[\mu_0^m(\mathbf{A}^{(m-1)}, \mathbf{x}\backslash z_m)\middle|\mathbf{A}^{(m-2)}, \overline{\mathbf{x}}^{(m-2)}\right]\\
&=: \mu_0^{m-1}(\mathbf{A}^{(m-2)}, \overline{\mathbf{x}}^{(m-2)}).
\end{aligned}
$$

Finally, for $i+1 \in \{m-1, \cdots, 3\}$, suppose

$$\nu_0^{i+1}(\mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{x}}^{(i)}) = \mu_0^{i+1}(\mathbf{A}^{(i)}, \overline{\mathbf{x}}^{(i)}).$$

Then,

$$\begin{aligned}
\nu_0^{i}(\mathbf{A}^{(i-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(i-1)}) &= \mathbb{E}\left[\nu_0^{i+1}(\mathbf{A}^{(i)}(\mathbf{x}), \overline{\mathbf{x}}^{(i)}) \middle| \mathbf{A}^{(i-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(i-1)}\right] \\
&= \mathbb{E}\left[\mu_0^{i+1}(\mathbf{A}^{(i)}, \overline{\mathbf{x}}^{(i)}) \middle| \mathbf{A}^{(i-1)}(\mathbf{x}), \overline{\mathbf{x}}^{(i-1)}\right] \\
&= \mathbb{E}\left[\mu_0^{i+1}(\mathbf{A}^{(i)}(z_i), \overline{\mathbf{x}}^{(i)}) \middle| \mathbf{A}^{(i-1)}(z_i), \overline{\mathbf{X}}^{(i-1)}(z_i) = \overline{\mathbf{x}}^{(i-1)}\right] \\
&= \mathbb{E}_{P^i}\left[\mu_0^{i+1}(\mathbf{A}^{(i)}, \overline{\mathbf{x}}^{(i)}) \middle| \mathbf{A}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)}\right] \\
&= \mu_0^{i}(\mathbf{A}^{(i-1)}, \overline{\mathbf{x}}^{(i-1)}).
\end{aligned}$$

□

C. Proofs

C.1. Preliminaries

Lemma C.1 (**Continuous Mapping Theorem for** $L_2(P)$)**.** *Let X_n, X denote a random sequence defined on a metric space S. Suppose a function $g : S \to S'$ (where S' is another metric space) is continuous almost everywhere. Suppose g is bounded. Then,*

$$X_n \overset{L_2(P)}{\to} X \implies g(X_n) \overset{L_2(P)}{\to} g(X).$$

Proof of Lemma C.1. We first note that $X_n \overset{L_2(P)}{\to} X$ implies $X_n \overset{p}{\to} X$. Then, by continuous mapping theorem, $g(X_n) \overset{p}{\to} g(X)$. Then,

$$\lim_{n\to\infty} \|g(X_n) - g(X)\|^2 = \lim_{n\to\infty} \int_{\mathcal{X}} |g(X_n) - g(X)|^2 \ d[P] \overset{*}{=} \int_{\mathcal{X}} \lim_{n\to\infty} |g(X_n) - g(X)|^2 \ d[P] = 0,$$

where the equation $\overset{*}{=}$ holds by dominated convergence theorem in $L_2(P)$ space, which is applicable since $g(X_n), g(X)$ are bounded functions (from the given condition) and $X_n \overset{p}{\to} X$. □

Lemma C.2 (**Asymptotic Unbiasedness implies Consistency**)**.** *Suppose an estimator T_N is asymptotically unbiased to μ; i.e., $\mathbb{E}_P[T_N - \mu] \to 0$ as $N \to \infty$. Suppose an estimator has vanishing variance; i.e., var$(T_N) \to 0$ as $N \to \infty$. Then, T_N is a consistent estimator of μ.*

Proof of Lemma C.2. By Markov inequality,

$$P(|T_N - \mu| > \epsilon) = P((T_N - \mu)^2 > \epsilon^2) \leq \mathbb{E}_P\left[(T_N - \mu)^2\right] / \epsilon^2.$$

Also, for $\mu_N \coloneqq \mathbb{E}_P[T_N]$,

$$\begin{aligned}
\mathbb{E}_P\left[(T_N - \mu)^2\right] &\leq 2\mathbb{E}_P\left[(T_N - \mu_N)^2\right] + 2(\mu_N - \mu)^2 \\
&= 2\mathbb{V}_P[T_N] + 2(\mu_N - \mu)^2 \\
&\to 0.
\end{aligned}$$

where $\text{var}(T_N) + (\mu_N - \mu) \to 0$ by the given assumptions that $\text{var}(T_N) \to 0$ and $\mathbb{E}_P[T_N - \mu] = \mu_N - \mu \to 0$ as $N \to \infty$. □

Lemma C.3 (**Decomposition** (Kennedy et al., 2020, Lemma 2)). *Let $f_\eta \equiv f(\mathbf{V};\eta)$ denote a finite and continuous functional and η denote its nuisances. For some samples $D \sim P$, let $T \equiv \mathbb{E}_D[f_\eta]$. Let $\theta_0 \equiv \mathbb{E}_P[f_{\eta_0}]$ for some η_0. Let $\mathbb{E}_{D-P}[f_\eta] \equiv \mathbb{E}_D[f_\eta] - \mathbb{E}_P[f_\eta]$. Then, the following decomposition holds:*

$$\mathbb{E}_D[f_\eta] - \theta_0 = \mathbb{E}_{D-P}[f_{\eta_0}] + \mathbb{E}_{D-P}[f_\eta - f_{\eta_0}] + \mathbb{E}_P[f_\eta - f_{\eta_0}]. \tag{C.1}$$

Suppose further that

1. *Samples used for estimating η are independent and separate from D; and*
2. $\|\eta - \eta_0\| = o_P(1)$.

Then, Eq. (C.1) reduces to

$$\mathbb{E}_D[f_\eta] - \theta_0 = R + \mathbb{E}_P[f_\eta - f_{\eta_0}], \tag{C.2}$$

where R is a random variable such that $\sqrt{n}R$ converges in distribution to a mean-zero normal random variable, where $n \equiv |D|$.

Proof of Lemma C.3. We first prove the equality in Eq. (C.1).

$$\begin{aligned}\mathbb{E}_D[f_\eta] - \theta_0 &= \mathbb{E}_D[f_\eta] - \mathbb{E}_P[f_{\eta_0}] \\ &= \mathbb{E}_{D-P}[f_\eta] + \mathbb{E}_P[f_\eta - f_{\eta_0}] \\ &= \underbrace{\mathbb{E}_{D-P}[f_{\eta_0}]}_{\equiv A} + \underbrace{\mathbb{E}_{D-P}[f_\eta - f_{\eta_0}]}_{\equiv B} + \mathbb{E}_P[f_\eta - f_{\eta_0}].\end{aligned}$$

We now prove Eq. (C.2).

- A converges in distribution to the zero-mean normal distribution at $\sqrt{n}$ rate by the central limit theorem.
- We note that a given condition $\|\eta - \eta_0\| = o_P(1)$ implies $\|f_\eta - f_{\eta_0}\| = o_P(1)$ by continuous mapping theorem for $L_2(P)$ in Lemma C.1. In particular, Lemma C.1 is applicable since f_η, f_{η_0} is a bounded and continuous function, and $\|\eta - \eta_0\| = o_P(1)$. Then, B converges to zero at $o_P(1/\sqrt{N})$ rate by (Kennedy et al., 2020, Lemma 2).

Finally, define $R \equiv A + B$. Then, the proof completes by applying Slutsky's theorem. ☐

C.2. Proof of Theorem 1

Definition 1 (**Adjustment criterion for Treatment-Treatment Interaction (AC-TTI)**). A set $\{C_1, W\}$ is said to satisfy the *adjustment criterion for treatment-treatment interaction (AC-TTI)* w.r.t $\{(X_1, X_2), Y\}$ in G if

1. $(\{C_1, W\} \perp\!\!\!\perp X_2 | X_1)_{G_{\overline{X_1, X_2}}}$; there are no direct paths from X_2 to $\{C_1, W\}$ in $G_{\overline{X_1, X_2}}$; and
2. $(Y \perp\!\!\!\perp X_1 | C_1, W, X_2)_{G_{\underline{X_1}\overline{X_2}}}$; the back-door paths between X_1 and Y are blocked by $\{C_1, W\}$ in $G_{\overline{X_2}}$.

Assumption 1 (**Positivity Assumption for AC-TTI**). *$P_{x_1}(C_1, W), P_{x_2}(C_1, W), P_{x_2}(X_1|C_1, W)$ are strictly positive distributions for $\forall x_1, x_2 \in \mathfrak{D}_{X_1, X_2}$.*

Theorem 1 (**Identification through AC-TTI**). *Suppose AC-TTI in Def. 1 and Assumption 1 hold. Then, $\mathbb{E}[Y|do(x_1, x_2)]$ is identifiable from $P_{rand(X_1)}(C_1, W)$ and $P_{rand(X_2)}(C_1, W, X_1, Y)$ and the expression is:*

$$\mathbb{E}[Y|do(x_1, x_2)] = \mathbb{E}_{P_{x_1}}\left[\mathbb{E}_{P_{x_2}}[Y|C_1, W, x_1]\right]. \tag{1}$$

Proof of Theorem 1.

$$
\begin{aligned}
\mathbb{E}\left[Y|do(x_1,x_2)\right] &= \mathbb{E}\left[\mathbb{E}\left[Y|do(x_1,x_2),C_1,W\right]|do(x_1,x_2)\right] \\
&\overset{1}{=} \mathbb{E}\left[\mathbb{E}\left[Y|do(x_2),x_1,C_1,W\right]|do(x_1,x_2)\right] \\
&= \mathbb{E}\left[\mathbb{E}_{P_{x_2}}\left[Y|x_1,C_1,W\right]\Big|do(x_1,x_2)\right] \\
&\overset{2}{=} \mathbb{E}\left[\mathbb{E}_{P_{x_2}}\left[Y|x_1,C_1,W\right]\Big|do(x_1)\right] \\
&= \mathbb{E}_{P_{x_1}}\left[\mathbb{E}_{P_{x_2}}\left[Y|x_1,C_1,W\right]\right],
\end{aligned}
$$

where $\overset{1}{=}$ holds by the condition 2 which implies Rule 2 of do-calculus, and $\overset{2}{=}$ holds by the condition 1 in AC-TTI which implies Rule 3 of do-calculus. □

C.3. Proof of Theorem 2 and Corollary 2

Definition 2 (Nuisances for TTI). Nuisance functions for AC-TTI functional in Eq. (1) are defined as follow: For a fixed $x_1, x_2 \in \mathfrak{D}_{X_1,X_2}$ where x_1, x_2 are specified in Eq. (1), $\pi_0 := \pi_0(C_1, X_1, W) := \frac{P_{x_1}(W|C_1)}{P_{x_2}(W,X_1|C_1)}$. Also, $\mu_0 := \mu_0(C_1, X_1, W) := \mathbb{E}_{P_{x_2}}[Y|X_1, W, C_1]$. We will use $\pi := \pi(C_1, X_1, W) > 0$ and $\mu := \mu(C_1, X_1, W)$ to denote arbitrary[4] finite functions.

Definition 3 (Estimators for TTI). Let D_1 and D_2 denote two separate samples following the distribution $P_{\text{rand}(X_1)}(C_1, W)$ and $P_{\text{rand}(X_2)}(C_1, W, X_1, Y)$, respectively. For fixed $x_1, x_2 \in \mathfrak{D}_{X_1,X_2}$, we define D_{x_1} and D_{x_2} as subsamples of D_1 and D_2 such that $X_1 = x_1$ and $X_2 = x_2$. Let μ and π denote the nuisances as defined in Definition 2. We now introduce the {REG, PW, DML} estimators for the AC-TTI-functional specified in Equation (1) as follows:

$$
\begin{aligned}
T^{reg} &:= \mathbb{E}_{D_{x_1}}\left[\mu(W, C_1, x_1))\right], \\
T^{pw} &:= \mathbb{E}_{D_{x_2}}\left[\pi(W, C_1, X_1)\mathbb{1}_{x_1}(X_1)Y\right], \\
T^{dml} &:= \mathbb{E}_{D_{x_1}}\left[\pi\mathbb{1}_{x_1}(X_1)\{Y-\mu\}\right] + \mathbb{E}_{D_1}\left[\mu(W, C_1, x_1))\right].
\end{aligned}
$$

Assumption 2 (Sample-splitting). *Samples for training nuisances and evaluating the estimators equipped with the trained nuisance are separate and independent.*

Assumption 3 (L_2 consistency of nuisances). *Estimated nuisances are L_2 consistent; i.e., $\forall i \in \{1,2\}, \forall x_i \in \mathfrak{D}_{X_i}$,*

$$
\begin{aligned}
\|\mu(W, C_1, x_1) - \mu_0(W, C_1, x_1)\|_{P_{x_i}} &= o_{P_{x_i}}(1), \\
\|\pi(W, C_1, X_1) - \pi_0(W, C_1, X_1)\|_{P_{x_2}} &= o_{P_{x_2}}(1).
\end{aligned}
$$

Theorem 2 (Error analysis of the estimators). *Under Assumptions (1,2,3,4) and AC-TTI in Def. 1, the error of the estimators in Def. 3, denoted $\epsilon^{est} := T^{est} - \mathbb{E}\left[Y|do(x_1,x_2)\right]$ for est $\in \{reg, pw, dml\}$ are:*

$$
\begin{aligned}
\epsilon^{reg} &= R_1 + O_{P_{x_1}}\left(\|\mu - \mu_0\|\right), \\
\epsilon^{pw} &= R_2 + O_{P_{x_2}}\left(\|\pi - \pi_0\|\right), \\
\epsilon^{dml} &= R_1 + R_2 + O_{P_{x_2}}\left(\|\pi - \pi_0\|\,\|\mu - \mu_0\|\right),
\end{aligned}
$$

where R_i is a random variable such that $\sqrt{n_i}R_i$ converges in distribution to the mean-zero normal random variable, where $n_i := |D_{x_i}|$ for $i \in \{1,2\}$.

Proof of Theorem 2. We provide error analyses for each estimators:

Analysis for T^{reg}.

[4] Throughout the paper, μ, π may be understood as estimated nuisances for μ_0, π_0.

We first note that

$$\mathbb{E}_{P_{x_1}}\left[\mu_0(W,C_1,\mathbf{x})\right] = \mathbb{E}_{P_{x_1}}\left[\mathbb{E}_{P_{x_2}}\left[Y|W,C_1,x_1\right]\right] = \mathbb{E}\left[Y|do(x_1,x_2)\right],$$

where the last equation holds by Theorem 1. By Lemma C.3,

$$\begin{aligned}
&T^{reg} - \mathbb{E}\left[Y|do(x_1,x_2)\right] \\
&= T^{reg} - \mathbb{E}_{P_{x_1}}\left[\mu_0(W,C_1,\mathbf{x})\right] \\
&= \underbrace{\mathbb{E}_{P_{x_1}-D_{x_1}}\left[\mu_0(W,C_1,\mathbf{x})\right] + \mathbb{E}_{P_{x_1}-D_{x_1}}\left[\mu(W,C_1,\mathbf{x}) - \mu_0(W,C_1,\mathbf{x})\right]}_{:=R_1} + \mathbb{E}_{P_{x_1}}\left[\mu(W,C_1,\mathbf{x}) - \mu_0(W,C_1,\mathbf{x})\right] \\
&\overset{1}{=} R_1 + \mathbb{E}_{P_{x_1}}\left[\mu(W,C_1,\mathbf{x}) - \mu_0(W,C_1,\mathbf{x})\right] \\
&\overset{2}{=} R_1 + O_{P_{x_1}}\left(\|\mu_0 - \mu\|\right),
\end{aligned}$$

where

- $\overset{1}{=}$ holds by Lemma C.3.
- $\overset{2}{=}$ holds by Cauchy-Schwartz inequality.

Analysis for T^{pw}.

We first note that

$$\begin{aligned}
\mathbb{E}_{P_{x_2}}\left[\pi_0(W,C_1,X_1)\mathbb{1}_{x_1}(X_1)Y\right] &= \mathbb{E}_{P_{x_2}}\left[\pi_0(W,C_1,X_1)\mathbb{1}_{x_1}(X_1)\mu_0(W,C_1,X_1)\right] \\
&= \mathbb{E}_{P_{x_2}}\left[\frac{P_{x_1}(W|C_1)}{P_{x_2}(W,X_1|C_1)}\mathbb{1}_{x_1}(X_1)\mu_0(W,C_1,X_1)\right] \\
&\overset{3}{=} \mathbb{E}_{P_{x_2}}\left[\frac{P_{x_1}(W|C_1)P_{x_1}(C_1)}{P_{x_2}(W,X_1|C_1)P_{x_2}(C_1)}\mathbb{1}_{x_1}(X_1)\mu_0(W,C_1,X_1)\right] \\
&= \mathbb{E}_{P_{x_2}}\left[\frac{P_{x_1}(W,C_1)}{P_{x_2}(X_1|W,C_1)P_{x_2}(W,C_1)}\mathbb{1}_{x_1}(X_1)\mu_0(W,C_1,X_1)\right] \\
&= \mathbb{E}_{P_{x_2}}\left[\frac{P_{x_1}(W,C_1)}{P_{x_2}(W,C_1)}\mu_0(W,C_1,x_1)\right] \\
&= \mathbb{E}_{P_{x_1}}\left[\mu_0(W,C_1,x_1)\right] \\
&\overset{4}{=} \mathbb{E}\left[Y|do(x_1,x_2)\right],
\end{aligned}$$

where

- $\overset{3}{=}$ holds by Assumption 4.
- $\overset{4}{=}$ holds by Theorem 1.

By applying Lemma C.3,

$$\begin{aligned}
&T^{pw} - \mathbb{E}\left[Y|do(x_1,x_2)\right] \\
&= T^{pw} - \mathbb{E}_{P_{x_2}}\left[\pi_0(W,C_1,X_1)\mathbb{1}_{x_1}(X_1)Y\right] \\
&= \underbrace{\mathbb{E}_{P_{x_2}-D_{x_2}}\left[\pi_0(W,C_1,X_1)\mathbb{1}_{x_1}(X_1)Y\right] + \mathbb{E}_{P_{x_2}-D_{x_2}}\left[\{\pi_0(W,C_1,X_1)-\pi(W,C_1,X_1)\}\mathbb{1}_{x_1}(X_1)Y\right]}_{:=R_2} \\
&+ \mathbb{E}_{P_{x_2}}\left[\{\pi_0(W,C_1,X_1)-\pi(W,C_1,X_1)\}\,\mathbb{1}_{x_1}(X_1)Y\right] \\
&\overset{5}{=} R_2 + \mathbb{E}_{P_{x_2}}\left[\{\pi_0(W,C_1,X_1)-\pi(W,C_1,X_1)\}\,\mathbb{1}_{x_1}(X_1)Y\right] \\
&\overset{6}{=} R_2 + O_{P_{x_2}}\left(\|\pi_0-\pi\|\right),
\end{aligned}$$

- $\overset{5}{=}$ holds by Assumption 4.
- $\overset{6}{=}$ holds by Cauchy-Schwartz inequality and the setting where Y has a finite variance.

Analysis for T^{dml}.

Let

$$T^{dml} := \underbrace{\mathbb{E}_{D_{x_2}}\left[\pi(W,C_1,X_1)\mathbb{1}_{x_1}(X_1)\{Y-\mu(W,C_1,X_1)\}\right]}_{:=T^{dml,1}} + \underbrace{\mathbb{E}_{D_{x_1}}\left[\mu(W,C_1,x_1))\right]}_{:=T^{dml,2}}.$$

Let

$$\begin{aligned}
T_0^{dml,1} &:= \mathbb{E}_{P_{x_2}}\left[\pi_0(W,C_1,X_1)\mathbb{1}_{x_1}(X_1)\{Y-\mu_0(W,C_1,X_1)\}\right] \\
T_0^{dml,2} &:= \mathbb{E}_{P_{x_1}}\left[\mu_0(W,C_1,x_1))\right].
\end{aligned}$$

We note that $T_0^{dml} := T_0^{dml,1} + T_0^{dml,2} = \mathbb{E}\left[Y|do(x_1,x_2)\right]$. We first apply the Lemma C.3 to $T^{dml,1}$ and $T^{dml,2}$ separately.

$$\begin{aligned}
&T^{dml,1} - T_0^{dml,1} \\
&= \mathbb{E}_{P_{x_2}-D_{x_2}}\left[\pi_0(W,C_1,x_1)\mathbb{1}_{x_1}(X_1)\{Y-\mu_0(W,C_1,X_1)\}\right] \\
&+ \mathbb{E}_{P_{x_2}-D_{x_2}}\left[\pi_0(W,C_1,x_1)\mathbb{1}_{x_1}(X_1)\{Y-\mu_0(W,C_1,X_1)\} - \pi(W,C_1,x_1)\mathbb{1}_{x_1}(X_1)\{Y-\mu(W,C_1,X_1)\}\right] \\
&+ \mathbb{E}_{P_{x_2}}\left[\pi(W,C_1,x_1)\mathbb{1}_{x_1}(X_1)\{Y-\mu(W,C_1,X_1)\}\right] \\
&= R_2 + \mathbb{E}_{P_{x_2}}\left[\pi(W,C_1,x_1)\mathbb{1}_{x_1}(X_1)\{Y-\mu(W,C_1,X_1)\}\right],
\end{aligned}$$

where

$$\begin{aligned}
R_2 &:= \mathbb{E}_{P_{x_2}-D_{x_2}}\left[\pi_0(W,C_1,x_1)\mathbb{1}_{x_1}(X_1)\{Y-\mu_0(W,C_1,X_1)\}\right] \\
&\quad + \mathbb{E}_{P_{x_2}-D_{x_2}}\left[\pi_0(W,C_1,x_1)\mathbb{1}_{x_1}(X_1)\{Y-\mu_0(W,C_1,X_1)\} - \pi(W,C_1,x_1)\mathbb{1}_{x_1}(X_1)\{Y-\mu(W,C_1,X_1)\}\right].
\end{aligned}$$

Also, by the proof for analyzing the error of T^{reg},

$$T^{dml,2} - T_0^{dml,2} = R_1 + \mathbb{E}_{P_{x_1}}\left[\mu(W,C_1,x_1) - \mu_0(W,C_1,x_1)\right].$$

Then,

$$\begin{aligned}
&T^{dml} - \mathbb{E}\left[Y|do(x_1,x_2)\right] \\
&= T^{dml,1} + T^{dml,2} - T_0^{dml,1} - T_0^{dml,2} \\
&= R_1 + R_2 + \mathbb{E}_{P_{x_2}}\left[\pi(W,C_1,x_1)\mathbb{1}_{x_1}(X_1)\{Y-\mu(W,C_1,X_1)\}\right] + \mathbb{E}_{P_{x_1}}\left[\mu(W,C_1,x_1) - \mu_0(W,C_1,x_1)\right].
\end{aligned}$$

Note that

$$\begin{aligned}
&\mathbb{E}_{P_{x_1}}\left[\mu(W,C_1,x_1)-\mu_0(W,C_1,x_1)\right] \\
&=\mathbb{E}_{P_{x_1}}\left[\frac{\mathbb{1}_{x_1}(X_1)}{P_{x_1}(X_1|W,C_1)}\left\{\mu(W,C_1,X_1)-\mu_0(W,C_1,X_1)\right\}\right] \\
&=\mathbb{E}_{P_{x_2}}\left[\frac{P_{x_1}(W,C_1,X_1)}{P_{x_2}(W,C_1,X_1)}\frac{\mathbb{1}_{x_1}(X_1)}{P_{x_1}(X_1|W,C_1)}\left\{\mu(W,C_1,X_1)-\mu_0(W,C_1,X_1)\right\}\right] \\
&=\mathbb{E}_{P_{x_2}}\left[\frac{P_{x_1}(W,C_1)\mathbb{1}_{x_1}(X_1)}{P_{x_2}(W,C_1,X_1)}\left\{\mu(W,C_1,X_1)-\mu_0(W,C_1,X_1)\right\}\right] \\
&\overset{7}{=}\mathbb{E}_{P_{x_2}}\left[\frac{P_{x_1}(W|C_1)\mathbb{1}_{x_1}(X_1)}{P_{x_2}(W,X_1|C_1)}\left\{\mu(W,C_1,X_1)-\mu_0(W,C_1,X_1)\right\}\right] \\
&=\mathbb{E}_{P_{x_2}}\left[\pi_0(W,C_1,X_1)\mathbb{1}_{x_1}(X_1)\left\{\mu(W,C_1,X_1)-\mu_0(W,C_1,X_1)\right\}\right],
\end{aligned}$$

where $\overset{7}{=}$ holds by Assumption 4. Then,

$$\begin{aligned}
&\mathbb{E}_{P_{x_2}}\left[\pi(W,C_1,x_1)\mathbb{1}_{x_1}(X_1)\{Y-\mu(W,C_1,X_1)\}\right]+\mathbb{E}_{P_{x_1}}\left[\mu(W,C_1,x_1)-\mu_0(W,C_1,x_1)\right] \\
&=\mathbb{E}_{P_{x_2}}\left[\pi(W,C_1,x_1)\mathbb{1}_{x_1}(X_1)\{Y-\mu(W,C_1,X_1)\}\right]+\mathbb{E}_{P_{x_2}}\left[\pi_0(W,C_1,X_1)\mathbb{1}_{x_1}(X_1)\left\{\mu(W,C_1,X_1)-\mu_0(W,C_1,X_1)\right\}\right] \\
&=\mathbb{E}_{P_{x_2}}\left[\mathbb{1}_{x_1}(X_1)\left(\pi(W,C_1,x_1)\{\mu_0(W,C_1,X_1)-\mu(W,C_1,X_1)\}+\pi_0(W,C_1,X_1)\left\{\mu(W,C_1,X_1)-\mu_0(W,C_1,X_1)\right\}\right)\right] \\
&=\mathbb{E}_{P_{x_2}}\left[\mathbb{1}_{x_1}(X_1)\left\{\mu_0(W,C_1,X_1)-\mu(W,C_1,X_1)\right\}\left\{\pi_0(W,C_1,X_1)-\pi(W,C_1,X_1)\right\}\right] \\
&=O_{P_{x_2}}\left(\|\mu-\mu_0\|\,\|\pi-\pi_0\|\right).
\end{aligned}$$

Therefore,

$$T^{dml}-\mathbb{E}\left[Y|do(x_1,x_2)\right]=R_1+R_2+O_{P_{x_2}}\left(\|\mu-\mu_0\|\,\|\pi-\pi_0\|\right).$$

□

Corollary 2 (**Doubly robustness of the DML estimators** (Corollary of Thm. 2)**).** *Suppose Assumptions (1,2,3,4) and AC-TTI in Def. 1 hold. Suppose either* $\pi=\pi_0$ *or* $\mu=\mu_0$*. Then,* T^{dml} *is an unbiased estimator of* $\mathbb{E}\left[Y|do(x_1,x_2)\right]$.

