Identification and Estimation of "Causes of Effects" using Covariate-Mediator Information

Ryusei Shingaki **Manabu Kuroki**
Graduate School of Engineering Science, Yokohama National University
79-5 Tokiwadai, Hodogaya-ku, Yokohama 240-8501 Japan
shingaki.ryusei@gmail.com, manabu-kuroki-zm@ynu.ac.jp

Abstract

In this paper, we deal with the evaluation problem of "causes of effects" (CoE), which focuses on the likelihood that one event was the cause of another. To assess this likelihood, three types of probabilities of causation have been utilized: probability of necessity, probability of sufficiency, and probability of necessity and sufficiency. However, these usually cannot be estimated, even if "effects of causes" (EoC) is estimable from statistical data, regardless of how large the data is. To solve this problem, we propose novel identification conditions for CoE, using an intermediate variable together with covariate information. Additionally, we also propose a new method for estimating CoE that is applicable whenever they are identifiable through the proposed identification conditions.

1 INTRODUCTION

Statistical causal inference based on structural causal models (Pearl, 2009), which started with path analysis (Wright, 1923, 1934), has made significant progress in recent years, from both theoretical and practical aspects, in elucidating the cause-effect relationships behind statistical data. According to Holland (1986), "one difficulty that arises in talking about causation is the variety of questions that are subsumed under the heading." Such questions include the evaluation problems of "effects of causes" (EoC) and "causes of effects" (CoE). Roughly speaking, the focus of EoC is on predicting what will happen if a cause occurs, whereas that of CoE is on explaining why the effect occurred (Dammann, 2020), and the likelihood that one event was the cause of another (Pearl, 2009). Since experimental studies solve the evaluation problem of EoC through causal risks, we are concerned with the evaluation problem of the CoE from statistical data.

Probabilities of causation (PCs) have often been utilized to evaluate CoE from statistical data. Here, PCs are evaluated from three viewpoints, namely, "necessity causation," "sufficiency causation," and "necessity-and-sufficiency causation," and they have been utilized in various fields such as medical science (Beyea and Greenland, 1999; Greenland, 1987; Robins and Greenland, 1989), risk analysis (Cai and Kuroki, 2005; Cox, 1984), legal reasoning (Dawid et al., 2017), social science (Dawid et al., 2022), statistical science (Kuroki and Cai, 2011; VanderWeele, 2012), and artificial intelligence (Mueller et al., 2022; Pearl, 2009; Tian and Pearl, 2000). Recently, in the field of explainable artificial intelligence (XAI), it has been pointed out that "necessity causation," "sufficiency causation," and "necessity-and-sufficiency causation" are the fundamental components of successful explanation, and that PCs play an important role in evaluating these concepts for probabilistic aspects (Galhotra et al., 2021; Kommiya Mothilal et al., 2021; Watson et al., 2021).

Pearl (2009) and Tian and Pearl (2000) developed formal semantics for PCs based on structural causal models. Additionally, they presented formal definitions of the probability of necessity (PN), the probability of sufficiency (PS), and the probability of necessity and sufficiency (PNS). Furthermore, Lu et al. (2023) defined the posterior total and direct causal effects for the case with multiple causes that may affect each other, which are equivalent to the probability of necessity and probability of causation when there is only a single cause. These PCs are formulated based on the joint probabilities of two potential outcome variables. Since one usually cannot simultaneously observe the

Proceedings of the 27$^{\text{th}}$ International Conference on Artificial Intelligence and Statistics (AISTATS) 2024, Valencia, Spain. PMLR: Volume 238.

results of the same subjects treated and untreated in reality, PCs are not estimable, even for successful experimental studies (Pearl, 2009). To solve this problem, Tian and Pearl (2000) showed how to bound PCs from statistical data obtained in experimental and observational studies. Their bounds, called Tian-Pearl bounds in this paper, are sharp in the sense that the PCs must be within their bounds under a minimal set of assumptions about the data generation process.

Kuroki and Cai (2011) used covariate information to derive bounds for PCs that were narrower than the Tian-Pearl bounds. Recently, Dawid et al. (2022), Dawid and Musio (2022), Mueller et al. (2022), and Murtas et al. (2017) showed that intermediate variables (also called mediators) are also useful for deriving narrower bounds for PCs in experimental studies. However, note that these bounds are still too wide for a practical evaluation of PCs.

Tian and Pearl (2000) also stated that PCs are identifiable if the monotonicity (e.g., the no-defiance) can be assumed and the causal risks are also identifiable, and Pearl (2009) showed that specific functional relationships between cause and effect lead to the identification of PCs. Additionally, referring to the effect restoration method (Kuroki and Pearl, 2014), Shingaki and Kuroki (2021) derived the identification conditions for the joint probabilities of two potential outcome variables, such as PCs, through the observation of two proxy covariates for a single set of potential outcome variables. Furthermore, in the context of natural direct and indirect effects (Pearl, 2001), and under the assumption of no unmeasured confounding, Robins and Richardson (2011) stated that the joint