Proof of Corollary 2. Let π and μ denote the limiting estimator for π_0 and μ_0.

$$T^{dml}:=\underbrace{\mathbb{E}_{D_{x_2}}\left[\pi(W,C_1,X_1)\mathbb{1}_{x_1}(X_1)\{Y-\mu(W,C_1,X_1)\}\right]}_{:=T^{dml,1}}+\underbrace{\mathbb{E}_{D_{x_1}}\left[\mu(W,C_1,x_1))\right]}_{:=T^{dml,2}}.$$

Let

$$\begin{aligned}
T_0^{dml,1}&:=\mathbb{E}_{P_{x_2}}\left[\pi_0(W,C_1,X_1)\mathbb{1}_{x_1}(X_1)\{Y-\mu_0(W,C_1,X_1)\}\right] \\
T_0^{dml,2}&:=\mathbb{E}_{P_{x_1}}\left[\mu_0(W,C_1,x_1))\right].
\end{aligned}$$

Under the assumption that

$$\begin{aligned}
\mathbb{E}_{P_{x_2}}\left[T^{dml,1}\right]&=\mathbb{E}_{P_{x_2}}\left[\pi(W,C_1,X_1)\mathbb{1}_{x_1}(X_1)\left\{Y-\mu(W,C_1,X_1)\right\}\right] \\
\mathbb{E}_{P_{x_1}}\left[T^{dml,2}\right]&=\mathbb{E}_{P_{x_1}}\left[\mu(W,C_1,x_1)\right] \\
&=\mathbb{E}_{P_{x_2}}\left[\pi_0(W,C_1,X_1)\mathbb{1}_{x_1}(X_1)\mu(W,C_1,X_1)\right].
\end{aligned}$$

Then,

$$
\begin{aligned}
&\mathbb{E}_{P_{x_2}}\left[T^{dml,1}\right] + \mathbb{E}_{P_{x_1}}\left[T^{dml,2}\right] - \mathbb{E}_{P_{x_2}}\left[T_0^{dml,1}\right] + \mathbb{E}_{P_{x_1}}\left[T_0^{dml,2}\right] \\
&= \mathbb{E}_{P_{x_2}}\left[\pi(W,C_1,X_1)\mathbb{1}_{x_1}(X_1)\{Y - \mu(W,C_1,X_1)\} + \pi_0(W,C_1,X_1)\mathbb{1}_{x_1}(X_1)\left\{\mu(W,C_1,X_1) - \mu_0(W,C_1,X_1)\right\}\right] \\
&= \mathbb{E}_{P_{x_2}}\left[\mathbb{1}_{x_1}(X_1)\left\{\mu_0(W,C_1,X_1) - \mu(W,C_1,X_1)\right\}\left\{\pi_0(W,C_1,X_1) - \pi(W,C_1,X_1)\right\}\right] \\
&= O_{P_{x_2}}\left(\|\mu - \mu_0\|\,\|\pi - \pi_0\|\right) \\
&= 0,
\end{aligned}
$$

where the last equation holds under the given condition. □

C.4. Proof of Theorem 3

Definition 4 (Adjustment criterion for combining two experiments (AC-gTTI)). A set of variables $\mathbf{A}$ is said to satisfy *adjustment criterion for generalized TTI (AC-gTTI)* w.r.t an ordered set $\mathbf{X}$ and Y in G if

1. $Z_1 \subseteq \mathbf{X}$ and $(\mathbf{A} \perp\!\!\!\perp \mathbf{X}\backslash Z_1 | Z_1)_{G_{\overline{\mathbf{X}}}}$; there are no direct paths from $\mathbf{X}\backslash Z_1$ to $\mathbf{A}$ in $G_{\overline{\mathbf{X}}}$; and

2. $Z_2 \subseteq \mathbf{X}$ and $(Y \perp\!\!\!\perp \mathbf{X}\backslash Z_2 | \mathbf{A}, Z_2)_{G_{\underline{\mathbf{X}\backslash Z_2}\overline{Z_2}}}$; the back-door paths between $\mathbf{X}\backslash Z_2$ and Y are blocked by $\mathbf{A}$ in $G_{\overline{Z_2}}$.

Assumption 5 (Positivity Assumption for AC-gTTI). *$P_{z_1}(\mathbf{A}), P_{z_2}(\mathbf{A}), P_{z_2}(\mathbf{X}\backslash Z_2|\mathbf{A})$ are strictly positive distributions for $\forall z_1, z_2 \in \mathfrak{D}_{Z_1,Z_2}$.*

Theorem 3 (Identification through AC-gTTI). *Suppose AC-gTTI in Def. 4 and Assumption 5 hold. Then, the query $\mathbb{E}\left[Y|do(\mathbf{x})\right]$ is identifiable from $P_{rand(Z_1)}(\mathbf{A})$ and $P_{rand(Z_2)}(\mathbf{A},\mathbf{X},Y)$ and given as follow:*

$$
\mathbb{E}\left[Y|do(\mathbf{x})\right] = \mathbb{E}_{P_{z_1}}\left[\mathbb{E}_{P_{z_2}}\left[Y|\mathbf{A},\mathbf{x}\backslash\mathbf{z}_2\right]\right]. \tag{2}
$$

Proof of Theorem 3.

$$
\begin{aligned}
\mathbb{E}\left[Y|do(\mathbf{x})\right] &= \mathbb{E}\left[\mathbb{E}\left[Y|do(\mathbf{x}),\mathbf{A}\right]|do(\mathbf{x})\right] \\
&\overset{1}{=} \mathbb{E}\left[\mathbb{E}\left[Y|do(z_2),\mathbf{x}\backslash z_2,\mathbf{A}\right]|do(\mathbf{x})\right] \\
&= \mathbb{E}\left[\mathbb{E}_{P_{z_2}}\left[Y|\mathbf{x}\backslash z_2,\mathbf{A}\right]\Big|do(\mathbf{x})\right] \\
&\overset{2}{=} \mathbb{E}\left[\mathbb{E}_{P_{z_2}}\left[Y|\mathbf{x}\backslash z_2,\mathbf{A}\right]\Big|do(z_1)\right] \\
&= \mathbb{E}_{P_{z_1}}\left[\mathbb{E}_{P_{z_2}}\left[Y|\mathbf{x}\backslash z_2,\mathbf{A}\right]\right],
\end{aligned}
$$

where $\overset{1}{=}$ holds by the condition 2 which implies Rule 2 of do-calculus, and $\overset{2}{=}$ holds by the condition 1 in AC-gTTI which implies Rule 3 of do-calculus. □

C.5. Proof of Theorem 4 and Corollary 4

Definition 5 (Nuisances for gTTI). Nuisance functions for estimating AC-gTTI functional in Eq. (2) are defined as follow: For a fixed $z_1, z_2 \in \mathfrak{D}_{Z_1,Z_2}$ where z_1, z_2 are specified in Eq. (2), $\pi_0 \coloneqq \pi_0(\mathbf{A},\mathbf{X}) \coloneqq \frac{P_{z_1}(\mathbf{A})}{P_{z_2}(\mathbf{A},\mathbf{X}\backslash Z_2)}$, and $\mu_0 \coloneqq \mu_0(\mathbf{A},\mathbf{X}) \coloneqq \mathbb{E}_{P_{z_2}}\left[Y|\mathbf{X}\backslash Z_2,\mathbf{A}\right]$. We will use $\pi \coloneqq \pi(\mathbf{A},\mathbf{X}) > 0$ and $\mu \coloneqq \mu(\mathbf{A},\mathbf{X})$ to denote an arbitrary finite function.

Definition 6 (Estimators for gTTI). Let D_1, D_2 denote two sample sets following distributions $P_{\text{rand}(Z_1)}(\mathbf{A})$ and $P_{\text{rand}(Z_2)}(\mathbf{A},\mathbf{X},Y)$, respectively. For a fixed $z_1, z_2 \in \mathfrak{D}_{Z_1,Z_2}$, we define D_{z_1} and D_{z_2} as subsamples of D_1 and D_2 such that $Z_1 = z_1$ and $Z_2 = z_2$. Let μ, π denote nuisances defined in Def. 5. Then, {REG, PW, DML} estimators for

AC-gTTI functional defined as follow:

$$
\begin{aligned}
T^{reg} &:= \mathbb{E}_{D_{z_1}}\left[\mu(\mathbf{A},\mathbf{x}))\right],\\
T^{pw} &:= \mathbb{E}_{D_{z_2}}\left[\pi(\mathbf{A},\mathbf{X})\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right],\\
T^{dml} &:= \mathbb{E}_{D_{z_2}}\left[\pi\mathbb{1}_{\mathbf{x}}(\mathbf{X})\{Y-\mu\}\right] + \mathbb{E}_{D_{z_1}}\left[\mu(\mathbf{A},\mathbf{x}))\right].
\end{aligned}
$$

Assumption 2 (Sample-splitting). *Samples for training nuisances and evaluating the estimators equipped with the trained nuisance are separate and independent*

Assumption 6 (L_2 consistency of nuisances). *Estimated nuisances are L_2 consistent; i.e., $\forall i \in \{1,2\}, \forall z_i \in \mathfrak{D}_{Z_i}$,*

$$
\begin{aligned}
\|\mu(\mathbf{A},\mathbf{x}) - \mu_0(\mathbf{A},\mathbf{x})\|_{P_{z_i}} &= o_{P_{z_i}}(1),\\
\|\pi(\mathbf{A},\mathbf{X}) - \pi_0(\mathbf{A},\mathbf{X})\|_{P_{z_2}} &= o_{P_{z_2}}(1).
\end{aligned}
$$

Theorem 4 (Error analysis of the estimators). *Under Assumptions (2,5,6) and AC-gTTI in Def. 4, the error of the estimators in Def. 6, denoted $\epsilon^{est} := T^{est} - \mathbb{E}\left[Y|do(\mathbf{x})\right]$ for est $\in \{reg, pw, dml\}$, are:*

$$
\begin{aligned}
\epsilon^{reg} &= R_1 + O_{P_{z_1}}\left(\|\mu-\mu_0\|\right),\\
\epsilon^{pw} &= R_2 + O_{P_{z_2}}\left(\|\pi-\pi_0\|\right),\\
\epsilon^{dml} &= R_1 + R_2 + O_{P_{z_2}}\left(\|\pi-\pi_0\|\,\|\mu-\mu_0\|\right),
\end{aligned}
$$

where R_i is a random variable such that $\sqrt{n_i}R_i$ converges in distribution to the mean-zero normal random variable, where $n_i := |D_{z_i}|$.

Proof of Theorem 4. We provide error analyses for each estimators:

Analysis for T^{reg}.

We first note that

$$
\mathbb{E}_{P_{z_1}}\left[\mu_0(\mathbf{A},\mathbf{x})\right] = \mathbb{E}_{P_{z_1}}\left[\mathbb{E}_{P_{z_2}}\left[Y|\mathbf{A},\mathbf{x}\backslash\mathbf{z}_2\right]\right] = \mathbb{E}\left[Y|do(\mathbf{x})\right],
$$

where the last equation holds by Theorem 3. By Lemma C.3,

$$
\begin{aligned}
&T^{reg} - \mathbb{E}\left[Y|do(\mathbf{x})\right]\\
&= T^{reg} - \mathbb{E}_{P_{z_1}}\left[\mu_0(\mathbf{A},\mathbf{x})\right]\\
&= \underbrace{\mathbb{E}_{P_{z_1}-D_1}\left[\mu_0(\mathbf{A},\mathbf{x})\right] + \mathbb{E}_{P_{z_1}-D_1}\left[\mu(\mathbf{A},\mathbf{x}) - \mu_0(\mathbf{A},\mathbf{x})\right]}_{:=R_1} + \mathbb{E}_{P_{z_1}}\left[\mu(\mathbf{A},\mathbf{x}) - \mu_0(\mathbf{A},\mathbf{x})\right]\\
&\stackrel{1}{=} R_1 + \mathbb{E}_{P_{z_1}}\left[\mu(\mathbf{A},\mathbf{x}) - \mu_0(\mathbf{A},\mathbf{x})\right]\\
&\stackrel{2}{=} R_1 + O_{P_{z_1}}\left(\|\mu_0-\mu\|\right),
\end{aligned}
$$

where

- $\stackrel{1}{=}$ holds by Lemma C.3.
- $\stackrel{2}{=}$ holds by Cauchy-Schwartz inequality.

Analysis for T^{pw}.

We first note that

$$\begin{aligned}\mathbb{E}_{P_{z_2}}\left[\pi_0(\mathbf{A},\mathbf{X})\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right] &= \mathbb{E}_{P_{z_2}}\left[\pi_0(\mathbf{A},\mathbf{X})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\mu_0(\mathbf{A},\mathbf{X})\right] \\ &= \mathbb{E}_{P_{z_2}}\left[\frac{P_{z_1}(\mathbf{A})}{P_{z_2}(\mathbf{A},\mathbf{X}\backslash Z_2)}\mathbb{1}_{\mathbf{x}}(\mathbf{X})\mu_0(\mathbf{A},\mathbf{X})\right] \\ &= \mathbb{E}_{P_{z_1}}\left[\mathbb{1}_{\mathbf{x}}(\mathbf{X})\mu_0(\mathbf{A},\mathbf{X})\right] \\ &= \mathbb{E}_{P_{z_1}}\left[\mu_0(\mathbf{A},\mathbf{x})\right] \\ &\stackrel{3}{=} \mathbb{E}\left[Y|do(\mathbf{x})\right],\end{aligned}$$

where $\stackrel{3}{=}$ holds by Theorem 3. By applying Lemma C.3,

$$\begin{aligned}&T^{pw} - \mathbb{E}\left[Y|do(\mathbf{x})\right] \\ &= T^{pw} - \mathbb{E}_{P_{z_2}}\left[\pi_0(\mathbf{A},\mathbf{X})\mathbb{1}_{\mathbf{X}}(\mathbf{X})Y\right] \\ &= \underbrace{\mathbb{E}_{P_{z_2}-D_2}\left[\pi_0(\mathbf{A},\mathbf{X})\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right] + \mathbb{E}_{P_{z_2}-D_2}\left[\{\pi_0(\mathbf{A},\mathbf{X}) - \pi(\mathbf{A},\mathbf{X})\}\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right]}_{:=R_2} \\ &+ \mathbb{E}_{P_{z_2}}\left[\{\pi_0(\mathbf{A},\mathbf{X}) - \pi(\mathbf{A},\mathbf{X})\}\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right] \\ &\stackrel{4}{=} R_2 + \mathbb{E}_{P_{z_2}}\left[\{\pi_0(\mathbf{A},\mathbf{X}) - \pi(\mathbf{A},\mathbf{X})\}\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right] \\ &\stackrel{5}{=} R_2 + O_{P_{z_2}}\left(\|\pi_0 - \pi\|\right),\end{aligned}$$

- $\stackrel{4}{=}$ holds by Lemmas (C.1, C.3).
- $\stackrel{5}{=}$ holds by Cauchy-Schwartz inequality and the setting where Y has a finite variance.

Analysis for T^{dml}.

Let

$$T^{dml} := \underbrace{\mathbb{E}_{D_2}\left[\pi(\mathbf{A},\mathbf{X})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\{Y - \mu(\mathbf{A},\mathbf{X})\}\right]}_{:=T^{dml,1}} + \underbrace{\mathbb{E}_{D_1}\left[\mu(\mathbf{A},\mathbf{x}))\right]}_{:=T^{dml,2}}.$$

Let

$$\begin{aligned}T_0^{dml,1} &:= \mathbb{E}_{P_{z_2}}\left[\pi_0(\mathbf{A},\mathbf{X})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\{Y - \mu_0(\mathbf{A},\mathbf{X})\}\right] \\ T_0^{dml,2} &:= \mathbb{E}_{P_{z_1}}\left[\mu_0(\mathbf{A},\mathbf{x}))\right].\end{aligned}$$

We note that $T_0^{dml} := T_0^{dml,1} + T_0^{dml,2} = \mathbb{E}\left[Y|do(\mathbf{x})\right]$. We first apply the Lemma C.3 to $T^{dml,1}$ and $T^{dml,2}$ separately.

$$\begin{aligned}&T^{dml,1} - T_0^{dml,1} \\ &= \mathbb{E}_{P_{z_2}-D_2}\left[\pi_0(\mathbf{A},\mathbf{x})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\{Y - \mu_0(\mathbf{A},\mathbf{X})\}\right] \\ &+ \mathbb{E}_{P_{z_2}-D_2}\left[\pi_0(\mathbf{A},\mathbf{x})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\{Y - \mu_0(\mathbf{A},\mathbf{X})\} - \pi(\mathbf{A},\mathbf{x})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\{Y - \mu(\mathbf{A},\mathbf{X})\}\right] \\ &+ \mathbb{E}_{P_{z_2}}\left[\pi(\mathbf{A},\mathbf{x})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\{Y - \mu(\mathbf{A},\mathbf{X})\}\right] \\ &= R_2 + \mathbb{E}_{P_{z_2}}\left[\pi(\mathbf{A},\mathbf{x})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\{Y - \mu(\mathbf{A},\mathbf{X})\}\right],\end{aligned}$$

where

$$\begin{aligned}R_2 := &\,\mathbb{E}_{P_{z_2}-D_2}\left[\pi_0(\mathbf{A},\mathbf{x})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\{Y - \mu_0(\mathbf{A},\mathbf{X})\}\right] \\ &+ \mathbb{E}_{P_{z_2}-D_2}\left[\pi_0(\mathbf{A},\mathbf{x})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\{Y - \mu_0(\mathbf{A},\mathbf{X})\} - \pi(\mathbf{A},\mathbf{x})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\{Y - \mu(\mathbf{A},\mathbf{X})\}\right].\end{aligned}$$

Also, by the proof for analyzing the error of T^{reg},

$$T^{dml,2} - T_0^{dml,2} = R_1 + \mathbb{E}_{P_{z_1}}\left[\mu(\mathbf{A},\mathbf{x}) - \mu_0(\mathbf{A},\mathbf{x})\right].$$

Finally,

$$\begin{aligned}
&T^{dml} - \mathbb{E}\left[Y|do(\mathbf{x})\right] \\
&= T^{dml,1} + T^{dml,2} - T_0^{dml,1} - T_0^{dml,2} \\
&= R_1 + R_2 + \mathbb{E}_{P_{z_2}}\left[\pi(\mathbf{A},\mathbf{x})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\{Y - \mu(\mathbf{A},\mathbf{X})\}\right] + \mathbb{E}_{P_{z_1}}\left[\mu(\mathbf{A},\mathbf{x}) - \mu_0(\mathbf{A},\mathbf{x})\right].
\end{aligned}$$

We note that R_i for $i \in \{1,2\}$ is a variable such that $\sqrt{n_i}R_i$ converges in distribution to the normal random variable, where $n_i := |D_{z_i}|$, by Lemmas (C.1, C.3). Note that

$$\begin{aligned}
&\mathbb{E}_{P_{z_1}}\left[\mu(\mathbf{A},\mathbf{x}) - \mu_0(\mathbf{A},\mathbf{x})\right] \\
&= \mathbb{E}_{P_{z_1}}\left[\frac{\mathbb{1}_{\mathbf{x}}(\mathbf{X})}{P_{z_1}(\mathbf{X}|\mathbf{A})}\left\{\mu(\mathbf{A},\mathbf{X}) - \mu_0(\mathbf{A},\mathbf{X})\right\}\right] \\
&= \mathbb{E}_{P_{z_2}}\left[\frac{P_{z_1}(\mathbf{A},\mathbf{X})}{P_{z_2}(\mathbf{A},\mathbf{X})}\frac{\mathbb{1}_{\mathbf{x}}(\mathbf{X})}{P_{z_1}(\mathbf{X}|\mathbf{A})}\left\{\mu(\mathbf{A},\mathbf{X}) - \mu_0(\mathbf{A},\mathbf{X})\right\}\right] \\
&= \mathbb{E}_{P_{z_2}}\left[\frac{P_{z_1}(\mathbf{A})\mathbb{1}_{\mathbf{x}}(\mathbf{X})}{P_{z_2}(\mathbf{A},\mathbf{X}\setminus\mathbf{Z}_2)}\left\{\mu(\mathbf{A},\mathbf{X}) - \mu_0(\mathbf{A},\mathbf{X})\right\}\right] \\
&= \mathbb{E}_{P_{z_2}}\left[\pi_0(\mathbf{A},\mathbf{X})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\left\{\mu(\mathbf{A},\mathbf{X}) - \mu_0(\mathbf{A},\mathbf{X})\right\}\right].
\end{aligned}$$

Then,

$$\begin{aligned}
&\mathbb{E}_{P_{z_2}}\left[\pi(\mathbf{A},\mathbf{x})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\{Y - \mu(\mathbf{A},\mathbf{X})\}\right] + \mathbb{E}_{P_{z_1}}\left[\mu(\mathbf{A},\mathbf{x}) - \mu_0(\mathbf{A},\mathbf{x})\right] \\
&= \mathbb{E}_{P_{z_2}}\left[\pi(\mathbf{A},\mathbf{x})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\{Y - \mu(\mathbf{A},\mathbf{X})\}\right] + \mathbb{E}_{P_{z_2}}\left[\pi_0(\mathbf{A},\mathbf{X})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\left\{\mu(\mathbf{A},\mathbf{X}) - \mu_0(\mathbf{A},\mathbf{X})\right\}\right] \\
&= \mathbb{E}_{P_{z_2}}\left[\mathbb{1}_{\mathbf{x}}(\mathbf{X})\left(\pi(\mathbf{A},\mathbf{x})\{\mu_0(\mathbf{A},\mathbf{X}) - \mu(\mathbf{A},\mathbf{X})\} + \pi_0(\mathbf{A},\mathbf{X})\left\{\mu(\mathbf{A},\mathbf{X}) - \mu_0(\mathbf{A},\mathbf{X})\right\}\right)\right] \\
&= \mathbb{E}_{P_{z_2}}\left[\mathbb{1}_{\mathbf{x}}(\mathbf{X})\left\{\mu_0(\mathbf{A},\mathbf{X}) - \mu(\mathbf{A},\mathbf{X})\right\}\left\{\pi_0(\mathbf{A},\mathbf{X}) - \pi(\mathbf{A},\mathbf{X})\right\}\right] \\
&= O_{P_{z_2}}\left(\|\mu - \mu_0\|\,\|\pi - \pi_0\|\right).
\end{aligned}$$

Therefore,

$$T^{dml} - \mathbb{E}\left[Y|do(\mathbf{x})\right] = R_1 + R_2 + O_{P_{z_2}}\left(\|\mu - \mu_0\|\,\|\pi - \pi_0\|\right).$$

□

Corollary 4 (Doubly robustness of the DML estimators (Corollary of Thm. 4)**).** *Suppose Assumptions (2,5,6) and AC-gTTI in Def. 4 hold. Suppose either $\pi = \pi_0$ or $\mu = \mu_0$. Then, T^{dml} is an unbiased estimator of* $\mathbb{E}\left[Y|do(\mathbf{x})\right]$.

Proof of Corollary 4. Let π and μ denote the limiting estimator for π_0 and μ_0.

$$T^{dml} := \underbrace{\mathbb{E}_{D_2}\left[\pi(\mathbf{A},\mathbf{X})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\{Y - \mu(\mathbf{A},\mathbf{X})\}\right]}_{:=T^{dml,1}} + \underbrace{\mathbb{E}_{D_1}\left[\mu(\mathbf{A},\mathbf{x}))\right]}_{:=T^{dml,2}}.$$

Let

$$\begin{aligned}
T_0^{dml,1} &:= \mathbb{E}_{P_{z_2}}\left[\pi_0(\mathbf{A},\mathbf{X})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\{Y - \mu_0(\mathbf{A},\mathbf{X})\}\right] \\
T_0^{dml,2} &:= \mathbb{E}_{P_{z_1}}\left[\mu_0(\mathbf{A},\mathbf{x}))\right].
\end{aligned}$$

Under the assumption that samples are i.i.d.,

$$\begin{aligned}\mathbb{E}_{P_{z_2}}\left[T^{dml,1}\right] &= \mathbb{E}_{P_{z_2}}\left[\pi(\mathbf{A},\mathbf{X})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\left\{Y-\mu(\mathbf{A},\mathbf{X})\right\}\right] \\ \mathbb{E}_{P_{z_1}}\left[T^{dml,2}\right] &= \mathbb{E}_{P_{z_1}}\left[\mu(\mathbf{A},\mathbf{x})\right] \\ &= \mathbb{E}_{P_{z_2}}\left[\pi_0(\mathbf{A},\mathbf{X})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\mu(\mathbf{A},\mathbf{X})\right].\end{aligned}$$

Then,

$$\begin{aligned}&\mathbb{E}_{P_{z_2}}\left[T^{dml,1}\right]+\mathbb{E}_{P_{z_1}}\left[T^{dml,2}\right]-\mathbb{E}_{P_{z_2}}\left[T_0^{dml,1}\right]+\mathbb{E}_{P_{z_1}}\left[T_0^{dml,2}\right] \\ &= \mathbb{E}_{P_{z_2}}\left[\pi(\mathbf{A},\mathbf{X})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\{Y-\mu(\mathbf{A},\mathbf{X})\}+\pi_0(\mathbf{A},\mathbf{X})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\left\{\mu(\mathbf{A},\mathbf{X})-\mu_0(\mathbf{A},\mathbf{X})\right\}\right] \\ &= \mathbb{E}_{P_{z_2}}\left[\mathbb{1}_{\mathbf{x}}(\mathbf{X})\left\{\mu_0(\mathbf{A},\mathbf{X})-\mu(\mathbf{A},\mathbf{X})\right\}\left\{\pi_0(\mathbf{A},\mathbf{X})-\pi(\mathbf{A},\mathbf{X})\right\}\right] \\ &= 0,\end{aligned}$$

where the last equation holds under the given condition. □

C.6. Proof of Theorem 5

Definition 7 (Adjustment criterion for Multiple Treatment Interaction (AC-MTI)). An ordered set $\{C_1, W_1, C_2, W_2, \cdots, C_{m-1}, W_{m-1}\}$ satisfies *adjustment criterion for multiple treatment interaction (AC-MTI)* w.r.t. $\{\mathbf{X}, Y\}$ for $\mathbf{X} = \{X_i\}_{i=1}^m$ in G if, for $i = 1, 2, \cdots, m$,

1. $\{X_j\}_{j>i}$ is non-ancestor of $\{\mathbf{X}^{(i)}, \mathbf{W}^{(i)}, \mathbf{C}^{(i)}\}$; and

2. $(Y \perp\!\!\!\perp X_i | \mathbf{C}^{(i-1)}, \mathbf{X}^{(i-1)}, \mathbf{W}^{(i)}, \mathbf{X}^{>i})_{G_{\underline{X_i},\overline{\mathbf{X}^{>i}}}}$; the back-door paths between X_i and Y are blocked by $\mathbf{C}^{(i-1)}, \mathbf{X}^{(i-1)}, \mathbf{W}^{(i)}, \mathbf{X}^{>i}$ in the graph $G_{\overline{\mathbf{X}^{>i}}}$.

Assumption 7 (Positivity Assumption for AC-MTI). *$\{P_{x_i}(W_i, C_i | \mathbf{W}^{(i-1)}, \mathbf{C}^{(i-1)}, \mathbf{X}^{(i-1)})\}_{i=1}^m$, $P_{x_{i+1}}(X_i | \mathbf{W}^{(i)}, \mathbf{C}^{(i)}, \mathbf{X}^{(i-1)})$ for $i = 1, \cdots, m-1$ are strictly positive $\forall \mathbf{x} \in \mathfrak{D}_{\mathbf{X}}$.*

Theorem 5 (Identification through AC-MTI). *Suppose AC-MTI in Def. 7 and Assumption 7 hold. Then, $\mathbb{E}[Y(\mathbf{x})]$ is identifiable from $\{P_{rand(X_i)}(\mathbf{C}^{(i)}, \mathbf{W}^{(i)}, \mathbf{X}^{(i-1)})\}_{i=1}^m$ as follow: Let $\mu_0^m := \mathbb{E}_{P_{x_m}}\left[Y | \mathbf{W}^{(m-1)}, \mathbf{C}^{(m-1)}, \mathbf{X}^{(m-1)}\right]$, and for $i = m-1, \cdots, 2$,*

$$\mu_0^i := \mathbb{E}_{P_{x_i}}\left[\overline{\mu}_0^{i+1} | \mathbf{W}^{(i-1)}, \mathbf{C}^{(i-1)}, \mathbf{X}^{(i-1)}\right],$$

where $\overline{\mu}_0^{i+1} := \mu_0^{i+1}(\mathbf{W}^{(i)}, \mathbf{C}^{(i)}, x_i, \mathbf{X}^{(i-1)})$. Then,

$$\mathbb{E}[Y(\mathbf{x})] = \mathbb{E}_{P_{x_1}}\left[\mu_0^2(W_1, C_1, x_1)\right]. \tag{3}$$

Proof of Theorem 5. Let $A_i := \{W_i, C_i\}$ in this proof. Then, it suffices to show the following equation: For all $i = 1, 2, \cdots, m-1$,

$$\mathbb{E}\left[Y \middle| do(\mathbf{x}^{\geq i}), \mathbf{A}^{(i-1)}, \mathbf{x}^{(i-1)}\right] = \mathbb{E}\left[\mathbb{E}\left[Y \middle| do(\mathbf{x}^{\geq i+1}), \mathbf{A}^{(i)}, \mathbf{x}^{(i)}\right] \middle| do(x_i), \mathbf{A}^{(i-1)}, \mathbf{x}^{(i-1)}\right].$$

It holds as follow:

$$\begin{aligned}
&\mathbb{E}\left[Y \middle| do(\mathbf{x}^{\geq i}), \mathbf{A}^{(i-1)}, \mathbf{x}^{(i-1)}\right] \\
&\stackrel{1}{=} \mathbb{E}\left[\mathbb{E}\left[Y \middle| do(\mathbf{x}^{\geq i}), \mathbf{A}^{(i)}, \mathbf{x}^{(i-1)}\right] \middle| do(\mathbf{x}^{\geq i}), \mathbf{A}^{(i-1)}, \mathbf{x}^{(i-1)}\right] \\
&\stackrel{2}{=} \mathbb{E}\left[\mathbb{E}\left[Y \middle| do(\mathbf{x}^{\geq i+1}), \mathbf{A}^{(i)}, \mathbf{x}^{(i)}\right] \middle| do(\mathbf{x}^{\geq i}), \mathbf{A}^{(i-1)}, \mathbf{x}^{(i-1)}\right] \\
&\stackrel{3}{=} \mathbb{E}\left[\mathbb{E}\left[Y \middle| do(\mathbf{x}^{\geq i+1}), \mathbf{A}^{(i)}, \mathbf{x}^{(i)}\right] \middle| do(x_i), \mathbf{A}^{(i-1)}, \mathbf{x}^{(i-1)}\right],
\end{aligned}$$

where

- $\stackrel{1}{=}$ holds by marginalizing over $\mathbf{A}_i$.
- $\stackrel{2}{=}$ holds as follow:

  $$\mathbb{E}\left[Y \middle| do(\mathbf{x}^{\geq i}), \mathbf{A}^{(i)}, \mathbf{x}^{(i-1)}\right] = \mathbb{E}\left[Y \middle| do(\mathbf{x}^{\geq i+1}), \mathbf{A}^{(i)}, \mathbf{x}^{(i)}\right],$$

  since $(Y \perp\!\!\!\perp X_i | \mathbf{A}^{(i)}, \mathbf{X}^{(i-1)}, \mathbf{X}^{\geq i+1})_{G_{\underline{X_i}, \overline{\mathbf{X}^{\geq i+1}}}}$ by the given condition and the positivity condition.
- $\stackrel{3}{=}$ holds because $\mathbf{X}^{\geq i+1}$ is not an ancestor of $\mathbf{A}^{(i)}, \mathbf{X}^{(i)}$.

□

C.7. Proof of Theorem 6 and Corollary 6

Definition 8 (Nuisances for MTI). Nuisance functions for AC-MTI are defined as follows: For a fixed $\mathbf{x} := \{x_1, \cdots, x_m\} \in \mathfrak{D}_{\mathbf{X}}$, let $\{\mu^i\}_{i=2}^m$ and $\{\overline{\mu}^i\}_{i=2}^m$ be the nuisances defined in Thm. 5. For $i = 1, \cdots, m-1$, $\pi_0^i := \frac{P_{x_i}(W_i | C_i, \mathbf{C}^{(i-1)}, \mathbf{X}^{(i-1)}, \mathbf{W}^{(i-1)})}{P_{x_m}(W_i, X_i | C_i, \mathbf{C}^{(i-1)}, \mathbf{X}^{(i-1)}, \mathbf{W}^{(i-1)})}$, and $\pi_0^{(i)} := \prod_{j=1}^{i} \pi_0^j(\mathbf{W}^{(j)}, \mathbf{C}^{(j)}, \mathbf{X}^{(j)})$. We will use $\pi^i(\mathbf{W}^{(i)}, \mathbf{C}^{(i)}, \mathbf{X}^{(i)}) > 0$ and $\mu^i(\mathbf{W}^{(i-1)}, \mathbf{C}^{(i-1)}, \mathbf{X}^{(i-1)})$ for any arbitrary[5] finite functions.

Definition 9 (AC-MTI estimators). Let D_i denote samples following $P_{\text{rand}(X_i)}(\mathbf{C}^{(i)}, \mathbf{W}^{(i)}, \mathbf{X}^{(i)})$ for $i = 1, 2, \cdots, m$. For a fixed $x_i \in \mathfrak{D}_{X_i}$, let D_{x_i} denote the subsamples of D_i such that $X_i = x_i$. Let $A_i := \{W_i, C_i\}$ and $V_i := \{A_i, X_i\}$. Let $\mu^{m+1} := Y$. Let $\mathbb{1}_{\mathbf{x}}^{i-1} := \mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)})$ for $i = 2, \cdots, m$. Then {REG, PW, DML} estimators are defined as follow:

$$\begin{aligned}
T^{reg} &:= \mathbb{E}_{D_{x_1}}\left[\mu^2(W_1, C_1, x_1))\right], \\
T^{pw} &:= \mathbb{E}_{D_{x_m}}\left[\pi^{(m-1)} \mathbb{1}_{\mathbf{x}}(\mathbf{X}) Y\right], \\
T^{dml} &:= \sum_{i=2}^{m} \mathbb{E}_{D_{x_i}}\left[\pi^{(i-1)} \mathbb{1}_{\mathbf{x}}^{i-1} \{\overline{\mu}^{i+1} - \mu^i\}\right] + \mathbb{E}_{D_{x_1}}\left[\overline{\mu}^2\right].
\end{aligned}$$

Assumption 2 (Sample-splitting). *Samples for training nuisances and evaluating the estimators equipped with the trained nuisance are separate and independent*

Assumption 8 (L_2 consistency of nuisances). *Estimated nuisances are L_2-consistent; Specifically,*

$$\begin{aligned}
\|\mu^{i+1} - \mu_0^{i+1}\|_{P_{x_i}} &= o_{P_{x_i}}(1), \ \forall i \in \{1, 2, \cdots, m-1\} \\
\|\mu^i - \mu_0^i\|_{P_{x_i}} &= o_{P_{x_i}}(1), \ \forall i \in \{2, \cdots, m\} \\
\|\pi^i - \pi^i\|_{P_{x_{i+1}}} &= o_{P_{x_{i+1}}}(1), \ \forall i \in \{1, \cdots, m-1\}.
\end{aligned}$$

[5]Throughout the paper, $\mu^i, \overline{\mu}^i, \pi^i$ may be understood as estimated nuisances for $\mu_0^i, \overline{\mu}_0^i, \pi_0^i$.

Assumption 9 (Multiple experiments represent the same population). *For any fixed $i, j \in \{1, 2, \cdots, m-1\}$ s.t. $j > i$ and any fixed $x_i, x_j \in \mathfrak{D}_{X_i, X_j}$, the baseline covariates C_i's distribution satisfies the following: $P_{x_i}(C_i | \mathbf{C}^{(i-1)}, \mathbf{X}^{(j-1)}, \mathbf{W}^{(j-1)}) = P_{x_j}(C_i | \mathbf{C}^{(i-1)}, \mathbf{X}^{(i-1)}, \mathbf{W}^{(i-1)})$.*

Lemma C.4 (Error analysis of the REG estimator for MTI). *Suppose Assumptions (2,8) hold. Let T^{reg} denote the estimator defined in Def. 9. Then,*

$$T^{reg} - \mathbb{E}\left[Y | do(\mathbf{x})\right] = R_1 + O_{P_{x_1}}(\left\|\mu^2 - \mu_0^2\right\|),$$

where R_1 is the random variable such that $\sqrt{n_1} R_1$ converges in distribution to the mean-zero normal random variable, where $n_1 \coloneqq |D_{x_1}|$.

Proof of Lemma C.4. We first note that, by Theorem 5,

$$\mathbb{E}_{P_{x_1}}\left[\mu_0^2(W_1, C_1, x_1)\right] = \mathbb{E}\left[Y | do(\mathbf{x})\right].$$

By Lemma C.3,

$$\begin{aligned}
&T^{reg} - \mathbb{E}\left[Y | do(\mathbf{x})\right] \\
&= T^{reg} - \mathbb{E}_{P_{x_1}}\left[\mu_0^2(W_1, C_1, x_1)\right] \\
&\overset{1}{=} \underbrace{\mathbb{E}_{P_{x_1} - D_{x_1}}\left[\mu_0^2(W_1, C_1, x_1)\right] + \mathbb{E}_{P_{x_1} - D_{x_1}}\left[\mu_0^2(W_1, C_1, x_1) - \mu^2(W_1, C_1, x_1)\right]}_{\coloneqq R_1} \\
&+ \mathbb{E}_{P_{x_1}}\left[\mu_0^2(W_1, C_1, x_1) - \mu^2(W_1, C_1, x_1)\right] \\
&= R_1 + \mathbb{E}_{P_{x_1}}\left[\mu_0^2(W_1, C_1, x_1) - \mu^2(W_1, C_1, x_1)\right] \\
&= R_1 + O_{P_{x_1}}(\|\mu^2 - \mu_0^2\|).
\end{aligned}$$

$\overset{1}{=}$ holds by Lemmas (C.1, C.3), and the last equation holds by Cauchy-Schwartz inequality. □

Lemma C.5 (Error analysis of the PW estimator for MTI). *Suppose Assumptions (2,8,9) hold. Let T^{pw} denote the estimator defined in Def. 9. Then,*

$$T^{pw} - \mathbb{E}\left[Y | do(\mathbf{x})\right] = R_m + O_{P_{x_m}}(\|\pi^{(m-1)} - \pi_0^{(m-1)}\|),$$

where R_m is the random variable such that $\sqrt{n_m} R_m$ converges in distribution to the mean-zero normal random variable, where $n_m \coloneqq |D_{x_m}|$.

Proof of Lemma C.5. Throughout the proof, we set $A_i \coloneqq \{C_i, W_i\}$ for all $i = 1, 2, \cdots, m$. We will use $V_i \coloneqq \{A_i, X_i\}$. In the proof, we tentatively assume

$$\mathbb{E}_{P_{x_m}}\left[\pi_0^{(m-1)}(\mathbf{W}^{(m-1)}, \mathbf{C}^{(m-1)}, \mathbf{X}^{(m-1)}) \mathbb{1}_{\mathbf{x}}(\mathbf{X}) Y\right] = \mathbb{E}\left[Y | do(\mathbf{x})\right]. \tag{C.3}$$

Then, by Lemma C.3,

$$
\begin{aligned}
&T^{pw} - \mathbb{E}\left[Y|do(\mathbf{x})\right] \\
&= T^{pw} - \mathbb{E}_{P_{x_m}}\left[\pi_0^{(m-1)}(\mathbf{W}^{(m-1)}, \mathbf{C}^{(m-1)}, \mathbf{X}^{(m-1)})\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right] \\
&\overset{1}{=} \underbrace{\mathbb{E}_{P_{x_m}-D_{x_m}}\left[\pi_0^{(m-1)}(\mathbf{V}^{(m-1)})\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right] + \mathbb{E}_{P_{x_m}-D_{x_m}}\left[\left\{\pi_0^{(m-1)}(\mathbf{V}^{(m-1)}) - \pi^{(m-1)}(\mathbf{V}^{(m-1)})\right\}\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right]}_{:=R_m} \\
&+ \mathbb{E}_{P_{x_m}}\left[\left\{\pi_0^{(m-1)}(\mathbf{V}^{(m-1)}) - \pi^{(m-1)}(\mathbf{V}^{(m-1)})\right\}\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right] \\
&= R_m + \mathbb{E}_{P_{x_m}}\left[\left\{\pi_0^{(m-1)}(\mathbf{V}^{(m-1)}) - \pi^{(m-1)}(\mathbf{V}^{(m-1)})\right\}\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right] \\
&= R_m + O_{P_{x_m}}(\|\pi^{(m-1)} - \pi_0^{(m-1)}\|),
\end{aligned}
$$

where $\overset{1}{=}$ holds by Lemmas (C.1, C.3). The last equation holds by Cauchy-Schwartz inequality.

We now prove Eq. (C.3). We first show the following: For $i = 2, \cdots, m$,

$$
\mathbb{E}\left[Y|do(\mathbf{x})\right] = \mathbb{E}_{P_{x_i}}\left[\prod_{j=1}^{i-1}\frac{P_{x_j}(\mathbf{A}^{(j)}, \mathbf{X}^{(j-1)})}{P_{x_{j+1}}(\mathbf{A}^{(j)}, \mathbf{X}^{(j)})}\mu_0^i(\mathbf{A}^{(i-1)}, \mathbf{X}^{(i-1)})\mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)})\right]. \tag{C.4}
$$

It holds for $i = 2$ as follow:

$$
\mathbb{E}_{P_{x_2}}\left[\frac{P_{x_1}(A_1)}{P_{x_2}(A_1, X_1)}\mu_0^2(A_1, X_1)\mathbb{1}_{x_1}(X_1)\right] = \mathbb{E}_{P_{x_1}}\left[\mu_0^2(A_1, x_1)\right] = \mathbb{E}\left[Y|do(\mathbf{x})\right],
$$

where the last equation holds by Lemma C.4. Now, we make the following induction hypothesis: For some $i-1 \in \{2, 3, \cdots, m-1\}$, suppose

$$
\mathbb{E}\left[Y|do(\mathbf{x})\right] \overset{\text{induction hypothesis}}{=} \mathbb{E}_{P_{x_{i-1}}}\left[\prod_{j=1}^{i-2}\frac{P_{x_j}(\mathbf{A}^{(j)}, \mathbf{X}^{(j-1)})}{P_{x_{j+1}}(\mathbf{A}^{(j)}, \mathbf{X}^{(j)})}\mu_0^{i-1}(\mathbf{A}^{(i-2)}, \mathbf{X}^{(i-2)})\mathbb{1}_{\mathbf{x}^{(i-2)}}(\mathbf{X}^{(i-2)})\right].
$$

Then,

$$
\begin{aligned}
&\mathbb{E}_{P_{x_{i-1}}}\left[\prod_{j=1}^{i-2}\frac{P_{x_j}(\mathbf{A}^{(j)}, \mathbf{X}^{(j-1)})}{P_{x_{j+1}}(\mathbf{A}^{(j)}, \mathbf{X}^{(j)})}\mu_0^{i-1}(\mathbf{A}^{(i-2)}, \mathbf{X}^{(i-2)})\mathbb{1}_{\mathbf{x}^{(i-2)}}(\mathbf{X}^{(i-2)})\right] \\
&= \mathbb{E}_{P_{x_{i-1}}}\left[\prod_{j=1}^{i-2}\frac{P_{x_j}(\mathbf{A}^{(j)}, \mathbf{X}^{(j-1)})}{P_{x_{j+1}}(\mathbf{A}^{(j)}, \mathbf{X}^{(j)})}\mathbb{E}_{P_{x_{i-1}}}\left[\mu_0^i(\mathbf{A}^{(i-1)}, x_{i-1}, \mathbf{X}^{(i-2)})|\mathbf{A}^{(i-2)}, \mathbf{X}^{(i-2)}\right]\mathbb{1}_{\mathbf{x}^{(i-2)}}(\mathbf{X}^{(i-2)})\right] \\
&\overset{1}{=} \mathbb{E}_{P_{x_{i-1}}}\left[\prod_{j=1}^{i-2}\frac{P_{x_j}(\mathbf{A}^{(j)}, \mathbf{X}^{(j-1)})}{P_{x_{j+1}}(\mathbf{A}^{(j)}, \mathbf{X}^{(j)})}\mu_0^i(\mathbf{A}^{(i-1)}, x_{i-1}, \mathbf{X}^{(i-2)})\mathbb{1}_{\mathbf{x}^{(i-2)}}(\mathbf{X}^{(i-2)})\right] \\
&\overset{2}{=} \mathbb{E}_{P_{x_{i-1}}}\left[\prod_{j=1}^{i-2}\frac{P_{x_j}(\mathbf{A}^{(j)}, \mathbf{X}^{(j-1)})}{P_{x_{j+1}}(\mathbf{A}^{(j)}, \mathbf{X}^{(j)})}\mu_0^i(\mathbf{A}^{(i-1)}, \mathbf{X}^{(i-1)})\mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)})\right] \\
&= \mathbb{E}_{P_{x_i}}\left[\prod_{j=1}^{i-2}\frac{P_{x_j}(\mathbf{A}^{(j)}, \mathbf{X}^{(j-1)})}{P_{x_{j+1}}(\mathbf{A}^{(j)}, \mathbf{X}^{(j)})}\frac{P_{x_{i-1}}(\mathbf{A}^{(i-1)}, \mathbf{X}^{(i-2)})}{P_{x_i}(\mathbf{A}^{(i-1)}, \mathbf{X}^{(i-1)})}\mu_0^i(\mathbf{A}^{(i-1)}, \mathbf{X}^{(i-1)})\mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)})\right] \\
&= \mathbb{E}_{P_{x_i}}\left[\prod_{j=1}^{i-1}\frac{P_{x_j}(\mathbf{A}^{(j)}, \mathbf{X}^{(j-1)})}{P_{x_{j+1}}(\mathbf{A}^{(j)}, \mathbf{X}^{(j)})}\mu_0^i(\mathbf{A}^{(i-1)}, \mathbf{X}^{(i-1)})\mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)})\right],
\end{aligned}
$$

where

- $\stackrel{1}{=}$ holds by the law of total expectation.
- $\stackrel{2}{=}$ holds since the expectation is over $P_{x_{i-1}}$.

Therefore, Eq. (C.4) holds. By plugging $i = m$, we have

$$\begin{aligned}
\mathbb{E}\left[Y|do(\mathbf{x})\right] &= \mathbb{E}_{P_{x_m}}\left[\prod_{j=1}^{m-1}\frac{P_{x_j}(\mathbf{A}^{(j)},\mathbf{X}^{(j-1)})}{P_{x_{j+1}}(\mathbf{A}^{(j)},\mathbf{X}^{(j)})}\mu_0^m(\mathbf{A}^{(m-1)},\mathbf{X}^{(m-1)})\mathbb{1}_{\mathbf{x}^{(m-1)}}(\mathbf{X}^{(m-1)})\right] \\
&= \mathbb{E}_{P_{x_m}}\left[\prod_{j=1}^{m-1}\frac{P_{x_j}(\mathbf{A}^{(j)},\mathbf{X}^{(j-1)})}{P_{x_{j+1}}(\mathbf{A}^{(j)},\mathbf{X}^{(j)})}\mathbb{E}_{P_{x_m}}\left[Y|\mathbf{A}^{(m-1)},\mathbf{X}^{(m-1)}\right]\mathbb{1}_{\mathbf{x}^{(m-1)}}(\mathbf{X}^{(m-1)})\right] \\
&= \mathbb{E}_{P_{x_m}}\left[\prod_{j=1}^{m-1}\frac{P_{x_j}(\mathbf{A}^{(j)},\mathbf{X}^{(j-1)})}{P_{x_{j+1}}(\mathbf{A}^{(j)},\mathbf{X}^{(j)})}\mathbb{1}_{\mathbf{x}^{(m-1)}}(\mathbf{X}^{(m-1)})Y\right].
\end{aligned}$$

Finally,

$$\begin{aligned}
\prod_{j=1}^{m-1}\frac{P_{x_j}(\mathbf{A}^{(j)},\mathbf{X}^{(j-1)})}{P_{x_{j+1}}(\mathbf{A}^{(j)},\mathbf{X}^{(j)})} &= \frac{P_{x_1}(\mathbf{A}^{(1)})P_{x_2}(\mathbf{A}^{(2)},\mathbf{X}^{(1)})\cdots P_{x_{m-1}}(\mathbf{A}^{(m-1)},\mathbf{X}^{(m-2)})}{P_{x_2}(\mathbf{A}^{(1)},\mathbf{X}^{(1)})P_{x_3}(\mathbf{A}^{(2)},\mathbf{X}^{(2)})\cdots P_{x_{m-1}}(\mathbf{A}^{(m-2)},\mathbf{X}^{(m-2)})P_{x_m}(\mathbf{A}^{(m-1)},\mathbf{X}^{(m-1)})} \\
&= \frac{1}{P_{x_m}(\mathbf{A}^{(m-1)},\mathbf{X}^{(m-1)})}\prod_{j=1}^{m-1}P_{x_j}(A_j|\mathbf{A}^{(j-1)},\mathbf{X}^{(j-1)}) \\
&= \prod_{j=1}^{m-1}\frac{P_{x_j}(A_j|\mathbf{A}^{(j-1)},\mathbf{X}^{(j-1)})}{P_{x_m}(A_j,X_j|\mathbf{A}^{(j-1)},\mathbf{X}^{(j-1)})} \\
&= \prod_{j=1}^{m-1}\frac{P_{x_j}(C_j,W_j|\mathbf{A}^{(j-1)},\mathbf{X}^{(j-1)})}{P_{x_m}(C_j,W_j,X_j|\mathbf{A}^{(j-1)},\mathbf{X}^{(j-1)})} \\
&= \prod_{j=1}^{m-1}\frac{P_{x_j}(W_j|C_j,\mathbf{A}^{(j-1)},\mathbf{X}^{(j-1)})P_{x_j}(C_j|\mathbf{A}^{(j-1)},\mathbf{X}^{(j-1)})}{P_{x_m}(W_j,X_j|C_j,\mathbf{A}^{(j-1)},\mathbf{X}^{(j-1)})P_{x_m}(C_j|\mathbf{A}^{(j-1)},\mathbf{X}^{(j-1)})} \\
&= \prod_{j=1}^{m-1}\frac{P_{x_j}(W_j|C_j,\mathbf{A}^{(j-1)},\mathbf{X}^{(j-1)})}{P_{x_m}(W_j,X_j|C_j,\mathbf{A}^{(j-1)},\mathbf{X}^{(j-1)})} \\
&= \pi_0^{(m-1)}(\mathbf{A}^{(m-1)},\mathbf{X}^{(m-1)}).
\end{aligned}$$

□

Lemma C.6 (Bias Analysis of the DML estimator for MTI). *Suppose Assumptions (2,8,9) hold. For $i = 1, 2, \cdots, m$, let $A_i := \{C_i, W_i\}$ and $V_i := \{A_i, X_i\}$. For $i = 1, \cdots, m$, let $B_i := \{A_i, X_{i-1}\}$ where $X_0 := \emptyset$. Let $T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^m)$ be defined as follow:*

$$\begin{aligned}
&T^{dml}(\{\pi^k\}_{k=1}^{m-1},\{\mu^k\}_{k=2}^m) \\
&:= \sum_{i=2}^{m}\mathbb{E}_{P_{x_i}}\left[\pi^{(i-1)}(\mathbf{V}^{(i-1)})\mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)})\left\{\mu^{i+1}(\mathbf{B}^{(i)},x_i)-\mu^i(\mathbf{B}^{(i-1)},X_{i-1})\right\}\right]+\mathbb{E}_{P_{x_1}}\left[\mu^2(\mathbf{B}^{(1)},x_1)\right]. \quad \text{(C.5)}
\end{aligned}$$

Then,

$$T^{dml}(\{\pi^k\}_{k=1}^{m-1},\{\mu^k\}_{k=2}^{m}) - \mathbb{E}\left[Y|do(\mathbf{x})\right] = \sum_{i=2}^{m} O_{P_{x_i}}\left(\left\|\mu^i-\mu_0^i\right\|\left\|\pi^{i-1}-\pi_0^{i-1}\right\|\right) \tag{C.6}$$

Proof. We follow the proof technique used in (Rotnitzky et al., 2017). We first note that

$$T^{dml}(\{\pi_0^k\}_{k=1}^{m-1},\{\mu_0^k\}_{k=2}^{m}) = \mathbb{E}\left[Y|do(\mathbf{x})\right]. \tag{C.7}$$

It's easy to witness Eq. (C.7) because, for $i=2,3,\cdots,m$,

$$\begin{aligned}
&\mathbb{E}_{P_{x_i}}\left[\pi_0^{(i-1)}(\mathbf{V}^{(i-1)})\mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)})\left\{\mu_0^{i+1}(\mathbf{B}^{(i)},x_i)-\mu_0^{i}(\mathbf{B}^{(i-1)},X_{i-1})\right\}\right]\\
&\stackrel{1}{=}\mathbb{E}_{P_{x_i}}\left[\mathbb{E}_{P_{x_i}}\left[\pi_0^{(i-1)}(\mathbf{V}^{(i-1)})\mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)})\left\{\mu_0^{i+1}(\mathbf{B}^{(i)},x_i)-\mu_0^{i}(\mathbf{B}^{(i-1)},X_{i-1})\right\}\middle|\mathbf{B}^{(i-1)},X_{i-1}\right]\right]\\
&=\mathbb{E}_{P_{x_i}}\left[\pi_0^{(i-1)}(\mathbf{V}^{(i-1)})\mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)})\left\{\mathbb{E}_{P_{x_i}}\left[\mu_0^{i+1}(\mathbf{B}^{(i)},x_i)|\mathbf{B}^{(i-1)},X_{i-1}\right]-\mu_0^{i}(\mathbf{B}^{(i-1)},X_{i-1})\right\}\right]\\
&=\mathbb{E}_{P_{x_i}}\left[\pi_0^{(i-1)}(\mathbf{V}^{(i-1)})\mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)})\left\{\mu_0^{i}(\mathbf{B}^{(i-1)},X_{i-1})-\mu_0^{i}(\mathbf{B}^{(i-1)},X_{i-1})\right\}\right]\\
&=0,
\end{aligned}$$

where the equation $\stackrel{1}{=}$ holds by the law of total expectation. Therefore,

$$T^{dml}(\{\pi_0^k\}_{k=1}^{m-1},\{\mu_0^k\}_{k=2}^{m}) = \mathbb{E}_{P_{x_1}}\left[\mu_0^2(\mathbf{B}^{(1)},x_1)\right] = \mathbb{E}\left[Y|do(\mathbf{x})\right],$$

where the second equation holds by Lemma C.4. Therefore, it suffices to prove the following to show Eq. (C.6):

$$T^{dml}(\{\pi^k\}_{k=1}^{m-1},\{\mu^k\}_{k=2}^{m}) - T^{dml}(\{\pi_0^k\}_{k=1}^{m-1},\{\mu_0^k\}_{k=2}^{m}) = \sum_{i=2}^{m} O_{P_{x_i}}\left(\left\|\mu^i-\mu_0^i\right\|\left\|\pi^{i-1}-\pi_0^{i-1}\right\|\right). \tag{C.8}$$

For $i=1,2,\cdots,m-1$, we define a quantity

$$\omega_0^i(\mathbf{B}^{(i)}) \coloneqq \frac{P_{x_i}(\mathbf{B}^{(i)})}{P_{x_m}(\mathbf{B}^{(i)})}.$$

We note that $\omega_0^i(\mathbf{B}^{(i)})$ is related with π as follow:

$$\omega_0^i(\mathbf{B}^{(i)}) = \pi_0^i(\mathbf{V}^{(i)})P_{x_m}(X_i|\mathbf{B}^{(i)}). \tag{C.9}$$

To witness, consider the following:

$$\begin{aligned}
\omega_0^i(\mathbf{B}^{(i)}) &= \frac{P_{x_i}(A_i|\mathbf{V}^{(i-1)})P_{x_i}(\mathbf{V}^{(i-1)})}{P_{x_m}(A_i|\mathbf{V}^{(i-1)})P_{x_m}(\mathbf{V}^{(i-1)})}\\
&\stackrel{2}{=}\frac{P_{x_i}(A_i|\mathbf{V}^{(i-1)})}{P_{x_m}(A_i|\mathbf{V}^{(i-1)})}\\
&=\frac{P_{x_i}(W_i,C_i|\mathbf{W}^{(i-1)},\mathbf{C}^{(i-1)},\mathbf{X}^{(i-1)})}{P_{x_m}(W_i,C_i|\mathbf{W}^{(i-1)},\mathbf{C}^{(i-1)},\mathbf{X}^{(i-1)})}\\
&\stackrel{3}{=}\frac{P_{x_i}(W_i|C_i,\mathbf{W}^{(i-1)},\mathbf{C}^{(i-1)},\mathbf{X}^{(i-1)})}{P_{x_m}(W_i|C_i,\mathbf{W}^{(i-1)},\mathbf{C}^{(i-1)},\mathbf{X}^{(i-1)})}\\
&=\pi_0^i(\mathbf{W}^{(i)},\mathbf{C}^{(i)},\mathbf{X}^{(i)})\frac{P_{x_m}(W_i,X_i|C_i,\mathbf{W}^{(i-1)},\mathbf{C}^{(i-1)},\mathbf{X}^{(i-1)})}{P_{x_m}(W_i|C_i,\mathbf{W}^{(i-1)},\mathbf{C}^{(i-1)},\mathbf{X}^{(i-1)})}
\end{aligned}$$

$$= \pi_0^i(\mathbf{V}^{(i)}) P_{x_m}(X_i | \mathbf{B}^{(i)}),$$

where

- $\stackrel{2}{=}$ holds since X_i is non-descendent to $\mathbf{V}^{(i-1)}$, so that $P_{x_i}(\mathbf{V}^{(i-1)}) = P_{x_m}(\mathbf{V}^{(i-1)})$.
- $\stackrel{3}{=}$ holds by Assumption 9.

To simplify the notation, we sometimes simply denote $\omega_0^i(\mathbf{B}^{(i)})$ as ω_0^i; $\mu^i(\mathbf{B}^{(i-1)}, X_{i-1})$ as μ^i; $\mu^i(\mathbf{B}^{(i-1)}, x_{i-1})$ as $\overline{\mu}^i$; and $\pi^i(\mathbf{V}^{(i)})$ as π^i.

Then, $T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m})$ in Eq. (C.5) can be rewritten as

$$T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m}) = \sum_{i=2}^{m} \mathbb{E}_{P_{x_m}}\left[\omega_0^i \pi^{(i-1)} \mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)}) \left\{\overline{\mu}^{i+1} - \mu^i\right\} + \omega_0^1 \overline{\mu}^2\right], \tag{C.10}$$

where $\overline{\mu}^{m+1} \coloneqq Y$.

For each $k = 1, 2, \cdots, m$, we define a quantity Q_k as follow:

$$Q_k \coloneqq Q_k(\{\pi^j\}_{j=k}^{m-1}, \{\mu^j\}_{j=k+1}^{m}) \coloneqq \omega_0^k \overline{\mu}^{k+1} + \sum_{i=k+1}^{m} \omega_0^i \pi^{(k:i-1)} \mathbb{1}_{\mathbf{x}^{(k:i-1)}}(\mathbf{X}^{(k:i-1)}) \left\{\overline{\mu}^{i+1} - \mu^i\right\}. \tag{C.11}$$

Note $Q_m = Y$ and $\mathbb{E}_{P_{x_m}}[Q_1] = T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m})$ defined in Eq. (C.10). We note that

$$\begin{aligned}
\mathbb{E}_{P_{x_m}}\left[Q_1 - \omega_0^1 \overline{\mu}_0^2\right] &\stackrel{4}{=} T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m}) - \mathbb{E}_{P_{x_1}}\left[\mu_0^2(\mathbf{B}^{(1)}, x_1)\right] \\
&\stackrel{5}{=} T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m}) - \mathbb{E}\left[Y | do(\mathbf{x})\right] \\
&= \text{l.h.s. of Eq. (C.6)},
\end{aligned}$$

where

- $\stackrel{4}{=}$ holds since $\mathbb{E}_{P_{x_m}}\left[\omega_0^1(\mathbf{B}^{(1)}) \mu^2(\mathbf{B}^{(1)}, x_1)\right] = \mathbb{E}_{P_{x_1}}\left[\mu^1(\mathbf{B}^{(1)}, x_1)\right]$.
- $\stackrel{5}{=}$ holds by Lemma C.4.

Motivating from the fact that $\mathbb{E}_{P_{x_m}}\left[Q_1 - \omega_0^1 \overline{\mu}_0^2\right] = $ l.h.s. of Eq. (C.6), we establish a following induction hypothesis. For $\overline{P}_{x_m}^{i-1} \coloneqq P_{x_m}(\cdot | \mathbf{V}^{(i-1)})$, the induction hypothesis is given as follow:

$$\text{Hypothesis: } \mathbb{E}_{\overline{P}_{x_m}^{k-1}}\left[Q_k - \omega_0^k \overline{\mu}_0^{k+1}\right] = \sum_{i=k+1}^{m} O_{\overline{P}_{x_i}^{k-1}}\left(\left\|\mu^i - \mu_0^i\right\| \left\|\pi^{i-1} - \pi_0^{i-1}\right\|\right), \text{ for } k \in \{2, \cdots, m-1\} \tag{C.12}$$

We first verify the hypothesis Eq. (C.12) for $k = m - 1$.

$$\begin{aligned}
&\mathbb{E}_{\overline{P}_{x_m}^{m-2}}\left[Q_{m-1} - \omega_0^{m-1} \overline{\mu}_0^m\right] \\
&= \mathbb{E}_{\overline{P}_{x_m}^{m-2}}\left[\omega_0^{m-1} \overline{\mu}^m + \pi^{m-1} \mathbb{1}_{x_{m-1}}(X_{m-1}) \left\{Y - \mu^m\right\} - \omega_0^{m-1} \overline{\mu}_0^m\right] \\
&\stackrel{6}{=} \mathbb{E}_{\overline{P}_{x_m}^{m-2}}\left[\omega_0^{m-1} \overline{\mu}^m + \pi^{m-1} \mathbb{1}_{x_{m-1}}(X_{m-1}) \left\{\mu_0^m - \mu^m\right\} - \omega_0^{m-1} \overline{\mu}_0^m\right] \\
&= \mathbb{E}_{\overline{P}_{x_m}^{m-2}}\left[\omega_0^{m-1} \left\{\overline{\mu}^m - \overline{\mu}_0^m\right\} + \pi^{m-1} \mathbb{1}_{x_{m-1}}(X_{m-1}) \left\{\mu_0^m - \mu^m\right\}\right] \\
&\stackrel{7}{=} \mathbb{E}_{\overline{P}_{x_m}^{m-2}}\left[\omega_0^{m-1} \frac{\mathbb{1}_{x_{m-1}}(X_{m-1})}{P_{x_m}(X_{m-1} | \mathbf{B}^{(m-1)})} \left\{\mu^m - \mu_0^m\right\} + \pi^{m-1} \mathbb{1}_{x_{m-1}}(X_{m-1}) \left\{\mu_0^m - \mu^m\right\}\right]
\end{aligned}$$

$$\begin{aligned}
&\overset{8}{=} \mathbb{E}_{\overline{P}^{m-2}_{x_m}} \left[\pi_0^{m-1} \mathbb{1}_{x_{m-1}}(X_{m-1}) \left\{\mu^m - \mu_0^m\right\} + \pi^{m-1} \mathbb{1}_{x_{m-1}}(X_{m-1}) \left\{\mu_0^m - \mu^m\right\}\right] \\
&= \mathbb{E}_{\overline{P}^{m-2}_{x_m}} \left[\mathbb{1}_{x_{m-1}}(X_{m-1}) \left\{\mu^m - \mu_0^m\right\} \left\{\pi_0^{m-1} - \pi^{m-1}\right\}\right] \\
&= \mathbb{E}_{P_{x_m}} \left[\mathbb{1}_{x_{m-1}}(X_{m-1}) \left\{\mu^m - \mu_0^m\right\} \left\{\pi_0^{m-1} - \pi^{m-1}\right\} \middle| \mathbf{V}^{(m-2)}\right] \\
&\overset{9}{=} O_{\overline{P}^{m-2}_{x_m}} \left(\|\mu^m - \mu_0^m\| \left\|\pi^{m-1} - \pi_0^{m-1}\right\|\right),
\end{aligned}$$

where

- $\overset{6}{=}$ holds by the total law of expectation.
- $\overset{7}{=}$ holds since

$$\mathbb{E}_{P_{x_m}} \left[\mu^{m-1}(\mathbf{B}^{(m-1)}, x_{m-1}) \middle| \mathbf{V}^{(m-2)}\right] = \mathbb{E}_{P_{x_m}} \left[\mu^{m-1}(\mathbf{B}^{(m-1)}, X_{m-1}) \frac{\mathbb{1}_{x_{m-1}}(X_{m-1})}{P_{x_m}(X_{m-1}|\mathbf{B}^{(m-1)})} \middle| \mathbf{V}^{(m-2)}\right].$$

- $\overset{8}{=}$ holds by the definition of ω_0^{m-1}.
- $\overset{9}{=}$ holds by applying Cauchy-Schwarz inequality.

Now, we suppose Eq. (C.12) holds for some $k+1 \in \{2, \cdots, m-1\}$. Then, we will show that Eq. (C.12) holds for k. Toward this end, we first rewrite Q_k in Eq. (C.11) in a recursive form. For any $k+1 \in \{2, \cdots, m-1\}$, the following relation can be derived from Eq. (C.11):

$$\pi^k \mathbb{1}_{x_k}(X_k) \left\{Q_{k+1} - \omega_0^{k+1} \overline{\mu}^{k+2}\right\} = \sum_{i=k+2}^{m} \omega_0^i \pi^{(k:i-1)} \mathbb{1}_{\mathbf{x}^{(k:i-1)}}(\mathbf{X}^{(k:i-1)}) \left\{\overline{\mu}^{i+1} - \mu^i\right\}.$$

Therefore, for each $k = 1, 2, \cdots, m-1$,

$$Q_k(\{\pi^j\}_{j=k}^{m-1}, \{\mu^j\}_{j=k+1}^m) = \omega_0^k \overline{\mu}^{k+1} + \omega_0^{k+1} \pi^k \mathbb{1}_{x_k}(X_k) \left\{\overline{\mu}^{k+2} - \mu^{k+1}\right\} + \pi^k \mathbb{1}_{x_k}(X_k) \left\{Q_{k+1} - \omega_0^{k+1} \overline{\mu}^{k+2}\right\}.$$

Then,

$$\begin{aligned}
&\mathbb{E}_{\overline{P}^{k-1}_{x_m}} \left[Q_k - \omega_0^k \overline{\mu}_0^{k+1}\right] \\
&= \mathbb{E}_{\overline{P}^{k-1}_{x_m}} \left[\omega_0^k \overline{\mu}^{k+1} + \omega_0^{k+1} \pi^k \mathbb{1}_{x_k}(X_k) \left\{\overline{\mu}^{k+2} - \mu^{k+1}\right\} + \pi^k \mathbb{1}_{x_k}(X_k) \left\{Q_{k+1} - \omega_0^{k+1} \overline{\mu}^{k+2}\right\} - \omega_0^k \overline{\mu}_0^{k+1}\right] \\
&= \mathbb{E}_{\overline{P}^{k-1}_{x_m}} \left[\omega_0^k \overline{\mu}^{k+1} + \omega_0^{k+1} \pi^k \mathbb{1}_{x_k}(X_k) \left\{\overline{\mu}^{k+2} - \mu^{k+1}\right\} + \omega_0^{k+1} \pi^k \mathbb{1}_{x_k}(X_k) \left\{\overline{\mu}_0^{k+2} - \overline{\mu}^{k+2}\right\} - \omega_0^k \overline{\mu}_0^{k+1}\right] \\
&+ \mathbb{E}_{\overline{P}^{k-1}_{x_m}} \left[\pi^k \mathbb{1}_{x_k}(X_k) \left\{Q_{k+1} - \omega_0^{k+1} \overline{\mu}_0^{k+2}\right\}\right] \\
&= \mathbb{E}_{\overline{P}^{k-1}_{x_m}} \left[\omega_0^k \overline{\mu}^{k+1} + \omega_0^{k+1} \pi^k \mathbb{1}_{x_k}(X_k) \left\{\overline{\mu}_0^{k+2} - \mu^{k+1}\right\} - \omega_0^k \overline{\mu}_0^{k+1}\right] + \mathbb{E}_{\overline{P}^{k-1}_{x_m}} \left[\pi^k \mathbb{1}_{x_k}(X_k) \left\{Q_{k+1} - \omega_0^{k+1} \overline{\mu}_0^{k+2}\right\}\right] \\
&= \mathbb{E}_{\overline{P}^{k-1}_{x_m}} \left[\omega_0^k \left\{\overline{\mu}^{k+1} - \overline{\mu}_0^{k+1}\right\} + \omega_0^{k+1} \pi^k \mathbb{1}_{x_k}(X_k) \left\{\overline{\mu}_0^{k+2} - \mu^{k+1}\right\}\right] + \mathbb{E}_{\overline{P}^{k-1}_{x_m}} \left[\pi^k \mathbb{1}_{x_k}(X_k) \left\{Q_{k+1} - \omega_0^{k+1} \overline{\mu}_0^{k+2}\right\}\right] \\
&\overset{10}{=} \mathbb{E}_{\overline{P}^{k-1}_{x_m}} \left[\omega_0^k \left\{\overline{\mu}^{k+1} - \overline{\mu}_0^{k+1}\right\} + \pi^k \mathbb{1}_{x_k}(X_k) \left\{\mu_0^{k+1} - \mu^{k+1}\right\}\right] + \mathbb{E}_{\overline{P}^{k-1}_{x_m}} \left[\pi^k \mathbb{1}_{x_k}(X_k) \left\{Q_{k+1} - \omega_0^{k+1} \overline{\mu}_0^{k+2}\right\}\right] \\
&\overset{11}{=} \mathbb{E}_{\overline{P}^{k-1}_{x_m}} \left[\pi_0^k \mathbb{1}_{x_k}(X_k) \left\{\mu^{k+1} - \mu_0^{k+1}\right\} + \pi^k \mathbb{1}_{x_k}(X_k) \left\{\mu_0^{k+1} - \mu^{k+1}\right\}\right] + \mathbb{E}_{\overline{P}^{k-1}_{x_m}} \left[\pi^k \mathbb{1}_{x_k}(X_k) \left\{Q_{k+1} - \omega_0^{k+1} \overline{\mu}_0^{k+2}\right\}\right] \\
&= \mathbb{E}_{\overline{P}^{k-1}_{x_m}} \left[\mathbb{1}_{x_k}(X_k) \left\{\mu^{k+1} - \mu_0^{k+1}\right\} \left\{\pi_0^k - \pi^k\right\}\right] + \mathbb{E}_{\overline{P}^{k-1}_{x_m}} \left[\pi^k \mathbb{1}_{x_k}(X_k) \left\{Q_{k+1} - \omega_0^{k+1} \overline{\mu}_0^{k+2}\right\}\right] \\
&\overset{12}{=} \mathbb{E}_{\overline{P}^{k-1}_{x_m}} \left[\omega_0^{k+1} \mathbb{1}_{x_k}(X_k) \left\{\mu^{k+1} - \mu_0^{k+1}\right\} \left\{\pi_0^k - \pi^k\right\}\right] + \mathbb{E}_{\overline{P}^{k-1}_{x_m}} \left[\pi^k \mathbb{1}_{x_k}(X_k) \left\{Q_{k+1} - \omega_0^{k+1} \overline{\mu}_0^{k+2}\right\}\right] \\
&\overset{13}{=} \mathbb{E}_{\overline{P}^{k-1}_{x_{k+1}}} \left[\mathbb{1}_{x_k}(X_k) \left\{\mu^{k+1} - \mu_0^{k+1}\right\} \left\{\pi_0^k - \pi^k\right\}\right] + \mathbb{E}_{\overline{P}^{k-1}_{x_m}} \left[\pi^k \mathbb{1}_{x_k}(X_k) \left\{Q_{k+1} - \omega_0^{k+1} \overline{\mu}_0^{k+2}\right\}\right]
\end{aligned}$$

$$
\begin{aligned}
&\overset{14}{=} \mathbb{E}_{\overline{P}^{k-1}_{x_{k+1}}}\left[\mathbb{1}_{x_k}(X_k)\left\{\mu^{k+1}-\mu_0^{k+1}\right\}\left\{\pi_0^k-\pi^k\right\}\right]+\sum_{i=k+2}^{m} O_{\overline{P}^k_{x_i}}\left(\left\|\mu^i-\mu_0^i\right\|\left\|\pi^{i-1}-\pi_0^{i-1}\right\|\right)\\
&\overset{15}{=} O_{\overline{P}_{x^{k-1}_{k+1}}}\left(\left\|\mu^{k+1}-\mu_0^{k+1}\right\|\left\|\pi^k-\pi_0^k\right\|\right)+\sum_{i=k+2}^{m} O_{\overline{P}^k_{x_i}}\left(\left\|\mu^i-\mu_0^i\right\|\left\|\pi^{i-1}-\pi_0^{i-1}\right\|\right)\\
&=\sum_{i=k+1}^{m} O_{\overline{P}^{k-1}_{x_i}}\left(\left\|\mu^i-\mu_0^i\right\|\left\|\pi^{i-1}-\pi_0^{i-1}\right\|\right),
\end{aligned}
$$

where

- $\overset{10}{=}$ holds since

$$
\begin{aligned}
&\mathbb{E}_{P_{x_m}}\left[\omega_0^{k+1}(\mathbf{B}^{(k+1)})\pi^k(\mathbf{V}^{(k)})\mathbb{1}_{x_k}(X_k)\mu_0^{k+2}(\mathbf{B}^{(k+1)},x_{k+1})\middle|\mathbf{V}^{(k-1)}\right]\\
&=\mathbb{E}_{P_{x_m}}\left[\mathbb{E}_{P_{x_m}}\left[\omega_0^{k+1}(\mathbf{B}^{(k+1)})\pi^k(\mathbf{V}^{(k)})\mathbb{1}_{x_k}(X_k)\mu_0^{k+2}(\mathbf{B}^{(k+1)},x_{k+1})\middle|\mathbf{V}^{(k)}\right]\middle|\mathbf{V}^{(k-1)}\right]\\
&=\mathbb{E}_{P_{x_m}}\left[\pi^k(\mathbf{V}^{(k)})\mathbb{1}_{x_k}(X_k)\mathbb{E}_{P_{x_m}}\left[\omega_0^{k+1}(\mathbf{B}^{(k+1)})\mu_0^{k+2}(\mathbf{B}^{(k+1)},x_{k+1})\middle|\mathbf{V}^{(k)}\right]\middle|\mathbf{V}^{(k-1)}\right]\\
&=\mathbb{E}_{P_{x_m}}\left[\pi^k(\mathbf{V}^{(k)})\mathbb{1}_{x_k}(X_k)\mathbb{E}_{P_{x_m}}\left[\frac{P_{x_{k+1}}(\mathbf{B}^{(k+1)})}{P_{x_m}(\mathbf{B}^{(k+1)})}\mu_0^{k+2}(\mathbf{B}^{(k+1)},x_{k+1})\middle|\mathbf{V}^{(k)}\right]\middle|\mathbf{V}^{(k-1)}\right]\\
&=\mathbb{E}_{P_{x_m}}\left[\pi^k(\mathbf{V}^{(k)})\mathbb{1}_{x_k}(X_k)\mathbb{E}_{P_{x_{k+1}}}\left[\mu_0^{k+2}(\mathbf{B}^{(k+1)},x_{k+1})\middle|\mathbf{V}^{(k)}\right]\middle|\mathbf{V}^{(k-1)}\right]\\
&=\mathbb{E}_{P_{x_m}}\left[\pi^k(\mathbf{V}^{(k)})\mathbb{1}_{x_k}(X_k)\mu_0^{k+1}(\mathbf{V}^{(k)})\middle|\mathbf{V}^{(k-1)}\right].
\end{aligned}
$$

- $\overset{11}{=}$ holds since

$$
\begin{aligned}
&\mathbb{E}_{P_{x_m}}\left[\omega_0^k(\mathbf{B}^{(k)})\left\{\mu^{k+1}(\mathbf{B}^{(k)},x_k)-\mu_0^{k+1}(\mathbf{B}^{(k)},x_k)\right\}\middle|\mathbf{V}^{(k-1)}\right]\\
&=\mathbb{E}_{P_{x_m}}\left[\omega_0^k(\mathbf{B}^{(k)})\frac{\mathbb{1}_{x_k}(X_k)}{P_{x_m}(X_k|\mathbf{B}^{(k)})}\left\{\mu^{k+1}(\mathbf{B}^{(k)},X_k)-\mu_0^{k+1}(\mathbf{B}^{(k)},X_k)\right\}\middle|\mathbf{V}^{(k-1)}\right]\\
&=\mathbb{E}_{P_{x_m}}\left[\pi_0^k(\mathbf{V}^{(k)})P_{x_m}(X_k|\mathbf{B}^{(k)})\frac{\mathbb{1}_{x_k}(X_k)}{P_{x_m}(X_k|\mathbf{B}^{(k)})}\left\{\mu^{k+1}(\mathbf{B}^{(k)},X_k)-\mu_0^{k+1}(\mathbf{B}^{(k)},X_k)\right\}\middle|\mathbf{V}^{(k-1)}\right]\\
&=\mathbb{E}_{P_{x_m}}\left[\pi_0^k(\mathbf{V}^{(k)})\mathbb{1}_{x_k}(X_k)\left\{\mu^{k+1}(\mathbf{B}^{(k)},X_k)-\mu_0^{k+1}(\mathbf{B}^{(k)},X_k)\right\}\middle|\mathbf{V}^{(k-1)}\right].
\end{aligned}
$$

- $\overset{12}{=}$ and $\overset{13}{=}$ hold since

$$
\begin{aligned}
&\mathbb{E}_{P_{x_m}}\left[\mathbb{1}_{x_k}(X_k)\left\{\mu^{k+1}(\mathbf{V}^{(k)})-\mu_0^{k+1}(\mathbf{V}^{(k)})\right\}\left\{\pi_0^k(\mathbf{V}^{(k)})-\pi^k(\mathbf{V}^{(k)})\right\}\middle|\mathbf{V}^{(k-1)}\right]\\
&=\mathbb{E}_{P_{x_m}}\left[\omega_0^{k+1}(\mathbf{V}^{(k)})\mathbb{1}_{x_k}(X_k)\left\{\mu^{k+1}(\mathbf{V}^{(k)})-\mu_0^{k+1}(\mathbf{V}^{(k)})\right\}\left\{\pi_0^k(\mathbf{V}^{(k)})-\pi^k(\mathbf{V}^{(k)})\right\}\middle|\mathbf{V}^{(k-1)}\right]\\
&=\mathbb{E}_{P_{x_m}}\left[\frac{P_{x_{k+1}}(\mathbf{V}^{(k)})}{P_{x_m}(\mathbf{V}^{(k)})}\mathbb{1}_{x_k}(X_k)\left\{\mu^{k+1}(\mathbf{V}^{(k)})-\mu_0^{k+1}(\mathbf{V}^{(k)})\right\}\left\{\pi_0^k(\mathbf{V}^{(k)})-\pi^k(\mathbf{V}^{(k)})\right\}\middle|\mathbf{V}^{(k-1)}\right]\\
&=\mathbb{E}_{P_{x_{k+1}}}\left[\mathbb{1}_{x_k}(X_k)\left\{\mu^{k+1}(\mathbf{V}^{(k)})-\mu_0^{k+1}(\mathbf{V}^{(k)})\right\}\left\{\pi_0^k(\mathbf{V}^{(k)})-\pi^k(\mathbf{V}^{(k)})\right\}\middle|\mathbf{V}^{(k-1)}\right],
\end{aligned}
$$

  where the second equation hold since

$$
\omega_0^{k+1}(\mathbf{V}^{(k)})=\frac{P_{x_{k+1}}(\mathbf{V}^{(k)})}{P_{x_m}(\mathbf{V}^{(k)})}=1
$$

  since X_{k+1}, X_m are non-descendants of $\mathbf{V}^{(k)}$ so that $P_{x_{k+1}}(\mathbf{V}^{(k)})=P_{x_m}(\mathbf{V}^{(k)})$.

- $\stackrel{14}{=}$ holds by the induction hypothesis.
- $\stackrel{15}{=}$ holds by Cauchy-Schwarz inequality.

Therefore, the induction hypothesis in Eq. (C.12) holds for all $k = 1, 2, \cdots m-1$. Therefore,

$$\text{l.h.s. of Eq. (C.6)} = \mathbb{E}_{P_{x_m}}\left[Q_1 - \omega_0^1 \overline{\mu}_0^2\right] = \sum_{i=2}^{m} O_{P_{x_i}}\left(\left\|\mu^i - \mu_0^i\right\| \left\|\pi^{i-1} - \pi_0^{i-1}\right\|\right),$$

where the second equation holds by plugging $k = 1$ into the verified hypothesis in Eq. (C.12). This completes the proof. $\square$

Lemma C.7 (**Error analysis of the DML estimator for MTI**). *Suppose Assumptions (2,8,9) hold. Let T^{dml} denote the estimator defined in Def. 9. Then,*

$$T^{dml} - \mathbb{E}\left[Y | do(\mathbf{x})\right] = \sum_{i=1}^{m} R_i + \sum_{i=2}^{m} O_{P_{x_i}}\left(\left\|\mu^i - \mu_0^i\right\| \left\|\pi^{i-1} - \pi_0^{i-1}\right\|\right),$$

where R_i for $i = 1, 2, \cdots, m$ are variables converging in mean-zero normal distribution at $n_i^{-1/2}$ rates.

Proof of Lemma C.7. Throughout the proof, we set $\mathbf{A}_i \coloneqq \{C_i, W_i\}$ for all $i = 1, 2, \cdots, m$. We will use $V_i \coloneqq \{A_i, X_i\}$ or all $i = 1, 2, \cdots, m$. We will use $B_i \coloneqq \{A_i, X_{i-1}\}$ or all $i = 1, 2, \cdots, m$, where $X_0 \coloneqq \emptyset$. To simplify the notation, we sometimes simply denote$\mu^i(\mathbf{B}^{(i-1)}, X_{i-1})$ as μ^i; $\mu^i(\mathbf{B}^{(i-1)}, x_{i-1})$ as $\overline{\mu}^i$; and $\pi^i(\mathbf{V}^{(i)})$ as π^i.

Let $T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m})$ be a quantity defined in Eq. (C.5). We first note that

$$T^{dml}(\{\pi_0^k\}_{k=1}^{m-1}, \{\mu_0^k\}_{k=2}^{m}) = \mathbb{E}\left[Y | do(\mathbf{x})\right]$$

by Eq. (C.7). Then, by Lemma C.3,

$$\begin{aligned}
&T^{dml} - \mathbb{E}\left[Y | do(\mathbf{x})\right] \\
&= T^{dml} - T^{dml}(\{\pi_0^k\}_{k=1}^{m-1}, \{\mu_0^k\}_{k=2}^{m}) \\
&= \sum_{i=2}^{m} \mathbb{E}_{P_{x_i} - D_{x_i}}\left[\pi_0^{(i-1)} \mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)}) \left\{\overline{\mu}_0^{i+1} - \mu_0^i\right\}\right] + \mathbb{E}_{P_{x_1} - D_{x_1}}\left[\overline{\mu}_0^2\right] && \text{(C.13)} \\
&+ \sum_{i=2}^{m} \mathbb{E}_{P_{x_i} - D_{x_i}}\left[\pi_0^{(i-1)} \mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)}) \left\{\overline{\mu}_0^{i+1} - \mu_0^i\right\} - \pi^{(i-1)} \mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)}) \left\{\overline{\mu}^{i+1} - \mu^i\right\}\right] + \mathbb{E}_{P_{x_1} - D_{x_1}}\left[\overline{\mu}_0^2 - \overline{\mu}_0^2\right] && \text{(C.14)} \\
&+ \sum_{i=2}^{m} \mathbb{E}_{P_{x_i}}\left[\pi^{(i-1)} \mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)}) \left\{\overline{\mu}^{i+1} - \mu^i\right\}\right] + \mathbb{E}_{P_{x_1}}\left[\overline{\mu}^2 - \overline{\mu}_0^2\right]. && \text{(C.15)}
\end{aligned}$$

We first note that

$$\text{Eq. (C.14)} = \sum_{i=1}^{m} O_{P_{x_i}}(n_i^{-1/2})$$

under Assumptions (2,8) by Lemma C.3.

Then,

$$\text{Eq. (C.13)} + \text{Eq. (C.14)} = \sum_{i=1}^{m} R_i,$$

where R_i for $i = 1, 2, \cdots, m$ are variables converging in mean-zero normal distribution, by the central limit theorem and Slutsky's theorem.

Finally

$$\begin{aligned}\text{Eq. (C.15)} &= T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m}) - \mathbb{E}\left[Y|do(\mathbf{x})\right] \\ &= \sum_{i=k+1}^{m} O_{\overline{P}^{k-1}_{x_i}}\left(\left\|\mu^i - \mu_0^i\right\| \left\|\pi^{i-1} - \pi_0^{i-1}\right\|\right),\end{aligned}$$

where the second equation holds by Lemma C.6. □

Theorem 6 (Error analysis of the estimators for MTI). *Under Assumptions (2,7,8,9) and AC-MTI in Def. 7, the error of the estimators in Def. 9, denoted* $\epsilon^{est} := T^{est} - \mathbb{E}\left[Y|do(\mathbf{x})\right]$ *for* $est \in \{reg, pw, dml\}$*, are:*

$$\begin{aligned}\epsilon^{reg} &= R_1 + O_{P_{x_1}}\left(\|\mu^1 - \mu_0^1\|\right), \\ \epsilon^{pw} &= R_m + O_{P_{x_m}}(\|\pi^{(m-1)} - \pi_0^{(m-1)}\|), \\ \epsilon^{dml} &= \sum_{i=1}^{m} R_i + \sum_{i=2}^{m} O_{P_{x_i}}\left(\|\mu^i - \mu_0^i\|\|\pi^{i-1} - \pi_0^{i-1}\|\right),\end{aligned}$$

where R_i is a random variable such that $\sqrt{n_i}R_i$ converges in distribution to the mean-zero normal random variable, where $n_i := |D_{x_i}|$ for $i \in \{1, \cdots, m\}$.

Proof of Theorem 6. The proof is complete by Lemmas (C.4, C.5, C.7). □

Corollary 6 (Multiply robustness of the DML estimators (Corollary of Thm. 6)**).** *Suppose Asumptions (2,7,8,9) and AC-MTI in Def. 7 hold. For $i = 2, \cdots, m-1$, suppose either $\pi^{i-1} = \pi_0^{i-1}$ or $\mu^i = \mu_0^i$. Then, T^{dml} in Def. 9 is an unbiased estimator of $\mathbb{E}\left[Y|do(\mathbf{x})\right]$.*

Proof of Corollary 6. Let $T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m})$ be a quantity defined in Eq. (C.5). Let

$$T^{dml,i} := \mathbb{E}_{D_i}\left[\pi^{(i-1)}(\mathbf{V}^{(i-1)})\mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)})\left\{\mu^{i+1}(\mathbf{B}^{(i)}, x_i) - \mu^i(\mathbf{V}^{(i)})\right\}\right],\ i = 2, \cdots, m$$

and

$$T^{dml,1} := \mathbb{E}_{D_1}\left[\mu^2(B_1, x_1)\right].$$

Under the assumption that samples are i.i.d.,

$$\sum_{i=1}^{m} \mathbb{E}_{P_{x_i}}\left[T^{dml,i}\right] = T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m}).$$

Then,

$$\begin{aligned}&\sum_{i=1}^{m} \mathbb{E}_{P_{x_i}}\left[T^{dml,i}\right] - \mathbb{E}\left[Y|do(\mathbf{x})\right] \\ &= T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m}) - \mathbb{E}\left[Y|do(\mathbf{x})\right] \\ &= \sum_{i=2}^{m} O_{P_{x_i}}\left(\left\|\mu^i - \mu_0^i\right\| \left\|\pi^{i-1} - \pi_0^{i-1}\right\|\right) \\ &= 0,\end{aligned}$$

where the third equation holds by Lemma C.6, and the last equation holds under the given condition. □

C.8. Proof of Theorem 7

Definition 10 (Adjustment criterion for gMTI (AC-gMTI)). Let $\mathbf{Z} := \{Z_1, \cdots, Z_m\} \subseteq \mathbf{X}$ denote the subset of treatments. Let $\{\ell_i\}_{i=1}^m \subseteq \{1, 2, \cdots, |\mathbf{X}|\}$ denote the index of $\mathbf{Z}$; i.e., $\mathbf{Z} = \{X_{\ell_1}, \cdots, X_{\ell_m}\}$. Let $\overline{X}_1 := \{X_j\}_{j \leq \ell_1}$, $\overline{X}_{m+1} := \{X_j\}_{j > \ell_m}$, and $\overline{X}_i := \{X_j\}_{\ell_{i-1} < j \leq \ell_i}$ for $i = 2, 3, \cdots, m$. An ordered set $\mathbf{A} := \{\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_m\}$ satisfies *adjustment criterion for combining multiple experiments (AC-gMTI)* w.r.t. $\{\mathbf{X}, Y\}$ in G if, for $i = 1, 2, \cdots, m-1$,

1. $(\mathbf{A}_i \perp\!\!\!\perp \overline{\mathbf{X}}^{>i-1} \setminus Z_i | \overline{\mathbf{X}}^{(i-1)}, \mathbf{A}^{(i-1)}, Z_i)_{G_{\overline{\overline{\mathbf{X}}^{>i-1}}}}$;

2. $(Y \perp\!\!\!\perp \overline{X}_i | \mathbf{A}^{(i)}, \overline{\mathbf{X}}^{(i-1)}, \overline{\mathbf{X}}^{>i})_{G_{\underline{\overline{X}_i}, \overline{\overline{\mathbf{X}}^{>i}}}}$; and

3. $(Y \perp\!\!\!\perp \overline{\mathbf{X}}^{\geq m} \setminus Z_m | \mathbf{A}^{(m-1)}, \overline{\mathbf{X}}^{(m-1)}, Z_m)_{G_{\overline{Z_m}, \underline{\overline{\mathbf{X}}^{\geq m} \setminus Z_m}}}$.

Assumption 10 (Positivity Assumption for AC-gMTI). *$P_{z_m}(\overline{X}_m \setminus Z_m, \overline{X}_{m+1} | \mathbf{A}^{(m-1)}, \overline{\mathbf{X}}^{(m-1)})$ and $\{P_{z_i}(\mathbf{A}_i | \mathbf{A}^{(i-1)}, \overline{\mathbf{X}}^{(i-1)}), P_{z_{i+1}}(\mathbf{A}_i | \mathbf{A}^{(i-1)}, \overline{\mathbf{X}}^{(i-1)})\}_{i=1}^{m-1}$, $\{P^{i+1}(\overline{X}_i | \mathbf{A}^{(i)}, \overline{\mathbf{X}}^{(i-1)})\}_{i=1}^{m-1}$ are strictly positive distributions $\forall i \in \{1, \cdots, m\}, \forall z_i \in \mathfrak{D}_{Z_i}$.*

Theorem 7 (Identification through AC-gMTI). *Suppose AC-gMTI in Def. 10 and Assumption 10 hold. Then, $\mathbb{E}[Y | do(\mathbf{x})]$ is identifiable from $\{P_{rand(Z_i)}(\mathbf{A}^{(i)}, \overline{\mathbf{X}}^{(i)})\}_{i=1}^m$ and given as follow:*

$$\mu_0^m := \mathbb{E}_{P_{z_m}}\left[Y | \mathbf{A}^{(m-1)}, \mathbf{X} \setminus Z_m\right]$$
$$\overline{\mu}_0^m := \mathbb{E}_{P_{z_m}}\left[Y | \mathbf{A}^{(m-1)}, \overline{\mathbf{x}}_{m-1:m+1}, \overline{\mathbf{X}}^{(m-2)}\right]$$
$$\mu_0^{m-1} := \mathbb{E}_{P_{z_{m-1}}}\left[\overline{\mu}_0^m \middle| \mathbf{A}^{(m-2)}, \overline{\mathbf{X}}^{(m-2)}\right],$$

where $\overline{X}_{m-1:m+1} := \{\overline{X}_{m-1}, \overline{X}_m, \overline{X}_{m+1}\}$. For $i = m-2, \cdots, 2$,

$$\mu_0^i := \mathbb{E}_{P_{z_i}}\left[\mu^{i+1}(\mathbf{A}^{(i)}, \overline{x}_i, \overline{\mathbf{X}}^{(i-1)}) \middle| \mathbf{A}^{(i-1)}, \overline{\mathbf{X}}^{(i-1)}\right],$$

and $\overline{\mu}_0^{i+1} := \mu_0^{i+1}(\mathbf{A}^{(i)}, \overline{x}_i, \overline{\mathbf{X}}^{(i-1)})$. Then,

$$\mathbb{E}[Y(\mathbf{x})] = \mathbb{E}_{P_{z_1}}\left[\overline{\mu}_0^2\right]. \tag{4}$$

Proof of Theorem 7. We first note that

$$\begin{aligned}
\mathbb{E}\left[Y | do(\mathbf{x} \setminus \overline{\mathbf{x}}^{(m-1)}), \overline{\mathbf{x}}^{(m-1)}, \mathbf{A}^{(m-1)}\right] &= \mathbb{E}\left[Y | do(\overline{x}_m \setminus z_m, z_m, \overline{x}_{m+1}), \overline{\mathbf{x}}^{(m-1)}, \mathbf{A}^{(m-1)}\right] \\
&\stackrel{1}{=} \mathbb{E}\left[Y | do(z_m), \overline{x}_m \setminus z_m, \overline{x}_{m+1}, \overline{\mathbf{x}}^{(m-1)}, \mathbf{A}^{(m-1)}\right] \\
&= \mathbb{E}\left[Y | do(z_m), \mathbf{x} \setminus z_m, \mathbf{A}^{(m-1)}\right] \\
&= \mu_0^m(\mathbf{A}^{(m-1)}, \mathbf{x} \setminus z_m),
\end{aligned}$$

where

- $\stackrel{1}{=}$ holds by the condition $\left(Y \perp\!\!\!\perp \{\overline{X}_m \setminus Z_m, \overline{X}_{m+1}\} | \mathbf{A}^{(m-1)}(\mathbf{x}), \overline{\mathbf{X}}^{(m-1)}, Z_m\right)_{G_{\overline{Z_m}, \underline{\overline{X}_m \setminus Z_m, \overline{X}_{m+1}}}}$ in Def. 10. Specifically, the condition is an application of Rule 2 of do-calculus (Pearl, 2000).

We also note that

$$\begin{aligned}
&\mathbb{E}\left[Y|do(\mathbf{x}\backslash\overline{\mathbf{x}}^{(m-2)}),\overline{\mathbf{x}}^{(m-2)},\mathbf{A}^{(m-2)}\right]\\
&=\mathbb{E}\left[\mathbb{E}\left[Y|do(\mathbf{x}\backslash\overline{\mathbf{x}}^{(m-2)}),\overline{\mathbf{x}}^{(m-2)},\mathbf{A}^{(m-1)}\right]\left|do(\mathbf{x}\backslash\overline{\mathbf{x}}^{(m-2)}),\overline{\mathbf{x}}^{(m-2)},\mathbf{A}^{(m-2)}\right.\right]\\
&\stackrel{2}{=}\mathbb{E}\left[\mathbb{E}\left[Y|do(\mathbf{x}\backslash\overline{\mathbf{x}}^{(m-1)}),\overline{\mathbf{x}}^{(m-1)},\mathbf{A}^{(m-1)}\right]\left|do(\mathbf{x}\backslash\overline{\mathbf{x}}^{(m-2)}),\overline{\mathbf{x}}^{(m-2)},\mathbf{A}^{(m-2)}\right.\right]\\
&=\mathbb{E}\left[\mu_0^m(\mathbf{A}^{(m-1)},\mathbf{x}\backslash z_m)\middle|do(\mathbf{x}\backslash\overline{\mathbf{x}}^{(m-2)}),\overline{\mathbf{x}}^{(m-2)},\mathbf{A}^{(m-2)}\right]\\
&=\mathbb{E}\left[\mu_0^m(\mathbf{A}^{(m-1)},\mathbf{x}\backslash z_m)\middle|do(\overline{\mathbf{x}}^{>m-1}\backslash z_{m-1},z_{m-1}),\overline{\mathbf{x}}^{(m-2)},\mathbf{A}^{(m-2)}\right]\\
&\stackrel{3}{=}\mathbb{E}\left[\mu_0^m(\mathbf{A}^{(m-1)},\mathbf{x}\backslash z_m)\middle|do(z_{m-1}),\overline{\mathbf{x}}^{(m-2)},\mathbf{A}^{(m-2)}\right]\\
&=\mu_0^{m-1}(\mathbf{A}^{(m-2)},\overline{\mathbf{x}}^{(m-2)}),
\end{aligned}$$

where

- $\stackrel{2}{=}$ holds since $(Y\perp\!\!\!\perp\overline{X}_i|\mathbf{A}^{(i)},\overline{\mathbf{X}}^{(i-1)},\overline{\mathbf{X}}^{>i})_{G_{\underline{\overline{X}_i},\overline{\overline{\mathbf{X}}^{>i}}}}$ in Def. 10 and the given positivity condition.
- $\stackrel{3}{=}$ hold since $\left(\mathbf{A}_i\perp\!\!\!\perp\overline{\mathbf{X}}^{>i-1}\backslash Z_i|\overline{\mathbf{X}}^{(i-1)},\mathbf{A}^{(i-1)},Z_i\right)_{G_{\overline{\overline{X}^{\geq i-1}}}}$ in Def. 10, $\overline{\mathbf{X}}^{>i-1}\backslash Z_i$ is non-ancestral to $\overline{\mathbf{X}}^{(i-1)},\mathbf{A}^{(i-1)}$ and the given positivity condition.

Finally, for $i+1\in\{m-2,\cdots,3\}$, suppose

$$\mathbb{E}\left[Y|do(\mathbf{x}\backslash\overline{\mathbf{x}}^{(i)}),\overline{\mathbf{x}}^{(i)},\mathbf{A}^{(i)}\right]=\mu_0^{i+1}(\mathbf{A}^{(i)},\overline{\mathbf{x}}^{(i)}).$$

Then,

$$\begin{aligned}
&\mathbb{E}\left[Y|do(\mathbf{x}\backslash\overline{\mathbf{x}}^{(i-1)}),\overline{\mathbf{x}}^{(i-1)},\mathbf{A}^{(i-1)}\right]\\
&=\mathbb{E}\left[\mathbb{E}\left[Y|do(\mathbf{x}\backslash\overline{\mathbf{x}}^{(i-1)}),\overline{\mathbf{x}}^{(i-1)},\mathbf{A}^{(i)}\right]\left|do(\mathbf{x}\backslash\overline{\mathbf{x}}^{(i-1)}),\overline{\mathbf{x}}^{(i-1)},\mathbf{A}^{(i-1)}\right.\right]\\
&\stackrel{4}{=}\mathbb{E}\left[\mathbb{E}\left[Y|do(\mathbf{x}\backslash\overline{\mathbf{x}}^{(i)}),\overline{\mathbf{x}}^{(i)},\mathbf{A}^{(i)}\right]\left|do(\mathbf{x}\backslash\overline{\mathbf{x}}^{(i-1)}),\overline{\mathbf{x}}^{(i-1)},\mathbf{A}^{(i-1)}\right.\right]\\
&=\mathbb{E}\left[\mu_0^{i+1}(\mathbf{A}^{(i)},\overline{\mathbf{x}}^{(i)})\middle|do(\mathbf{x}\backslash\overline{\mathbf{x}}^{(i-1)}),\overline{\mathbf{x}}^{(i-1)},\mathbf{A}^{(i-1)}\right]\\
&=\mathbb{E}\left[\mu_0^{i+1}(\mathbf{A}^{(i)},\overline{\mathbf{x}}^{(i)})\middle|do(\overline{x}^{\geq i}\backslash z_i,z_i),\overline{\mathbf{x}}^{(i-1)},\mathbf{A}^{(i-1)}\right]\\
&\stackrel{5}{=}\mathbb{E}\left[\mu_0^{i+1}(\mathbf{A}^{(i)},\overline{\mathbf{x}}^{(i)})\middle|do(z_i),\overline{\mathbf{x}}^{(i-1)},\mathbf{A}^{(i-1)}\right]\\
&=\mu_0^{i}(\mathbf{A}^{(i-1)},\overline{\mathbf{x}}^{(i-1)}),
\end{aligned}$$

where

- $\stackrel{4}{=}$ holds since $(Y\perp\!\!\!\perp\overline{X}_i|\mathbf{A}^{(i)},\overline{\mathbf{X}}^{(i-1)},\overline{\mathbf{X}}^{>i})_{G_{\underline{\overline{X}_i},\overline{\overline{\mathbf{X}}^{>i}}}}$ in Def. 10 and the given positivity condition.
- $\stackrel{5}{=}$ holds since $\left(\mathbf{A}_i\perp\!\!\!\perp\overline{\mathbf{X}}^{>i-1}\backslash Z_i|\overline{\mathbf{X}}^{(i-1)},\mathbf{A}^{(i-1)},Z_i\right)_{G_{\overline{\overline{X}^{\geq i-1}}}}$ in Def. 10, $\overline{\mathbf{X}}^{>i-1}\backslash Z_i$ is non-ancestral to $\overline{\mathbf{X}}^{(i-1)},\mathbf{A}^{(i-1)}$ and the given positivity condition.

Therefore, for all $i=m-2,\cdots,2$.

$$\mathbb{E}\left[Y|do(\mathbf{x}\backslash\overline{\mathbf{x}}^{(i)}),\overline{\mathbf{x}}^{(i)},\mathbf{A}^{(i)}\right]=\mu_0^{i+1}(\mathbf{A}^{(i)},\overline{\mathbf{x}}^{(i)}).$$

Finally,

$$\begin{aligned}
\mathbb{E}\left[Y|do(\mathbf{x})\right] &= \mathbb{E}\left[\mathbb{E}\left[Y|do(\mathbf{x}), A_1\right]|do(\mathbf{x})\right] \\
&\stackrel{6}{=} \mathbb{E}\left[\mathbb{E}\left[Y|do(\mathbf{x}\backslash x_1), x_1, A_1\right]|do(\mathbf{x})\right] \\
&= \mathbb{E}\left[\mu_0^2(A_1, \overline{x}_1)|do(\mathbf{x})\right] \\
&= \mathbb{E}\left[\mu_0^2(A_1, \overline{x}_1)|do(\mathbf{x}\backslash z_1, z_1)\right] \\
&\stackrel{7}{=} \mathbb{E}\left[\mu_0^2(A_1, \overline{x}_1)|do(z_1)\right] \\
&= \mathbb{E}_{P_{z_1}}\left[\mu_0^2(A_1, \overline{x}_1)\right],
\end{aligned}$$

where

- $\stackrel{6}{=}$ holds since $(Y \perp\!\!\!\perp \overline{X}_i|\mathbf{A}^{(i)}, \overline{\mathbf{X}}^{(i-1)}, \overline{\mathbf{X}}^{>i})_{G_{\underline{\overline{X}_i}, \overline{\overline{\mathbf{X}}^{>i}}}}$ in Def. 10 and the given positivity condition.
- $\stackrel{7}{=}$ holds since $\left(\mathbf{A}_i \perp\!\!\!\perp \overline{\mathbf{X}}^{>i-1}\backslash Z_i|\overline{\mathbf{X}}^{(i-1)}, \mathbf{A}^{(i-1)}, Z_i\right)_{G_{\overline{\overline{X}^{\geq i-1}}}}$ in Def. 10, $\overline{\mathbf{X}}^{>i-1}\backslash Z_i$ is non-ancestral to $\overline{\mathbf{X}}^{(i-1)}, \mathbf{A}^{(i-1)}$ and the given positivity condition.

□

C.9. Proof of Theorem 8 and Corollary 8

Definition 11 (Nuisances for AC-gMTI). Nuisance functions for AC-gMTI are defined as follows: For a fixed $\mathbf{z} := \{z_1, \cdots, z_m\} \in \mathfrak{D}_{\mathbf{Z}}$, let $\{\mu_0^i\}_{i=2}^m$ be the nuisances defined in Thm. 7. For $i = 1, \cdots, m-2$, $\pi_0^i := \frac{P_{z_i}(A_i|\mathbf{A}^{(i-1)}, \overline{\mathbf{X}}^{(i-1)})}{P_{z_m}(A_i, \overline{X}_i|\mathbf{A}^{(i-1)}, \overline{\mathbf{X}}^{(i-1)})}$, and $\pi_0^{(i)} := \prod_{j=1}^i \pi_0^j(\mathbf{A}^{(j)}, \overline{\mathbf{X}}^{(j)})$. Also, $\pi_0^{m-1} := \frac{P_{z_{m-1}}(A_{m-1}|\mathbf{A}^{(m-2)}, \overline{\mathbf{X}}^{(m-2)})}{P_{z_m}(A_{m-1}, \overline{X}_{m-1:m+1}|\mathbf{A}^{(m-2)}, \overline{\mathbf{X}}^{(m-2)})}$, and $pi_0^{(m-1)} := \pi_0^{(m-2)} \times \pi_0^{m-1}$, where $\overline{X}_{m-1:m+1} := \{\overline{X}_{m-1}, \overline{X}_m, \overline{X}_{m+1}\}$. For all $i = 1, 2, \cdots, m-1$, we will use $\pi^i(\mathbf{W}^{(i)}, \mathbf{C}^{(i)}, \mathbf{X}^{(i)}) > 0$ and μ^i and $\overline{\mu}^i$ to denote arbitrary finite functions.

Definition 12 (AC-gMTI estimators). Let D_i denote samples following $P_{\text{rand}(Z_i)}(\mathbf{V})$ for $i = 1, 2, \cdots, m$. For a fixed $z_i \in \mathfrak{D}_{Z_i}$, let D_{z_i} denote the subsamples of D_i such that $Z_i = z_i$. Let $\mu^{m+1} := Y$. Let $\mathbb{1}_{\mathbf{x}}^{i-1} := \mathbb{1}_{\overline{\mathbf{x}}^{(i-1)}}(\overline{\mathbf{X}}^{(i-1)})$. Then {REG, PW, DML} estimators are:

$$\begin{aligned}
T^{reg} &:= \mathbb{E}_{D_{z_1}}\left[\mu^2(A_1, \overline{x}_1))\right], \\
T^{pw} &:= \mathbb{E}_{D_{z_m}}\left[\pi^{(m-1)}(\mathbf{A}^{(m-1)}, \mathbf{X})\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right], \\
T^{dml} &:= \sum_{i=2}^m \mathbb{E}_{D_{z_i}}\left[\pi^{(i-1)}\mathbb{1}_{\mathbf{x}}^{i-1}\{\overline{\mu}^{i+1} - \mu^i\}\right] + \mathbb{E}_{D_{z_1}}\left[\overline{\mu}^2\right].
\end{aligned}$$

Assumption 11 (L_2 **consistency of nuisances).** *Estimated nuisances $\{\mu^i\}_{i=2}^m$ and $\{\pi^i\}_{i=1}^{m-1}$ are L_2 consistent; Specifically,*

$$\begin{aligned}
\|\mu^{i+1} - \mu_0^{i+1}\|_{P_{z_i}} &= o_{P_{z_i}}(1), \ \forall i \in \{1, 2, \cdots, m-1\} \\
\|\mu^i - \mu_0^i\|_{P_{z_i}} &= o_{P_{z_i}}(1), \ \forall i \in \{2, \cdots, m\} \\
\|\pi^i - \pi^i\|_{P_{z_{i+1}}} &= o_{P_{z_{i+1}}}(1), \ \forall i \in \{1, \cdots, m-1\}.
\end{aligned}$$

Lemma C.8 (Error analysis of the REG estimator for AC-gMTI). *Suppose Assumptions (2,11) hold. Let T^{reg} denote the estimator defined in Def. 12. Then,*

$$T^{reg} - \mathbb{E}\left[Y|do(\mathbf{x})\right] = R_1 + O_{P_{z_1}}(\left\|\mu^2 - \mu_0^2\right\|)$$

Proof of Lemma C.8. We first note that, by Theorem 7,

$$\mathbb{E}_{P_{z_1}}\left[\mu_0^2(A_1,\overline{x}_1)\right] = \mathbb{E}\left[Y|do(\mathbf{x})\right].$$

By Lemma C.3,

$$\begin{aligned}
&T^{reg} - \mathbb{E}\left[Y|do(\mathbf{x})\right] \\
&= T^{reg} - \mathbb{E}_{P_{z_1}}\left[\mu_0^2(A_1,\overline{x}_1)\right] \\
&= \underbrace{\mathbb{E}_{P_{z_1}-D_1}\left[\mu_0^2(A_1,\overline{x}_1)\right] + \mathbb{E}_{P_{z_1}-D_1}\left[\mu_0^2(A_1,\overline{x}_1) - \mu^2(A_1,\overline{x}_1)\right]}_{:=R_1} \\
&+ \mathbb{E}_{P_{z_1}}\left[\mu_0^2(A_1,\overline{x}_1) - \mu^2(A_1,\overline{x}_1)\right] \\
&= R_1 + \mathbb{E}_{P_{z_1}}\left[\mu_0^2(A_1,\overline{x}_1) - \mu^2(A_1,\overline{x}_1)\right] \\
&= R_1 + O_{P_{z_1}}(\|\mu^2 - \mu_0^2\|),
\end{aligned}$$

where R_1 is a variable such that $\sqrt{n_1}R_1$ converges in distribution to the normal random variable. The last equation holds by Cauchy-Schwartz inequality. □

Lemma C.9 (Error analysis of the PW estimator for AC-gMTI). *Suppose Assumptions (2,11) hold. Let T^{pw} denote the estimator defined in Def. 12. Then,*

$$T^{pw} - \mathbb{E}\left[Y|do(\mathbf{x})\right] = R_m + O_{P_{z_m}}(\|\pi^{(m-1)} - \pi_0^{(m-1)}\|)$$

Proof of Lemma C.9. In the proof, we will use $\tilde{X}_i := \overline{X}_i$ for $i = 1, 2, \cdots, m-2$, and $\tilde{X}_{m-1} := \{\overline{X}_{m-1}, \overline{X}_m, \overline{X}_{m+1}\}$. Therefore, $\tilde{X}_i$ partitions $\mathbf{X}$. In the proof, we tentatively assume

$$\mathbb{E}_{P_{z_m}}\left[\pi_0^{(m-1)}(\mathbf{V}^{(m-1)})\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right] = \mathbb{E}\left[Y|do(\mathbf{x})\right]. \tag{C.16}$$

Then, by Lemma C.3,

$$\begin{aligned}
&T^{pw} - \mathbb{E}\left[Y|do(\mathbf{x})\right] \\
&= T^{pw} - \mathbb{E}_{P_{z_m}}\left[\pi_0^{(m-1)}(\mathbf{V}^{(m-1)})\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right] \\
&= \underbrace{\mathbb{E}_{P_{z_m}-D_m}\left[\pi_0^{(m-1)}(\mathbf{V}^{(m-1)})\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right] + \mathbb{E}_{P_{z_m}-D_m}\left[\left\{\pi_0^{(m-1)}(\mathbf{V}^{(m-1)}) - \pi^{(m-1)}(\mathbf{V}^{(m-1)})\right\}\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right]}_{:=R_m} \\
&+ \mathbb{E}_{P_{z_m}}\left[\left\{\pi_0^{(m-1)}(\mathbf{V}^{(m-1)}) - \pi^{(m-1)}(\mathbf{V}^{(m-1)})\right\}\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right] \\
&= R_m + \mathbb{E}_{P_{x_m}}\left[\left\{\pi_0^{(m-1)}(\mathbf{V}^{(m-1)}) - \pi^{(m-1)}(\mathbf{V}^{(m-1)})\right\}\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right] \\
&= R_1 + O_{P_{z_m}}(\left\|\pi^{(m-1)} - \pi_0^{(m-1)}\right\|),
\end{aligned}$$

where R_m is a variable converging in distribution to the normal distribution at $\sqrt{n_m}$-rate. The last equation holds by Cauchy-Schwartz inequality.

We now prove Eq. (C.16). We first show the following: For $i = 2, \cdots, m$,

$$\mathbb{E}\left[Y|do(\mathbf{x})\right] = \mathbb{E}_{P_{z_i}}\left[\prod_{j=1}^{i-1}\frac{P_{z_j}(A_j|\mathbf{A}^{(j-1)}, \tilde{\mathbf{X}}^{(j-1)})}{P_{z_{j+1}}(A_j, \tilde{X}_j|\mathbf{A}^{(j-1)}, \tilde{\mathbf{X}}^{(j-1)})}\mu_0^i(\mathbf{A}^{(i-1)}, \tilde{\mathbf{X}}^{(i-1)})\mathbb{1}_{\tilde{\mathbf{x}}^{(i-1)}}(\tilde{\mathbf{X}}^{(i-1)})\right]. \tag{C.17}$$

It holds for $i = 2$ as follow:

$$\mathbb{E}_{P_{z_2}}\left[\frac{P_{z_1}(A_1)}{P_{z_2}(A_1,\tilde{X}_1)}\mu_0^2(A_1,\tilde{X}_1)\mathbb{1}_{\tilde{x}_1}(\tilde{X}_1)\right] = \mathbb{E}_{P_{z_1}}\left[\mu_0^2(A_1,\tilde{x}_1)\right] = \mathbb{E}\left[Y|do(\mathbf{x})\right],$$

where the last equation holds by Lemma C.8. Now, we make a following induction hypothesis: For some $i \in \{2,\cdots,m-1\}$, suppose

$$\mathbb{E}\left[Y|do(\mathbf{x})\right] \overset{\text{induction hypothesis}}{=} \mathbb{E}_{P_{z_{i-1}}}\left[\prod_{j=1}^{i-2}\frac{P_{z_j}(A_j|\mathbf{A}^{(j-1)},\tilde{\mathbf{X}}^{(j-1)})}{P_{z_{j+1}}(A_j,\tilde{X}_j|\mathbf{A}^{(j-1)},\tilde{\mathbf{X}}^{(j-1)})}\mu_0^{i-1}(\mathbf{A}^{(i-2)},\tilde{\mathbf{X}}^{(i-2)})\mathbb{1}_{\tilde{\mathbf{x}}^{(i-2)}}(\tilde{\mathbf{X}}^{(i-2)})\right].$$

Then,

$$\begin{aligned}
&\mathbb{E}_{P_{z_{i-1}}}\left[\prod_{j=1}^{i-2}\frac{P_{z_j}(A_j|\mathbf{A}^{(j-1)},\tilde{\mathbf{X}}^{(j-1)})}{P_{z_{j+1}}(A_j,\tilde{X}_j|\mathbf{A}^{(j-1)},\tilde{\mathbf{X}}^{(j-1)})}\mu_0^{i-1}(\mathbf{A}^{(i-2)},\tilde{\mathbf{X}}^{(i-2)})\mathbb{1}_{\tilde{\mathbf{x}}^{(i-2)}}(\tilde{\mathbf{X}}^{(i-2)})\right]\\
&= \mathbb{E}_{P_{z_{i-1}}}\left[\prod_{j=1}^{i-2}\frac{P_{z_j}(A_j|\mathbf{A}^{(j-1)},\tilde{\mathbf{X}}^{(j-1)})}{P_{z_{j+1}}(A_j,\tilde{X}_j|\mathbf{A}^{(j-1)},\tilde{\mathbf{X}}^{(j-1)})}\mathbb{E}_{P_{z_{i-1}}}\left[\mu_0^i(\mathbf{A}^{(i-1)},\tilde{x}_{i-1},\tilde{\mathbf{X}}^{(i-2)})|\mathbf{A}^{(i-2)},\tilde{\mathbf{X}}^{(i-2)}\right]\mathbb{1}_{\tilde{\mathbf{x}}^{(i-2)}}(\tilde{\mathbf{X}}^{(i-2)})\right]\\
&\overset{1}{=} \mathbb{E}_{P_{z_{i-1}}}\left[\prod_{j=1}^{i-2}\frac{P_{z_j}(A_j|\mathbf{A}^{(j-1)},\tilde{\mathbf{X}}^{(j-1)})}{P_{z_{j+1}}(A_j,\tilde{X}_j|\mathbf{A}^{(j-1)},\tilde{\mathbf{X}}^{(j-1)})}\mu_0^i(\mathbf{A}^{(i-1)},\tilde{x}_{i-1},\tilde{\mathbf{X}}^{(i-2)})\mathbb{1}_{\tilde{\mathbf{x}}^{(i-2)}}(\tilde{\mathbf{X}}^{(i-2)})\right]\\
&= \mathbb{E}_{P_{z_{i-1}}}\left[\prod_{j=1}^{i-2}\frac{P_{z_j}(A_j|\mathbf{A}^{(j-1)},\tilde{\mathbf{X}}^{(j-1)})}{P_{z_{j+1}}(A_j,\tilde{X}_j|\mathbf{A}^{(j-1)},\tilde{\mathbf{X}}^{(j-1)})}\frac{\mathbb{1}_{\tilde{x}_{i-1}}(\tilde{X}_{i-1})}{P_{z_{i-1}}(\tilde{X}_{i-1}\backslash Z_{i-1}|\mathbf{A}^{(i-1)},\tilde{\mathbf{X}}^{(i-2)})}\mu_0^i(\mathbf{A}^{(i-1)},\tilde{\mathbf{X}}^{(i-1)})\mathbb{1}_{\tilde{\mathbf{x}}^{(i-2)}}(\tilde{\mathbf{X}}^{(i-2)})\right]\\
&= \mathbb{E}_{P_{z_i}}\left[\prod_{j=1}^{i-2}\frac{P_{z_j}(A_j|\mathbf{A}^{(j-1)},\tilde{\mathbf{X}}^{(j-1)})}{P_{z_{j+1}}(A_j,\tilde{X}_j|\mathbf{A}^{(j-1)},\tilde{\mathbf{X}}^{(j-1)})}\frac{\mathbb{1}_{\tilde{\mathbf{x}}^{(i-1)}}(\tilde{\mathbf{X}}^{(i-1)})}{P_{z_{i-1}}(\tilde{X}_{i-1}\backslash Z_{i-1}|\mathbf{A}^{(i-1)},\tilde{\mathbf{X}}^{(i-2)})}\frac{P_{z_{i-1}}(\mathbf{A}^{(i-1)},\tilde{\mathbf{X}}^{(i-1)})}{P_{z_i}(\mathbf{A}^{(i-1)},\tilde{\mathbf{X}}^{(i-1)})}\mu_0^i(\mathbf{A}^{(i-1)},\tilde{\mathbf{X}}^{(i-1)})\right]\\
&= \mathbb{E}_{P_{z_i}}\left[\prod_{j=1}^{i-2}\frac{P_{z_j}(A_j|\mathbf{A}^{(j-1)},\tilde{\mathbf{X}}^{(j-1)})}{P_{z_{j+1}}(A_j,\tilde{X}_j|\mathbf{A}^{(j-1)},\tilde{\mathbf{X}}^{(j-1)})}\frac{P_{z_{i-1}}(A_{i-1}|\mathbf{A}^{(i-2)},\tilde{\mathbf{X}}^{(i-2)})}{P_{z_i}(A_{i-1},\tilde{X}_{i-1}|\mathbf{A}^{(i-2)},\tilde{\mathbf{X}}^{(i-2)})}\mu_0^i(\mathbf{A}^{(i-1)},\tilde{\mathbf{X}}^{(i-1)})\mathbb{1}_{\tilde{\mathbf{x}}^{(i-1)}}(\tilde{\mathbf{X}}^{(i-1)})\right]\\
&= \mathbb{E}_{P_{z_i}}\left[\prod_{j=1}^{i-1}\frac{P_{z_j}(A_j|\mathbf{A}^{(j-1)},\tilde{\mathbf{X}}^{(j-1)})}{P_{z_{j+1}}(A_j,\tilde{X}_j|\mathbf{A}^{(j-1)},\tilde{\mathbf{X}}^{(j-1)})}\mu_0^i(\mathbf{A}^{(i-1)},\tilde{\mathbf{X}}^{(i-1)})\mathbb{1}_{\tilde{\mathbf{x}}^{(i-1)}}(\tilde{\mathbf{X}}^{(i-1)})\right].
\end{aligned}$$

where $\overset{1}{=}$ holds by the law of total expectation. Therefore, Eq. (C.17) holds. By plugging $i = m$, we have

$$\begin{aligned}
\mathbb{E}\left[Y|do(\mathbf{x})\right] &= \mathbb{E}_{P_{z_m}}\left[\prod_{j=1}^{m-1}\frac{P_{z_j}(A_j|\mathbf{A}^{(j-1)},\tilde{\mathbf{X}}^{(j-1)})}{P_{z_m}(A_j,\tilde{X}_j|\mathbf{A}^{(j-1)},\tilde{\mathbf{X}}^{(j-1)})}\mu_0^m(\mathbf{A}^{(m-1)},\tilde{\mathbf{X}}^{(m-1)})\mathbb{1}_{\tilde{\mathbf{x}}^{(m-1)}}(\tilde{\mathbf{X}}^{(m-1)})\right]\\
&= \mathbb{E}_{P_{z_m}}\left[\pi_0^{m-1}(\mathbf{A}^{(m-1)},\tilde{\mathbf{X}}^{(m-1)})\mu_0^m(\mathbf{A}^{(m-1)},\tilde{\mathbf{X}}^{(m-1)})\mathbb{1}_{\mathbf{x}}(\mathbf{X})\right]\\
&= \mathbb{E}_{P_{z_m}}\left[\pi_0^{m-1}\mathbf{A}^{(m-1)},\tilde{\mathbf{X}}^{(m-1)})\mathbb{E}_{z_m}\left[Y|\mathbf{A}^{(m-1)},\tilde{\mathbf{X}}^{(m-1)}\right]\mathbb{1}_{\mathbf{x}}(\mathbf{X})\right]\\
&= \mathbb{E}_{P_{z_m}}\left[\pi_0^{m-1}(\mathbf{A}^{(m-1)},\tilde{\mathbf{X}}^{(m-1)})\mathbb{1}_{\mathbf{x}}(\mathbf{X})Y\right].
\end{aligned}$$

□

Lemma C.10 (Bias Analysis of the DML estimator for AC-gMTI). *Suppose Assumptions (2,8) hold. Let $\mu^{m+1} := Y$. Let $\tilde{X}_i := \overline{X}_i$ for $i = 1, 2, \cdots, m-2$, and $\tilde{X}_{m-1} := \{\overline{X}_{m-1}, \overline{X}_m, \overline{X}_{m+1}\}$. Let $V_i := \{A_i, \tilde{X}_i\}$ for $i = 1, 2, \cdots, m-1$.*

Let $B_i := \{A_i, \tilde{X}_{i-1}\}$ for $i = 1, 2, \cdots, m$ where $\tilde{X}_0 := \emptyset$. Let $T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m})$ be defined as follow:

$$\begin{aligned} &T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m}) \\ &:= \sum_{i=2}^{m} \mathbb{E}_{P_{z_i}}\left[\pi^{(i-1)}(\mathbf{V}^{(i-1)})\mathbb{1}_{\tilde{\mathbf{x}}^{(i-1)}}(\tilde{\mathbf{X}}^{(i-1)})\left\{\mu^{i+1}(\mathbf{B}^{(i)}, \tilde{x}_i) - \mu^i(\mathbf{B}^{(i-1)}, \tilde{X}_{i-1})\right\}\right] + \mathbb{E}_{P_{z_1}}\left[\mu^2(\mathbf{B}^{(1)}, \tilde{x}_1)\right]. \end{aligned} \tag{C.18}$$

Then,

$$T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m}) - \mathbb{E}\left[Y|do(\mathbf{x})\right] = \sum_{i=2}^{m} O_{P_{z_i}}\left(\left\|\mu^i - \mu_0^i\right\|\left\|\pi^{i-1} - \pi_0^{i-1}\right\|\right) \tag{C.19}$$

Proof of Lemma C.10. We follow the proof technique used in (Rotnitzky et al., 2017). To simplify the notation, we sometimes simply denote $\mu^i(\mathbf{B}^{(i-1)}, \tilde{X}_{i-1})$ as μ^i; $\mu^i(\mathbf{B}^{(i-1)}, \tilde{x}_{i-1})$ as $\overline{\mu}^i$; and $\pi^i(\mathbf{V}^{(i)})$ as π^i.

We first note that

$$T^{dml}(\{\pi_0^k\}_{k=1}^{m-1}, \{\mu_0^k\}_{k=2}^{m}) = \mathbb{E}\left[Y|do(\mathbf{x})\right]. \tag{C.20}$$

It's easy to witness Eq. (C.20) because, for $i = 2, 3, \cdots, m$,

$$\begin{aligned} &\mathbb{E}_{P_{z_i}}\left[\pi_0^{(i-1)}(\mathbf{V}^{(i-1)})\mathbb{1}_{\tilde{\mathbf{x}}^{(i-1)}}(\tilde{\mathbf{X}}^{(i-1)})\left\{\mu_0^{i+1}(\mathbf{B}^{(i)}, \tilde{x}_i) - \mu_0^i(\mathbf{B}^{(i-1)}, \tilde{X}_{i-1})\right\}\right] \\ &\stackrel{1}{=} \mathbb{E}_{P_{z_i}}\left[\mathbb{E}_{P_{z_i}}\left[\pi_0^{(i-1)}(\mathbf{V}^{(i-1)})\mathbb{1}_{\tilde{\mathbf{x}}^{(i-1)}}(\tilde{\mathbf{X}}^{(i-1)})\left\{\mu_0^{i+1}(\mathbf{B}^{(i)}, \tilde{x}_i) - \mu_0^i(\mathbf{B}^{(i-1)}, \tilde{X}_{i-1})\right\}\middle|\mathbf{B}^{(i-1)}, \tilde{X}_{i-1}\right]\right] \\ &= \mathbb{E}_{P_{z_i}}\left[\pi_0^{(i-1)}(\mathbf{V}^{(i-1)})\mathbb{1}_{\tilde{\mathbf{x}}^{(i-1)}}(\tilde{\mathbf{X}}^{(i-1)})\left\{\mathbb{E}_{P_{z_i}}\left[\mu_0^{i+1}(\mathbf{B}^{(i)}, \tilde{x}_i)|\mathbf{B}^{(i-1)}, \tilde{X}_{i-1}\right] - \mu_0^i(\mathbf{B}^{(i-1)}, X_{i-1})\right\}\right] \\ &= \mathbb{E}_{P_{z_i}}\left[\pi_0^{(i-1)}(\mathbf{V}^{(i-1)})\mathbb{1}_{\tilde{\mathbf{x}}^{(i-1)}}(\tilde{\mathbf{X}}^{(i-1)})\left\{\mu_0^i(\mathbf{B}^{(i-1)}, \tilde{X}_{i-1}) - \mu_0^i(\mathbf{B}^{(i-1)}, \tilde{X}_{i-1})\right\}\right] \\ &= 0, \end{aligned}$$

where the equation $\stackrel{1}{=}$ holds by the law of total expectation. Therefore,

$$T^{dml}(\{\pi_0^k\}_{k=1}^{m-1}, \{\mu_0^k\}_{k=2}^{m}) = \mathbb{E}_{P_{x_1}}\left[\mu_0^2(\mathbf{B}^{(1)}, \overline{x}_1)\right] = \mathbb{E}\left[Y|do(\mathbf{x})\right],$$

where the second equation holds by Lemma C.8. Therefore, it suffices to prove the following to show Eq. (C.19):

$$T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m}) - T^{dml}(\{\pi_0^k\}_{k=1}^{m-1}, \{\mu_0^k\}_{k=2}^{m}) = \sum_{i=2}^{m} O_{P_{z_i}}\left(\left\|\mu^i - \mu_0^i\right\|\left\|\pi^{i-1} - \pi_0^{i-1}\right\|\right). \tag{C.21}$$

For $i = 1, 2, \cdots, m-1$, we define a quantity

$$\omega_0^i(\mathbf{B}^{(i)}) := \frac{P_{z_i}(\mathbf{B}^{(i)})}{P_{z_m}(\mathbf{B}^{(i)})}.$$

We note that $\omega_0^i(\mathbf{B}^{(i)})$ is related with π as follow:

$$\omega_0^i(\mathbf{B}^{(i)}) = \pi_0^i(\mathbf{V}^{(i)})P_{z_m}(\tilde{X}_i|\mathbf{B}^{(i)}). \tag{C.22}$$

To witness, consider the following:

$$\begin{aligned}\omega_0^i(\mathbf{B}^{(i)}) &= \frac{P_{z_i}(A_i|\mathbf{V}^{(i-1)})P_{z_i}(\mathbf{V}^{(i-1)})}{P_{z_m}(A_i|\mathbf{V}^{(i-1)})P_{z_m}(\mathbf{V}^{(i-1)})} \\ &\stackrel{2}{=} \frac{P_{z_i}(A_i|\mathbf{V}^{(i-1)})}{P_{z_m}(A_i|\mathbf{V}^{(i-1)})} \\ &= \pi_0^i(\mathbf{V}^{(i)})\frac{P_{z_m}(A_i,\tilde{X}_i|\mathbf{V}^{(i-1)})}{P_{z_m}(A_i|\mathbf{V}^{(i-1)})} \\ &= \pi_0^i(\mathbf{V}^{(i)})P_{z_m}(\tilde{X}_i|\mathbf{B}^{(i)}),\end{aligned}$$

where

- $\stackrel{2}{=}$ holds since $\tilde{X}_i$ is non-descendent to $\mathbf{V}^{(i-1)}$, so that $P_{z_i}(\mathbf{V}^{(i-1)}) = P_{z_m}(\mathbf{V}^{(i-1)})$.

To simplify the notation, we sometimes simply denote $\omega_0^i(\mathbf{B}^{(i)})$ as ω_0^i; $\mu^i(\mathbf{B}^{(i-1)}, X_{i-1})$ as μ^i; $\mu^i(\mathbf{B}^{(i-1)}, \tilde{x}_{i-1})$ as $\overline{\mu}^i$; and $\pi^i(\mathbf{V}^{(i)})$ as π^i.

Then, $T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^m)$ in Eq. (C.18) can be rewritten as

$$T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^m) = \sum_{i=2}^m \mathbb{E}_{P_{z_m}}\left[\omega_0^i\pi^{(i-1)}\mathbb{1}_{\tilde{\mathbf{x}}^{(i-1)}}(\tilde{\mathbf{X}}^{(i-1)})\left\{\overline{\mu}^{i+1} - \mu^i\right\} + \omega_0^1\overline{\mu}^2\right], \tag{C.23}$$

where $\overline{\mu}^{m+1} := Y$.

For each $k = 1, 2, \cdots, m$, we define a quantity Q_k as follow:

$$Q_k := Q_k(\{\pi^j\}_{j=k}^{m-1}, \{\mu^j\}_{j=k+1}^m) := \omega_0^k\overline{\mu}^{k+1} + \sum_{i=k+1}^m \omega_0^i\pi^{(k:i-1)}\mathbb{1}_{\tilde{\mathbf{x}}^{(k:i-1)}}(\tilde{\mathbf{X}}^{(k:i-1)})\left\{\overline{\mu}^{i+1} - \mu^i\right\}. \tag{C.24}$$

Note $Q_m = Y$ and $\mathbb{E}_{P_{x_m}}[Q_1] = T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^m)$ defined in Eq. (C.23). We note that

$$\begin{aligned}\mathbb{E}_{P_{x_m}}\left[Q_1 - \omega_0^1\overline{\mu}_0^2\right] &\stackrel{4}{=} T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^m) - \mathbb{E}_{P_{z_1}}\left[\mu_0^2(\mathbf{B}^{(1)}, \tilde{x}_1)\right] \\ &\stackrel{5}{=} T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^m) - \mathbb{E}\left[Y|do(\mathbf{x})\right] \\ &= \text{l.h.s. of Eq. (C.19)},\end{aligned}$$

where

- $\stackrel{4}{=}$ holds since $\mathbb{E}_{P_{z_m}}\left[\omega_0^1(\mathbf{B}^{(1)})\mu^2(\mathbf{B}^{(1)}, \tilde{x}_1)\right] = \mathbb{E}_{P_{z_1}}\left[\mu^1(\mathbf{B}^{(1)}, \tilde{x}_1)\right]$.

- $\stackrel{5}{=}$ holds by Lemma C.8.

Motivating from the fact that $\mathbb{E}_{P_{z_m}}\left[Q_1 - \omega_0^1\overline{\mu}_0^2\right] = $ l.h.s. of Eq. (C.19), we establish a following induction hypothesis. For $\overline{P}_{z_m}^{i-1} := P_{z_m}(\cdot|\mathbf{V}^{(i-1)})$, the induction hypothesis is given as follow:

$$\text{Hypothesis: } \mathbb{E}_{\overline{P}_{z_m}^{k-1}}\left[Q_k - \omega_0^k\overline{\mu}_0^{k+1}\right] = \sum_{i=k+1}^m O_{\overline{P}_{z_i}^{k-1}}\left(\left\|\mu^i - \mu_0^i\right\|\left\|\pi^{i-1} - \pi_0^{i-1}\right\|\right), \text{ for } k \in \{2, \cdots, m-1\} \tag{C.25}$$

We first verify the hypothesis Eq. (C.25) for $k = m-1$.

$$\begin{aligned}
&\mathbb{E}_{\overline{P}^{m-2}_{z_m}}\left[Q_{m-1} - \omega_0^{m-1}\overline{\mu}_0^m\right] \\
&= \mathbb{E}_{\overline{P}^{m-2}_{z_m}}\left[\omega_0^{m-1}\overline{\mu}^m + \pi^{m-1}\mathbb{1}_{\tilde{x}_{m-1}}(\tilde{X}_{m-1})\left\{Y - \mu^m\right\} - \omega_0^{m-1}\overline{\mu}_0^m\right] \\
&\overset{6}{=} \mathbb{E}_{\overline{P}^{m-2}_{z_m}}\left[\omega_0^{m-1}\overline{\mu}^m + \pi^{m-1}\mathbb{1}_{\tilde{x}_{m-1}}(\tilde{X}_{m-1})\left\{\mu_0^m - \mu^m\right\} - \omega_0^{m-1}\overline{\mu}_0^m\right] \\
&= \mathbb{E}_{\overline{P}^{m-2}_{z_m}}\left[\omega_0^{m-1}\left\{\overline{\mu}^m - \overline{\mu}_0^m\right\} + \pi^{m-1}\mathbb{1}_{\tilde{x}_{m-1}}(\tilde{X}_{m-1})\left\{\mu_0^m - \mu^m\right\}\right] \\
&\overset{7}{=} \mathbb{E}_{\overline{P}^{m-2}_{z_m}}\left[\omega_0^{m-1}\frac{\mathbb{1}_{\tilde{x}_{m-1}}(\tilde{X}_{m-1})}{P_{z_m}(\tilde{X}_{m-1}|\mathbf{B}^{(m-1)})}\left\{\mu^m - \mu_0^m\right\} + \pi^{m-1}\mathbb{1}_{\tilde{x}_{m-1}}(\tilde{X}_{m-1})\left\{\mu_0^m - \mu^m\right\}\right] \\
&\overset{8}{=} \mathbb{E}_{\overline{P}^{m-2}_{z_m}}\left[\pi_0^{m-1}\mathbb{1}_{\tilde{x}_{m-1}}(\tilde{X}_{m-1})\left\{\mu^m - \mu_0^m\right\} + \pi^{m-1}\mathbb{1}_{\tilde{x}_{m-1}}(\tilde{X}_{m-1})\left\{\mu_0^m - \mu^m\right\}\right] \\
&= \mathbb{E}_{\overline{P}^{m-2}_{z_m}}\left[\mathbb{1}_{\tilde{x}_{m-1}}(\tilde{X}_{m-1})\left\{\mu^m - \mu_0^m\right\}\left\{\pi_0^{m-1} - \pi^{m-1}\right\}\right] \\
&= \mathbb{E}_{P_{x_m}}\left[\mathbb{1}_{\tilde{x}_{m-1}}(\tilde{X}_{m-1})\left\{\mu^m - \mu_0^m\right\}\left\{\pi_0^{m-1} - \pi^{m-1}\right\}\middle|\mathbf{V}^{(m-2)}\right] \\
&\overset{9}{=} O_{\overline{P}^{m-2}_{z_m}}\left(\|\mu^m - \mu_0^m\|\left\|\pi^{m-1} - \pi_0^{m-1}\right\|\right),
\end{aligned}$$

where

- $\overset{6}{=}$ holds by the total law of expectation.

- $\overset{7}{=}$ holds since

$$\mathbb{E}_{P_{z_m}}\left[\mu^{m-1}(\mathbf{B}^{(m-1)}, \tilde{x}_{m-1})\middle|\mathbf{V}^{(m-2)}\right] = \mathbb{E}_{P_{z_m}}\left[\mu^{m-1}(\mathbf{B}^{(m-1)}, \tilde{X}_{m-1})\frac{\mathbb{1}_{\tilde{x}_{m-1}}(\tilde{X}_{m-1})}{P_{z_m}(\tilde{X}_{m-1}|\mathbf{B}^{(m-1)})}\middle|\mathbf{V}^{(m-2)}\right].$$

- $\overset{8}{=}$ holds by the definition of ω_0^{m-1}.

- $\overset{9}{=}$ holds by applying Cauchy-Schwartz inequality.

Now, we suppose Eq. (C.25) holds for some $k+1 \in \{2, \cdots, m-1\}$. Then, we will show that Eq. (C.25) holds for k. Toward this end, we first rewrite Q_k in Eq. (C.24) in a recursive form. For any $k+1 \in \{2, \cdots, m-1\}$, the following relation can be derived from Eq. (C.24):

$$\pi^k\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{Q_{k+1} - \omega_0^{k+1}\overline{\mu}^{k+2}\right\} = \sum_{i=k+2}^{m}\omega_0^i\pi^{(k:i-1)}\mathbb{1}_{\tilde{\mathbf{x}}^{(k:i-1)}}(\tilde{\mathbf{X}}^{(k:i-1)})\left\{\overline{\mu}^{i+1} - \mu^i\right\}.$$

Therefore, for each $k = 1, 2, \cdots, m-1$,

$$Q_k(\{\pi^j\}_{j=k}^{m-1}, \{\mu^j\}_{j=k+1}^{m}) = \omega_0^k\overline{\mu}^{k+1} + \omega_0^{k+1}\pi^k\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{\overline{\mu}^{k+2} - \mu^{k+1}\right\} + \pi^k\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{Q_{k+1} - \omega_0^{k+1}\overline{\mu}^{k+2}\right\}.$$

Then,

$$
\begin{aligned}
&\mathbb{E}_{\overline{P}_{z_m}^{k-1}}\left[Q_k - \omega_0^k \overline{\mu}_0^{k+1}\right] \\
&= \mathbb{E}_{\overline{P}_{z_m}^{k-1}}\left[\omega_0^k \overline{\mu}^{k+1} + \omega_0^{k+1}\pi^k \mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{\overline{\mu}^{k+2} - \mu^{k+1}\right\} + \pi^k \mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{Q_{k+1} - \omega_0^{k+1}\overline{\mu}^{k+2}\right\} - \omega_0^k \overline{\mu}_0^{k+1}\right] \\
&= \mathbb{E}_{\overline{P}_{z_m}^{k-1}}\left[\omega_0^k \overline{\mu}^{k+1} + \omega_0^{k+1}\pi^k \mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{\overline{\mu}^{k+2} - \mu^{k+1}\right\} + \omega_0^{k+1}\pi^k \mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{\overline{\mu}_0^{k+2} - \overline{\mu}^{k+2}\right\} - \omega_0^k \overline{\mu}_0^{k+1}\right] \\
&+ \mathbb{E}_{\overline{P}_{z_m}^{k-1}}\left[\pi^k \mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{Q_{k+1} - \omega_0^{k+1}\overline{\mu}_0^{k+2}\right\}\right] \\
&= \mathbb{E}_{\overline{P}_{z_m}^{k-1}}\left[\omega_0^k \overline{\mu}^{k+1} + \omega_0^{k+1}\pi^k \mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{\overline{\mu}_0^{k+2} - \mu^{k+1}\right\} - \omega_0^k \overline{\mu}_0^{k+1}\right] + \mathbb{E}_{\overline{P}_{z_m}^{k-1}}\left[\pi^k \mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{Q_{k+1} - \omega_0^{k+1}\overline{\mu}_0^{k+2}\right\}\right] \\
&= \mathbb{E}_{\overline{P}_{z_m}^{k-1}}\left[\omega_0^k \left\{\overline{\mu}^{k+1} - \overline{\mu}_0^{k+1}\right\} + \omega_0^{k+1}\pi^k \mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{\overline{\mu}_0^{k+2} - \mu^{k+1}\right\}\right] + \mathbb{E}_{\overline{P}_{z_m}^{k-1}}\left[\pi^k \mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{Q_{k+1} - \omega_0^{k+1}\overline{\mu}_0^{k+2}\right\}\right] \\
&\overset{10}{=} \mathbb{E}_{\overline{P}_{z_m}^{k-1}}\left[\omega_0^k \left\{\overline{\mu}^{k+1} - \overline{\mu}_0^{k+1}\right\} + \pi^k \mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{\mu_0^{k+1} - \mu^{k+1}\right\}\right] + \mathbb{E}_{\overline{P}_{z_m}^{k-1}}\left[\pi^k \mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{Q_{k+1} - \omega_0^{k+1}\overline{\mu}_0^{k+2}\right\}\right] \\
&\overset{11}{=} \mathbb{E}_{\overline{P}_{z_m}^{k-1}}\left[\pi_0^k \mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{\mu^{k+1} - \mu_0^{k+1}\right\} + \pi^k \mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{\mu_0^{k+1} - \mu^{k+1}\right\}\right] + \mathbb{E}_{\overline{P}_{z_m}^{k-1}}\left[\pi^k \mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{Q_{k+1} - \omega_0^{k+1}\overline{\mu}_0^{k+2}\right\}\right] \\
&= \mathbb{E}_{\overline{P}_{z_m}^{k-1}}\left[\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{\mu^{k+1} - \mu_0^{k+1}\right\}\left\{\pi_0^k - \pi^k\right\}\right] + \mathbb{E}_{\overline{P}_{z_m}^{k-1}}\left[\pi^k \mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{Q_{k+1} - \omega_0^{k+1}\overline{\mu}_0^{k+2}\right\}\right] \\
&\overset{12}{=} \mathbb{E}_{\overline{P}_{z_m}^{k-1}}\left[\omega_0^{k+1}\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{\mu^{k+1} - \mu_0^{k+1}\right\}\left\{\pi_0^k - \pi^k\right\}\right] + \mathbb{E}_{\overline{P}_{z_m}^{k-1}}\left[\pi^k \mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{Q_{k+1} - \omega_0^{k+1}\overline{\mu}_0^{k+2}\right\}\right] \\
&\overset{13}{=} \mathbb{E}_{\overline{P}_{z_{k+1}}^{k-1}}\left[\mathbb{1}_{\tilde{x}_k}(\tilde{x}_k)\left\{\mu^{k+1} - \mu_0^{k+1}\right\}\left\{\pi_0^k - \pi^k\right\}\right] + \mathbb{E}_{\overline{P}_{z_m}^{k-1}}\left[\pi^k \mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{Q_{k+1} - \omega_0^{k+1}\overline{\mu}_0^{k+2}\right\}\right] \\
&\overset{14}{=} \mathbb{E}_{\overline{P}_{z_{k+1}}^{k-1}}\left[\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{\mu^{k+1} - \mu_0^{k+1}\right\}\left\{\pi_0^k - \pi^k\right\}\right] + \sum_{i=k+2}^{m} O_{\overline{P}_{z_i}^k}\left(\left\|\mu^i - \mu_0^i\right\|\left\|\pi^{i-1} - \pi_0^{i-1}\right\|\right) \\
&\overset{15}{=} O_{\overline{P}_{z_{k+1}^{k-1}}}\left(\left\|\mu^{k+1} - \mu_0^{k+1}\right\|\left\|\pi^k - \pi_0^k\right\|\right) + \sum_{i=k+2}^{m} O_{\overline{P}_{z_i}^k}\left(\left\|\mu^i - \mu_0^i\right\|\left\|\pi^{i-1} - \pi_0^{i-1}\right\|\right) \\
&= \sum_{i=k+1}^{m} O_{\overline{P}_{z_i}^{k-1}}\left(\left\|\mu^i - \mu_0^i\right\|\left\|\pi^{i-1} - \pi_0^{i-1}\right\|\right),
\end{aligned}
$$

where

- $\overset{10}{=}$ holds since

$$
\begin{aligned}
&\mathbb{E}_{P_{z_m}}\left[\omega_0^{k+1}(\mathbf{B}^{(k+1)})\pi^k(\mathbf{V}^{(k)})\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\mu_0^{k+2}(\mathbf{B}^{(k+1)}, x_{k+1})\middle|\mathbf{V}^{(k-1)}\right] \\
&= \mathbb{E}_{P_{z_m}}\left[\mathbb{E}_{P_{z_m}}\left[\omega_0^{k+1}(\mathbf{B}^{(k+1)})\pi^k(\mathbf{V}^{(k)})\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\mu_0^{k+2}(\mathbf{B}^{(k+1)}, \tilde{x}_{k+1})\middle|\mathbf{V}^{(k)}\right]\middle|\mathbf{V}^{(k-1)}\right] \\
&= \mathbb{E}_{P_{z_m}}\left[\pi^k(\mathbf{V}^{(k)})\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\mathbb{E}_{P_{z_m}}\left[\omega_0^{k+1}(\mathbf{B}^{(k+1)})\mu_0^{k+2}(\mathbf{B}^{(k+1)}, \tilde{x}_{k+1})\middle|\mathbf{V}^{(k)}\right]\middle|\mathbf{V}^{(k-1)}\right] \\
&= \mathbb{E}_{P_{z_m}}\left[\pi^k(\mathbf{V}^{(k)})\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\mathbb{E}_{P_{z_m}}\left[\frac{P_{z_{k+1}}(\mathbf{B}^{(k+1)})}{P_{z_m}(\mathbf{B}^{(k+1)})}\mu_0^{k+2}(\mathbf{B}^{(k+1)}, \tilde{x}_{k+1})\middle|\mathbf{V}^{(k)}\right]\middle|\mathbf{V}^{(k-1)}\right] \\
&= \mathbb{E}_{P_{z_m}}\left[\pi^k(\mathbf{V}^{(k)})\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\mathbb{E}_{P_{z_{k+1}}}\left[\mu_0^{k+2}(\mathbf{B}^{(k+1)}, \tilde{x}_{k+1})\middle|\mathbf{V}^{(k)}\right]\middle|\mathbf{V}^{(k-1)}\right] \\
&= \mathbb{E}_{P_{z_m}}\left[\pi^k(\mathbf{V}^{(k)})\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\mu_0^{k+1}(\mathbf{V}^{(k)})\middle|\mathbf{V}^{(k-1)}\right].
\end{aligned}
$$

- $\stackrel{11}{=}$ holds since

$$
\begin{aligned}
&\mathbb{E}_{P_{z_m}}\left[\omega_0^k(\mathbf{B}^{(k)})\left\{\mu^{k+1}(\mathbf{B}^{(k)},\tilde{x}_k)-\mu_0^{k+1}(\mathbf{B}^{(k)},\tilde{x}_k)\right\}\middle|\mathbf{V}^{(k-1)}\right]\\
&=\mathbb{E}_{P_{z_m}}\left[\omega_0^k(\mathbf{B}^{(k)})\frac{\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)}{P_{z_m}(\tilde{X}_k|\mathbf{B}^{(k)})}\left\{\mu^{k+1}(\mathbf{B}^{(k)},\tilde{X}_k)-\mu_0^{k+1}(\mathbf{B}^{(k)},\tilde{X}_k)\right\}\middle|\mathbf{V}^{(k-1)}\right]\\
&=\mathbb{E}_{P_{z_m}}\left[\pi_0^k(\mathbf{V}^{(k)})P_{z_m}(\tilde{X}_k|\mathbf{B}^{(k)})\frac{\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)}{P_{z_m}(\tilde{X}_k|\mathbf{B}^{(k)})}\left\{\mu^{k+1}(\mathbf{B}^{(k)},\tilde{X}_k)-\mu_0^{k+1}(\mathbf{B}^{(k)},\tilde{X}_k)\right\}\middle|\mathbf{V}^{(k-1)}\right]\\
&=\mathbb{E}_{P_{z_m}}\left[\pi_0^k(\mathbf{V}^{(k)})\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{\mu^{k+1}(\mathbf{B}^{(k)},\tilde{X}_k)-\mu_0^{k+1}(\mathbf{B}^{(k)},\tilde{X}_k)\right\}\middle|\mathbf{V}^{(k-1)}\right].
\end{aligned}
$$

- $\stackrel{12}{=}$ and $\stackrel{13}{=}$ hold since

$$
\begin{aligned}
&\mathbb{E}_{P_{z_m}}\left[\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{\mu^{k+1}(\mathbf{V}^{(k)})-\mu_0^{k+1}(\mathbf{V}^{(k)})\right\}\left\{\pi_0^k(\mathbf{V}^{(k)})-\pi^k(\mathbf{V}^{(k)})\right\}\middle|\mathbf{V}^{(k-1)}\right]\\
&=\mathbb{E}_{P_{z_m}}\left[\omega_0^{k+1}(\mathbf{V}^{(k)})\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{\mu^{k+1}(\mathbf{V}^{(k)})-\mu_0^{k+1}(\mathbf{V}^{(k)})\right\}\left\{\pi_0^k(\mathbf{V}^{(k)})-\pi^k(\mathbf{V}^{(k)})\right\}\middle|\mathbf{V}^{(k-1)}\right]\\
&=\mathbb{E}_{P_{z_m}}\left[\frac{P_{z_{k+1}}(\mathbf{V}^{(k)})}{P_{z_m}(\mathbf{V}^{(k)})}\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{\mu^{k+1}(\mathbf{V}^{(k)})-\mu_0^{k+1}(\mathbf{V}^{(k)})\right\}\left\{\pi_0^k(\mathbf{V}^{(k)})-\pi^k(\mathbf{V}^{(k)})\right\}\middle|\mathbf{V}^{(k-1)}\right]\\
&=\mathbb{E}_{P_{z_{k+1}}}\left[\mathbb{1}_{\tilde{x}_k}(\tilde{X}_k)\left\{\mu^{k+1}(\mathbf{V}^{(k)})-\mu_0^{k+1}(\mathbf{V}^{(k)})\right\}\left\{\pi_0^k(\mathbf{V}^{(k)})-\pi^k(\mathbf{V}^{(k)})\right\}\middle|\mathbf{V}^{(k-1)}\right],
\end{aligned}
$$

  where the second equation hold since

$$
\omega_0^{k+1}(\mathbf{V}^{(k)})=\frac{P_{z_{k+1}}(\mathbf{V}^{(k)})}{P_{z_m}(\mathbf{V}^{(k)})}=1
$$

  since Z_{k+1},Z_m are non-descendants of $\mathbf{V}^{(k)}$ so that $P_{z_{k+1}}(\mathbf{V}^{(k)})=P_{z_m}(\mathbf{V}^{(k)})$.

- $\stackrel{14}{=}$ holds by the induction hypothesis.

- $\stackrel{15}{=}$ holds by Cauchy-Schwartz inequality.

Therefore, the induction hypothesis in Eq. (C.25) holds for all $k=1,2,\cdots m-1$. Therefore,

$$
\text{l.h.s. of Eq. (C.19)}=\mathbb{E}_{P_{z_m}}\left[Q_1-\omega_0^1\overline{\mu}_0^2\right]=\sum_{i=2}^{m}O_{P_{z_i}}\left(\left\|\mu^i-\mu_0^i\right\|\left\|\pi^{i-1}-\pi_0^{i-1}\right\|\right),
$$

where the second equation holds by plugging $k=1$ into the verified hypothesis in Eq. (C.25). This completes the proof. □

Lemma C.11 (Error analysis of the DML estimator for AC-gMTI). *Suppose Assumptions (2,11) hold. Let T^{dml} denote the estimator defined in Def. 12. Then,*

$$
T^{dml}-\mathbb{E}\left[Y|do(\mathbf{x})\right]=\sum_{i=1}^{m}R_i+\sum_{i=2}^{m}O_{P_{z_i}}\left(\left\|\mu^i-\mu_0^i\right\|\left\|\pi^{i-1}-\pi_0^{i-1}\right\|\right),
$$

where R_i for $i=1,2,\cdots,m$ are variables converging in mean-zero normal distribution at $n_i^{-1/2}$ rates.

Proof of Lemma C.11. In the proof, we will use $\tilde{X}_i:=\overline{X}_i$ for $i=1,2,\cdots,m-2$, and $\tilde{X}_{m-1}:=\{\overline{X}_{m-1},\overline{X}_m,\overline{X}_{m+1}\}$. Therefore, $\tilde{X}_i$ partitions $\mathbf{X}$. We will use $B_i:=\{A_i,\tilde{X}_{i-1}\}$ or all $i=1,2,\cdots,m$, where $X_0:=\emptyset$. To simplify the notation, we sometimes simply denote $\mu^i(\mathbf{B}^{(i-1)},\tilde{X}_{i-1})$ as μ^i; $\mu^i(\mathbf{B}^{(i-1)},\tilde{x}_{i-1})$ as $\overline{\mu}^i$; and $\pi^i(\mathbf{V}^{(i)})$ as π^i.

Let $T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m})$ be a quantity defined in Eq. (C.18). We first note that

$$T^{dml}(\{\pi_0^k\}_{k=1}^{m-1}, \{\mu_0^k\}_{k=2}^{m}) = \mathbb{E}\left[Y|do(\mathbf{x})\right]$$

by Eq. (C.7). Then, by Lemma C.3,

$$\begin{aligned}
&T^{dml} - \mathbb{E}\left[Y|do(\mathbf{x})\right] \\
&= T^{dml} - T^{dml}(\{\pi_0^k\}_{k=1}^{m-1}, \{\mu_0^k\}_{k=2}^{m}) \\
&= \sum_{i=2}^{m} \mathbb{E}_{P_{x_i}-D_i}\left[\pi_0^{(i-1)} \mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)}) \left\{\overline{\mu}_0^{i+1} - \mu_0^i\right\}\right] + \mathbb{E}_{P_{x_1}-D_1}\left[\overline{\mu}_0^2\right] && \text{(C.26)} \\
&+ \sum_{i=2}^{m} \mathbb{E}_{P_{x_i}-D_i}\left[\pi_0^{(i-1)} \mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)}) \left\{\overline{\mu}_0^{i+1} - \mu_0^i\right\} - \pi^{(i-1)} \mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)}) \left\{\overline{\mu}^{i+1} - \mu^i\right\}\right] + \mathbb{E}_{P_{x_1}-D_1}\left[\overline{\mu}_0^2 - \overline{\mu}_0^2\right] && \text{(C.27)} \\
&+ \sum_{i=2}^{m} \mathbb{E}_{P_{x_i}}\left[\pi^{(i-1)} \mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)}) \left\{\overline{\mu}^{i+1} - \mu^i\right\}\right] + \mathbb{E}_{P_{x_1}}\left[\overline{\mu}^2 - \overline{\mu}_0^2\right]. && \text{(C.28)}
\end{aligned}$$

We first note that

$$\text{Eq. (C.14)} = \sum_{i=1}^{m} o_{P_{x_i}}(n_i^{-1/2})$$

under Assumptions (2,8) by Lemma C.3.

Then,

$$\text{Eq. (C.13)} + \text{Eq. (C.14)} = \sum_{i=1}^{m} R_i,$$

where R_i for $i = 1, 2, \cdots, m$ are variables converging in mean-zero normal distribution, by the central limit theorem and Slutsky's theorem.

Finally

$$\begin{aligned}
\text{Eq. (C.15)} &= T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m}) - \mathbb{E}\left[Y|do(\mathbf{x})\right] \\
&= \sum_{i=k+1}^{m} O_{\overline{P}_{x_i}^{k-1}}\left(\left\|\mu^i - \mu_0^i\right\| \left\|\pi^{i-1} - \pi_0^{i-1}\right\|\right),
\end{aligned}$$

where the second equation holds by Lemma C.6. □

Theorem 6 (Error analysis of the estimators for MTI). *Under Assumptions (2,7,8,9) and AC-MTI in Def. 7, the error of the estimators in Def. 9, denoted* $\epsilon^{est} := T^{est} - \mathbb{E}\left[Y|do(\mathbf{x})\right]$ *for* $est \in \{reg, pw, dml\}$*, are:*

$$\begin{aligned}
\epsilon^{reg} &= R_1 + O_{P_{x_1}}\left(\|\mu^1 - \mu_0^1\|\right), \\
\epsilon^{pw} &= R_m + O_{P_{x_m}}(\|\pi^{(m-1)} - \pi_0^{(m-1)}\|), \\
\epsilon^{dml} &= \sum_{i=1}^{m} R_i + \sum_{i=2}^{m} O_{P_{x_i}}\left(\|\mu^i - \mu_0^i\|\|\pi^{i-1} - \pi_0^{i-1}\|\right),
\end{aligned}$$

where R_i *is a random variable such that* $\sqrt{n_i}R_i$ *converges in distribution to the mean-zero normal random variable, where* $n_i := |D_{x_i}|$ *for* $i \in \{1, \cdots, m\}$.

Proof of Theorem 6. The proof is complete by Lemmas (C.4, C.5, C.7). ☐

Corollary 6 (Multiply robustness of the DML estimators (Corollary of Thm. 6)**).** *Suppose Asumptions (2,7,8,9) and AC-MTI in Def. 7 hold. For $i = 2, \cdots, m-1$, suppose either $\pi^{i-1} = \pi_0^{i-1}$ or $\mu^i = \mu_0^i$. Then, T^{dml} in Def. 9 is an unbiased estimator of $\mathbb{E}[Y|do(\mathbf{x})]$.*

Proof of Corollary 6. Let $T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m})$ be a quantity defined in Eq. (C.5). Let

$$T^{dml,i} := \mathbb{E}_{D_i}\left[\pi^{(i-1)}(\mathbf{V}^{(i-1)})\mathbb{1}_{\mathbf{x}^{(i-1)}}(\mathbf{X}^{(i-1)})\left\{\mu^{i+1}(\mathbf{B}^{(i)}, x_i) - \mu^i(\mathbf{V}^{(i)})\right\}\right],\ i = 2, \cdots, m$$

and

$$T^{dml,1} := \mathbb{E}_{D_1}\left[\mu^2(B_1, x_1)\right].$$

Under the assumption that samples are i.i.d.,

$$\sum_{i=1}^{m} \mathbb{E}_{P_{x_i}}\left[T^{dml,i}\right] = T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m}).$$

Then,

$$\begin{aligned}
&\sum_{i=1}^{m} \mathbb{E}_{P_{x_i}}\left[T^{dml,i}\right] - \mathbb{E}[Y|do(\mathbf{x})] \\
&= T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m}) - \mathbb{E}[Y|do(\mathbf{x})] \\
&= \sum_{i=2}^{m} O_{P_{x_i}}\left(\|\mu^i - \mu_0^i\|\, \|\pi^{i-1} - \pi_0^{i-1}\|\right) \\
&= 0,
\end{aligned}$$

where the third equation holds by Lemma C.6, and the last equation holds under the given condition. ☐

Theorem 8 (Error analysis of the AC-gMTI estimators). *Under Assumptions (2,10,11) and AC-gMTI in Def. 10, the error of the estimators in Def. 12, denoted $\epsilon^{est} := T^{est} - \mathbb{E}[Y|do(\mathbf{x})]$ for est $\in \{reg, pw, dml\}$, are:*

$$\begin{aligned}
\epsilon^{reg} &= R_1 + O_{P_{z_1}}(\|\mu^1 - \mu_0^1\|), \\
\epsilon^{pw} &= R_m + O_{P_{z_m}}(\|\pi^{(m-1)} - \pi_0^{(m-1)}\|), \\
\epsilon^{dml} &= \sum_{i=1}^{m} R_i + \sum_{i=2}^{m} O_{P_{z_i}}(\|\mu^i - \mu_0^i\|\|\pi^{i-1} - \pi_0^{i-1}\|),
\end{aligned}$$

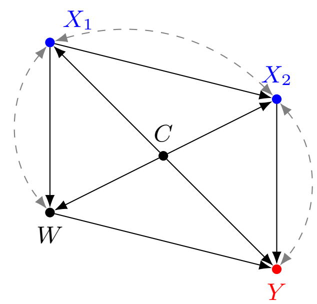
where R_i is a variable such that $\sqrt{n_i}R_i$ converges in distribution to the mean-zero normal random variable, where $n_i := |D_i|$ for $i \in \{1, \cdots, m\}$.

Proof of Theorem 8. The proof is complete by Lemmas (C.8, C.9, C.11). ☐

Corollary 8 (Multiply robustness of the DML estimators (Corollary of Thm. 8)**).** *Suppose Assumptions (2,10,11) and AC-gMTI in Def. 10 hold. For $i = 2, \cdots, m-1$, suppose either $\pi^{i-1} = \pi_0^{i-1}$ or $\mu^i = \mu_0^i$. Then, T^{dml} in Def. 12 is an unbiased estimator of $\mathbb{E}[Y|do(\mathbf{x})]$.*

Proof of Corollary 8. Let $T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m})$ be a quantity defined in Eq. (C.18). Let

$$T^{dml,i} := \mathbb{E}_{D_i}\left[\pi^{(i-1)}(\mathbf{V}^{(i-1)})\mathbb{1}_{\tilde{\mathbf{x}}^{(i-1)}}(\tilde{\mathbf{X}}^{(i-1)})\left\{\mu^{i+1}(\mathbf{B}^{(i)}, \tilde{x}_i) - \mu^i(\mathbf{V}^{(i)})\right\}\right],\ i = 2, \cdots, m$$



(a) Project STAR

Figure D.4: Example causal graphs for Section D. Nodes representing the treatment and outcome are marked in blue and red, respectively.

and

$$T^{dml,1} := \mathbb{E}_{D_1}\left[\mu^2(B_1, \tilde{x}_1)\right].$$

Under the assumption that samples are i.i.d.,

$$\sum_{i=1}^{m} \mathbb{E}_{P_{z_i}}\left[T^{dml,i}\right] = T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m}).$$

Then,

$$\begin{aligned}
&\sum_{i=1}^{m} \mathbb{E}_{P_{z_i}}\left[T^{dml,i}\right] - \mathbb{E}\left[Y|do(\mathbf{x})\right] \\
&= T^{dml}(\{\pi^k\}_{k=1}^{m-1}, \{\mu^k\}_{k=2}^{m}) - \mathbb{E}\left[Y|do(\mathbf{x})\right] \\
&= \sum_{i=2}^{m} O_{P_{z_i}}\left(\left\|\mu^i - \mu_0^i\right\| \left\|\pi^{i-1} - \pi_0^{i-1}\right\|\right) \\
&= 0,
\end{aligned}$$

where the third equation holds by Lemma C.10, and the last equation holds under the given condition. □

D. Project STAR: Estimating Joint Effects of Class Sizes to Academic Outcomes

We applied the proposed estimators to Project STAR dataset (Krueger & Whitmore, 2001; Schanzenbach, 2006). Project STAR is an experimental study investigating teacher/student ratios' impact on academic achievement for kindergarten through third-grade students. In the study, students were randomly assigned to three different class sizes: small-size classes, regular classes, and large-size classes. The objective was to evaluate how class size affects academic outcomes (Schanzenbach, 2006). In our analysis, we used the dataset introduced in the online complement of Stock et al. (2003).

Project STAR Dataset. We denote Project STAR dataset as D. The dataset D includes the following information: class size for kindergarten (X_1), the academic outcome in kindergarten (W), class size for third grade (X_2), the academic outcome in third grade (Y), and pre-treatment variables (C) including genders, age, ethnicity, qualification for free lunch, school types, and teacher's education levels. Since Project STAR is a longitudinal experimental study, the samples for variables $\{C, X_1, W\}$ follow a distribution $P_{\text{rand}(X_1)}(C, X_1, W)$, and the samples for variables $\{C, X_1, W, X_2, Y\}$ follow a distribution $P_{\text{rand}(X_1,X_2)}(C, X_1, W, X_2, Y)$.

Assumption on Dataset. We assume that the structural causal model $\mathcal{M}$ generating the dataset D induces a causal graph depicted in Figure **??**. Specifically, since Project STAR is a longitudinal experimental study randomizing X_1 and X_2, the submodel M_{x_1,x_2} for $x_1, x_2 \in \mathfrak{D}_{X_1,X_2}$ generates the dataset D.

Creation of Datasets from Marginal Experiments. In this empirical study, we create two datasets from this dataset: D_1 and D_2. The dataset D_1 is a random subsample of D only including $\{C, X_1, W\}$. Then, D_1 follows $P_{\text{rand}(X_1)}(C, X_1, W)$.

We now construct the dataset D_2 following the marginal experimental distribution $P_{\text{rand}(X_2)}(C, X_1, W, X_2, Y)$ by introducing the confounding bias between X_1 and W as follows. A specific procedure for introducing confounding bias from experimental studies follows an approach widely used in practice[6], which is described below. Among attributes in C, we chose specific covariates $C_{\text{bias}} \coloneqq \{\text{ethnicity, gender, free-lunch-eligibility}\}$. Next, we assign probabilities for $P_{\text{sample}}(x_1 | c_{\text{bias}})$ for $\forall x_1, c_{\text{bias}} \in \mathfrak{D}_{X_1, C_{\text{bias}}}$. Then, we construct the dataset D_2 as follows: $D_2 \coloneqq \{\}$, and for each samples in $D \coloneqq \{C_{(i)}, X_{1,(i)}, W_{(i)}, X_{2,(i)}, Y_{(i)}\}_{i=1}^{|D|}$, we repeat the following steps:

1. Generate the Bernouli random variable $B_{(i)}$ with parameter $P_{\text{sample}}(X_{1,(i)} | C_{\text{bias}.(i)})$.

2. If $B_{(i)} = 1$, include $\{C_{(i)}, X_{1,(i)}, W_{(i)}, X_{2,(i)}, Y_{(i)}\}$ in D_2.

Finally, we exclude the covariate 'ethnicity' from C in D_1 and D_2. By doing so, we introduce unmeasured confounding bias between X_1 and W in D_2. As a result, D_2 follows a marginal experimental distribution $P_{\text{rand}(X_2)}(C, X_1, W, X_2, Y)$. In this empirical study, the construction of estimators solely relied on the datasets D_1 and D_2, while the dataset D was exclusively leveraged to construct the ground-truth estimate. The following procedure outlines the specific steps on constructing the ground-truth estimate.

Goal. In this empirical study, we aim to study the joint effect of the class size for kindergarten (X_1) and the third grade (X_2) on the third grade's academic outcome (Y); i.e., $\mathbb{E}[Y | do(x_1, x_2)]$. Since D is a longitudinal experimental dataset following $P_{\text{rand}(X_1,X_2)}(C, X_1, W, X_2, Y)$, the ground-truth $\mathbb{E}[Y | do(x_1, x_2)]$ is estimated as $\mathbb{E}_D [Y \mathbb{1}_{x_1,x_2}(X_1, X_2)] / \mathbb{E}_D [\mathbb{1}_{x_1,x_2}(X_1, X_2)]$.

Causal Effect Identification. Identifying and estimating the causal effects $\mathbb{E}[Y | do(x_1, x_2)]$ falls under Task TTI. To witness, we first recall that the datasets D_1 and D_2 consist of samples that follow the distributions $P_{\text{rand}(X_1)}(C, X_1, W)$ and $P_{\text{rand}(X_2)}(C, X_1, W, X_2, Y)$, respectively. Furthermore, within each dataset, the samples D_{x_1} follow the distribution $P_{x_1}(C, W)$ and the samples D_{x_2} follow the distribution $P_{x_2}(C, X_1, W, Y)$.

We first observe that $\{C, W\}$ in the graph G (in Fig. D.4a) satisfies the AC-TTI in Def. 1 w.r.t $\{(X_1, X_2), Y\}$. Specifically,

1. $(\{C, W\} \perp\!\!\!\perp X_2 | X_1)_{G_{\overline{X_1, X_2}}}$; and

2. $(Y \perp\!\!\!\perp X_2 | C, W, X_1)_{G_{\underline{X_1} \overline{X_2}}}$.

Also, the positivity assumption in Assumption 1 is satisfied for D_1 and D_2. Therefore, according to Theorem 1, the joint treatment effects $\mathbb{E}[Y | do(x_1, x_2)]$ are identifiable and can be expressed as follows:

$$\mathbb{E}[Y | do(x_1, x_2)] = \mathbb{E}_{P_{x_1}} \left[\mathbb{E}_{P_{x_2}} [Y | C, W, x_1] \right]. \tag{D.1}$$

[6]The following procedure introduces confounding bias in an experimental dataset by resampling the dataset with a probability depending on the treatment X_1 and covariates C. The procedure has been used in prior research, such as (Hill, 2011; Louizos et al., 2017; Zhang & Bareinboim, 2019; Gentzel et al., 2021) for simulation purposes.

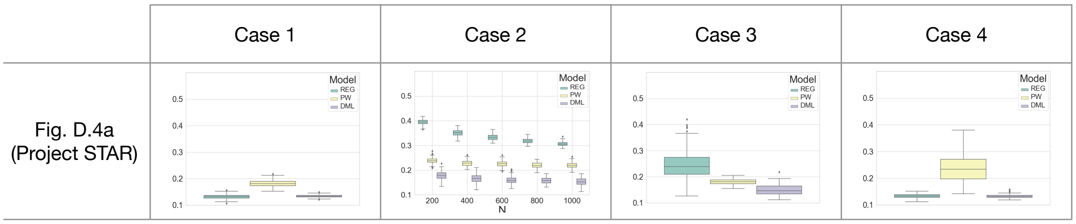


Figure D.5: AAE Plot for Fig. (D.4a) for Cases {1,2,3,4} depicted in the Experimental Setup in Sec. 5.

Causal Effect Estimation. We define the nuisance as follows: For the fixed $x_1, x_2 \in \mathfrak{D}_{X_1,X_2}$,

$$\mu_0(C, X_1, W) \coloneqq \mathbb{E}_{P_{x_2}}\left[Y | W, X_1, C\right], \tag{D.2}$$

$$\pi_0(C, X_1, W) \coloneqq \frac{P_{x_1}(W|C)}{P_{x_2}(W, X_1|C)}. \tag{D.3}$$

Then, besides Eq. (D.1), the causal effect can be expressed as follows:

$$\text{Eq. (D.1)} = \mathbb{E}_{P_{x_2}}\left[Y\pi_0(C, X_1, W)\mathbb{1}_{x_1}(X_1)\right], \text{ or }, \tag{D.4}$$

$$= \mathbb{E}_{P_{x_2}}\left[\pi_0(C, X_1, W)\mathbb{1}_{x_1}(X_1)\{Y - \mu_0(C, X_1, W)\}\right] + \mathbb{E}_{P_{x_1}}\left[\mu_0(C, x_1, W)\right]. \tag{D.5}$$

We then construct the regression-based, probability weighting-based, and double/debiased machine learning (DML) $T^{\text{reg}}, T^{\text{pw}}, T^{\text{dml}}$ using the following procedure.

1. For each fixed $x_i \in \mathfrak{D}_{X_i}$ and a sample set D_{x_i} for $i \in \{1, 2\}$, randomly split the sample as $D_{x_i,t}$ and $D_{x_i,e}$.
2. Use $\{D_{x_1,t}, D_{x_2,t}\}$ to train the model for learning nuisances in Eq. (D.2) and Eq. (D.3). Let $\mu(C, X_1, W)$ and $\pi(C, X_1, W)$ denote the learnt models. We use the XGBoost (Chen & Guestrin, 2016) to learn the model.
3. Then, each estimator is defined as follows:

$$T^{\text{reg}} \coloneqq \mathbb{E}_{D_{x_1,e}}\left[\mu(C, x_1, W)\right] \tag{D.6}$$

$$T^{\text{pw}} \coloneqq \mathbb{E}_{D_{x_2,e}}\left[\pi(C, X_1, W)\mathbb{1}_{x_1}(X_1)Y\right] \tag{D.7}$$

$$T^{\text{dml}} \coloneqq \mathbb{E}_{D_{x_2,e}}\left[\pi(C, X_1, W)\mathbb{1}_{x_1}(X_1)\{Y - \mu(C, X_1, W)\}\right] + \mathbb{E}_{D_{x_1,e}}\left[\mu(C, x_1, W)\right]. \tag{D.8}$$

With the following construction, the Assumption 2 is satisfied.

Experimental Results. As described in the Experimental Setup section (Sec. 5), we evaluated the AAE^{est} of estimators T^{est} for est $\in$ {reg, pw, dml} in Cases $\{1, 2, 3, 4\}$. The AAE plots for all cases can be seen in Fig. D.5. In this particular scenario, the sample size was not varied since the sample itself was externally given.

In Case 2, we introduced variation by adjusting the size of the converging noise ϵ, which follows a normal distribution $\text{Normal}(n^{-\alpha}, n^{-2\alpha})$ for $n \in \{200, 400, 600, 800, 1000\}$. It was observed that the DML estimator T^{dml} outperformed the other two estimators by achieving fast convergence, as demonstrated in Theorem 2. For Cases $\{3, 4\}$, the DML estimator T^{dml} exhibited doubly robust properties, as illustrated in Corollary 2.

E. Details of Experiments

As described in Sec. 5, we used the XGBoost (Chen & Guestrin, 2016) as a model for estimating nuisances $\mu, \pi, \{\mu^i\}_{i=2}^m, \{\pi^i\}_{i=1}^m$. We implemented the model using Python. In modeling nuisance using the XGBoost, we used the command `xgboost.XGBClassifier(eval_metric='logloss')`[7] to use the XGBoost with the default parameter settings. In implementing the PW estimators T^{pw} and the DML estimators T^{dml}, we use the clipped weight by trimming samples yielding weights lower than 10 percentile or greater than 90 percentile (Crump et al., 2009). For Tasks (TTI,MTI), d is chosen to be 10. For Task gTTI, d is chosen to be 5. For Task gMTI, d is chosen to be 1.

We split the dataset as training and test samples with a 5:5 ratio. The training samples are used only for running parameters of the XGBoost models, and the test samples are used only for evaluating the trained XGBoost models.

E.1. Designs of Simulations

This section provides the structural causal models used for generating the dataset. Specifically, we provide a part of the code for generating the dataset.

E.1.1. Task TTI

```
```Generate Exogeneous Variables```
# Generate U_C1_W (Latent confounders between C1, W)
U_C1_W = np.random.normal(0, 1, size=(n,))

# Generate U_C1_X1 (Latent confounders between C1, X1)
U_C1_X1 = np.random.normal(0, 1, size=(n,))

# Generate U_X1_W (Latent confounders between X1, W)
U_X1_W = np.random.normal(0, 1, size=(n,))

# Generate U_X1_X2 (Latent confounders between X1, X2)
U_X1_X2 = np.random.normal(0, 1, size=(n,))

# Generate U_X2_Y (Latent confounders between X2, Y)
U_X2_Y = np.random.normal(0, 1, size=(n,))

# Generate U_C2_X2 (Latent confounders between C2, X2)
U_C2_X2 = np.random.normal(0, 1, size=(n,))

# Generate U_C2_Y (Latent confounders between C2, Y)
U_C2_Y = np.random.normal(0, 1, size=(n,))

```Generate Endogenous Variables```
# SCM for Covariates C1
def f_C1(n,d,U_C1_X1, U_C1_W):
    C1 = np.zeros((n,d))
    for idx in range(0,d):
        C1[:,idx] = np.random.normal(0,1,size = (n,)) + U_C1_X1 + U_C1_W
    return(C1)

# SCM for Treatment X1
def f_X1(n,d,C1, U_C1_X1, U_X1_W, U_X1_X2):
    coeff = np.repeat(1,d)
    X1_linfun = np.dot(C1,coeff) + U_C1_X1 + U_X1_W + U_X1_X2
    X1_param = 1/(1+np.exp(-X1_linfun))
    X1 = np.round(X1_param)
    return(X1)
```

[7] Detailed parametrization of parameters including learning rates, maximum depth of the trees, etc. are explained in https://xgboost.readthedocs.io/en/stable/python/python_api.html#xgboost.XGBClassifier.

```
# SCM for Output W
def f_W(n, d, C1, X1, U_C1_W, U_X1_W):
    coeff1 = np.repeat(1,d)
    coeff2 = np.repeat(-1,d)
    W_linfun = np.dot(C1, coeff1) + np.dot(C1, coeff2) * X1 + U_C1_W + U_X1_W
    W_param = 1 / (1 + np.exp(-W_linfun))
    W = np.round(W_param)
    return (W)

# SCM for Covariates C2
def f_C2(n, d, C1,U_C2_X2, U_C2_Y):
    C2 = np.zeros((n, d))
    for idx in range(0, d):
        C2[:, idx] = (2*C1[:,idx]-1) + U_C2_X2 + U_C2_Y
    return (C2)

# SCM for Treatment X2
def f_X2(n, d, C2, X1, U_C2_X2, U_X2_Y, U_X1_X2):
    coeff1 = np.repeat(1, d)
    coeff2 = np.repeat(-1, d)
    X2_linfun = np.dot(C2, coeff1) +  np.dot(C2, coeff2) *X1 + U_C2_X2 + U_X2_Y + U_X1_X2
    X2_param = 1 / (1 + np.exp(-X2_linfun))
    X2 = np.round(X2_param)
    return (X2)

# SCM for Y
def f_Y(n, d, C2, X2, W, U_C2_Y, U_X2_Y):
    coeff1 = np.repeat(1, d)
    coeff2 = np.repeat(2, d)
    coeff3 = np.repeat(-1, d)

    Y_linfun = np.dot(C2, coeff1) + np.dot(C2, coeff2) *  X2  \
              + np.dot(C2, coeff3) *  W + U_C2_Y + U_X2_Y
    Y_param = 1 / (1 + np.exp(-Y_linfun))
    Y = np.round(Y_param)
    return (Y)
```

E.1.2. TASK gTTI

```
```Generate Exogeneous Variables```
# Generate U_X0_Z1 (Latent confounders between X0, Z1)
U_X0_Z1 = np.random.normal(0, 1, size=(n,))

## Generate U_X0_Z2 (Latent confounders between X0, Z2)
U_X0_Z2 = np.random.normal(0, 1, size=(n,))

## Generate U_Z1_W (Latent confounders between Z1, W)
U_Z1_W = np.random.normal(0, 1, size=(n,))

## Generate U_Z2_Y (Latent confounders between Z2, Y)
U_Z2_Y = np.random.normal(0, 1, size=(n,))
```
```

```
```Generate Endogenous Variables```
# SCM for Treatment C
def f_C(n,d):
    C = np.zeros((n, d))
    for idx in range(0, d):
        C[:, idx] = np.random.normal(0, 1, size=(n,))
    return (C)

# SCM for Treatment X0
def f_X0(n,d, U_X0_Z1 , U_X0_Z2):
    X0_linfun = U_X0_Z1 - U_X0_Z2 + 0.5 + np.random.normal(0, 1, size=(n,))
    X0_param = 1/(1+np.exp(-X0_linfun))
    X0 = np.round(X0_param)
    return(X0)

# SCM for Treatment Z1
def f_Z1(n, d, C, X0, U_X0_Z1, U_Z1_W):
    coeff1 = np.repeat(1, d)
    Z1_linfun = np.dot(C, coeff1)*(2*X0-1) +  U_X0_Z1 + U_Z1_W + X0 \
                + np.random.normal(0, 1, size=(n,))
    Z1_param = 1 / (1 + np.exp(-Z1_linfun))
    Z1 = np.round(Z1_param)
    return (Z1)

# SCM for W
def f_W(n,d,C,Z1,U_Z1_W):
    coeff1 = np.repeat(-0.5, d)
    W_linfun = np.dot(C, coeff1)*(2*Z1-1)  + U_Z1_W + np.random.normal(0, 1, size=(n,))
    W_param = 1 / (1 + np.exp(-W_linfun))
    W = np.round(W_param)
    return (W)

# SCM for Treatment Z2
def f_Z2(n, d, C, X0, Z1, U_X0_Z2, U_Z2_Y):
    coeff1 = np.repeat(-1, d)
    coeff2 = np.repeat(0.5, d)
    U = 0.5*(U_X0_Z2  + U_Z2_Y)
    Z2_linfun = np.dot(C,coeff1)*(2*X0-1) + np.dot(C,coeff2)*(2*Z1-1) + U \
                + np.random.normal(0, 1, size=(n,))
    Z2_param = 1 / (1 + np.exp(-Z2_linfun))
    Z2 = np.round(Z2_param)
    return (Z2)

# SCM for Y
def f_Y(n, d, C, X0, Z2, W, U_Z2_Y):
    coeff1 = np.repeat(-1, d)
    coeff3 = np.repeat(-0.5, d)
    coeff4 = np.repeat(2, d)
    U = 0.5 * U_Z2_Y

    Y_linfun = np.dot(C, coeff1) * (2*X0-1) + np.dot(C, coeff3) * (2*Z2-1) \
              + np.dot(C, coeff4) * (2*W-1) + \
               U  + np.random.normal(0, 1,size=(n,))
```
```

```
    Y_param = 1 / (1 + np.exp(-Y_linfun))
    Y = np.round(Y_param)
    return (Y)
```

E.1.3. TASK MTI

```
''' Generate Exogeneous Variables '''
# Generate U_C1_W1 (Latent confounder between C1, W1)
U_C1_W1 = np.random.normal(0, 1, size=(n,))

## Generate U_C1_X1 (Latent confounder between C1, X1)
U_C1_X1 = np.random.normal(0, 1, size=(n,))

## Generate U_X1_W1 (Latent confounder between X1, W1)
U_X1_W1 = np.random.normal(0, 1, size=(n,))

## Generate U_C2_W2 (Latent confounder between C2, W2)
U_C2_W2 = np.random.normal(0, 1, size=(n,))

## Generate U_C2_X2 (Latent confounder between C2, X2)
U_C2_X2 = np.random.normal(0, 1, size=(n,))

## Generate U_X2_W2 (Latent confounder between X2, W2)
U_X2_W2 = np.random.normal(0, 1, size=(n,))

## Generate U_C3_Y  (Latent confounder between C3, Y)
U_C3_Y = np.random.normal(0, 1, size=(n,))

## Generate U_C3_X3  (Latent confounder between C3, X3)
U_C3_X3 = np.random.normal(0, 1, size=(n,))

## Generate U_X3_Y  (Latent confounder between X3, Y)
U_X3_Y = np.random.normal(0, 1, size=(n,))

''' Generate Endogenous Variables '''
# SCM for Covariates C1
def f_C1(n,d,U_C1_X1,U_C1_W1):
    C1 = np.zeros((n,d))
    for idx in range(0,d):
        C1[:,idx] = np.random.normal(0,1,size = (n,)) + U_C1_X1 + U_C1_W1
    return(C1)

# SCM for Treatment X1
def f_X1(n,d,C1, U_C1_X1, U_X1_W1):
    coeff = np.repeat(1,d)
    X1_linfun = np.dot(C1,coeff) + U_C1_X1 + U_X1_W1
    X1_param = 1/(1+np.exp(-X1_linfun))
    X1 = np.round(X1_param)
    return(X1)

# SCM for Output W1
def f_W1(n, d, C1, X1, U_C1_W1, U_X1_W1):
```

```
    coeff1 = np.repeat(1,d)
    coeff2 = np.repeat(-1,d)
    W1_linfun = np.dot(C1, coeff1) + np.dot(C1, coeff2) * X1 + U_C1_W1 + U_X1_W1
    W1_param = 1 / (1 + np.exp(-W1_linfun))
    W1 = np.round(W1_param)
    return (W1)

# SCM for Covariates C2
def f_C2(n, d, C1,U_C2_X2, U_C2_W2):
    C2 = np.zeros((n, d))
    for idx in range(0, d):
        C2[:, idx] = (2*C1[:,idx]-1) + U_C2_X2 + U_C2_W2
    return (C2)

# SCM for Treatment X2
def f_X2(n, d, C2, U_C2_X2, U_X2_W2):
    coeff1 = np.repeat(1, d)
    X2_linfun = np.dot(C2, coeff1) + U_C2_X2 + U_X2_W2
    X2_param = 1 / (1 + np.exp(-X2_linfun))
    X2 = np.round(X2_param)
    return (X2)

# SCM for Output W2
def f_W2(n, d, C2, X2, W1, U_C2_W2, U_X2_W2):
    coeff1 = np.repeat(1, d)
    coeff2 = np.repeat(2, d)
    coeff3 = np.repeat(-1, d)

    W2_linfun = np.dot(C2, coeff1) + np.dot(C2, coeff2) *  X2 + np.dot(C2, coeff3) *  W1 \
            + U_C2_W2 + U_X2_W2
    W2_param = 1 / (1 + np.exp(-W2_linfun))
    W2 = np.round(W2_param)
    return (W2)

# SCM for Covariates C3
def f_C3(n, d, C2, U_C3_X3, U_C3_Y):
    C3 = np.zeros((n, d))
    for idx in range(0, d):
        C3[:, idx] = (2 * C2[:, idx] - 1) + U_C3_X3 + U_C3_Y
    return (C3)

# SCM for Treatment X3
def f_X3(n, d, C3, U_C3_X3, U_X3_Y):
    coeff1 = np.repeat(1, d)
    X3_linfun = np.dot(C3, coeff1) + U_C3_X3 + U_X3_Y
    X3_param = 1 / (1 + np.exp(-X3_linfun))
    X3 = np.round(X3_param)
    return (X3)

# SCM for Output Y
def f_Y(n, d, C3, X3, W2, U_C3_Y, U_X3_Y):
    coeff1 = np.repeat(1, d)
    coeff2 = np.repeat(2, d)
    coeff3 = np.repeat(-1, d)
```

```
    Y_linfun = np.dot(C3, coeff1) + np.dot(C3, coeff2) * X3 + np.dot(C3, coeff3)  * W2  \
                + U_C3_Y + U_X3_Y
    Y_param = 1 / (1 + np.exp(-Y_linfun))
    Y = np.round(Y_param)
    return (Y)
```

E.1.4. TASK gMTI

```
```Generate Exogeneous Variables```

# Generate U_X0_Z1 (Latent Confounders between X0, Z1)
U_X0_Z1 = np.random.normal(0, 1, size=(n,))

# Generate U_X0_Z2 (Latent Confounders between X0, Z2)
U_X0_Z2 = np.random.normal(0, 1, size=(n,))

# Generate U_X0_Z3 (Latent Confounders between X0, Z3)
U_X0_Z3 = np.random.normal(0, 1, size=(n,))

# Generate U_Z1_W1 (Latent Confounders between Z1, W)
U_Z1_W1 = np.random.normal(0, 1, size=(n,))

# Generate U_Z2_W2 (Latent Confounders between Z2, W2)
U_Z2_W2 = np.random.normal(0, 1, size=(n,))

# Generate U_Z3_Y (Latent Confounders between Z3, Y)
U_Z3_Y = np.random.normal(0, 1, size=(n,))

```Generate Endogenous Variables```
# SCM for Covariate C1
def f_C1(n,d):
    C1 = np.zeros((n, d))
    for idx in range(0, d):
        C1[:, idx] = np.random.normal(0, 1, size=(n,))
    return (C1)

# SCM for Treatment X0
def f_X0(n,d, U_X0_Z1 , U_X0_Z2):
    X0_linfun = U_X0_Z1 - U_X0_Z2 + 0.5 + np.random.normal(0, 1, size=(n,))
    X0_param = 1/(1+np.exp(-X0_linfun))
    X0 = np.round(X0_param)
    return(X0)

# SCM for Treatment Z1
def f_Z1(n, d, C1, X0, U_X0_Z1, U_Z1_W1):
    coeff1 = np.repeat(1, d)
    Z1_linfun = np.dot(C1, coeff1) * (2 * X0 - 1) + U_X0_Z1 + U_Z1_W1  \
               - X0 + np.random.normal(0, 1, size=(n,))
    Z1_param = 1 / (1 + np.exp(-Z1_linfun))
    Z1 = np.round(Z1_param)
    return (Z1)
```

```
# SCM for Outcome W1
def f_W1(n,d,C1,Z1,U_Z1_W1):
    coeff1 = np.repeat(-0.5, d)
    W1_linfun = np.dot(C1, coeff1)*(2*Z1-1)  + U_Z1_W1 + np.random.normal(0, 1, size=(n,))
    W1_param = 1 / (1 + np.exp(-W1_linfun))
    W1 = np.round(W1_param)
    return (W1)

# SCM for Covariate C2
def f_C2(n, d):
    C2 = np.zeros((n, d))
    for idx in range(0, d):
        C2[:,idx] = np.random.normal(0, 1, size=(n,))
    return (C2)


# SCM for Treatment Z2
def f_Z2(n, d, C1, X0, Z1, C2, U_X0_Z2, U_Z2_W2):
    coeff1 = np.repeat(-1, d)
    coeff2 = np.repeat(0.5, d)
    U = 0.5*(U_X0_Z2  + U_Z2_W2)
    Z2_linfun = np.dot(C1,coeff1)*(2*X0-1) + np.dot(C2,coeff2)*(2*Z1-1)  \
              + U + np.random.normal(0, 1, size=(n,))
    Z2_param = 1 / (1 + np.exp(-Z2_linfun))
    Z2 = np.round(Z2_param)
    return (Z2)

# SCM for Outcome W2
def f_W2(n, d, C1, C2, Z2, W1, U_Z2_W2):
    coeff1 = np.repeat(-1, d)
    coeff2 = np.repeat(-0.5, d)
    coeff3 = np.repeat(2, d)
    U = 0.5 * U_Z2_W2

    W2_linfun = np.dot(C1, coeff1) * (2*X0-1) + np.dot(C2, coeff2) * (2*Z2-1) \
               + np.dot(C1 + C2, coeff3) * (2*W1-1) + \
               U + np.random.normal(0, 1,size=(n,))

    W2_param = 1 / (1 + np.exp(-W2_linfun))
    W2 = np.round(W2_param)
    return (W2)

# SCM for Treatment Z3
def f_Z3(n, d, C2, X0, Z2, U_X0_Z3, U_Z3_Y):
    coeff1 = np.repeat(-1, d)
    coeff2 = np.repeat(-0.5, d)
    coeff3 = np.repeat(2, d)
    U = 0.5 * ( U_X0_Z3 + U_Z3_Y )
    Z3_linfun = np.dot(C2, coeff1) + np.dot(C2, coeff2) * (2*X0-1)  \
               + np.dot(C2, coeff3) * (2*Z2-1) +  U*(2*X0-1)*(2*Z2-1) \
            +np.random.normal(0, 1,size=(n,))
    Z3_param = 1 / (1 + np.exp(-Z3_linfun))
    Z3 = np.round(Z3_param)
    return (Z3)
```

```
# SCM for Outcome Y
def f_Y(n, d, C2, X0, Z3, W2, U_Z3_Y):
    coeff1 = np.repeat(-1, d)
    coeff2 = np.repeat(-0.5, d)
    coeff3 = np.repeat(2, d)
    U = 0.5 * U_Z3_Y

    Y_linfun = np.dot(C2, coeff1) * (2 * X0 - 1) + np.dot(C2, coeff2) * (2 * Z3 - 1) \
            + np.dot(C2, coeff3) * (2 * W2 - 1) + U + np.random.normal(0, 1, size=(n,))

    Y_param = 1 / (1 + np.exp(-Y_linfun))
    Y = np.round(Y_param)
    return (Y)
```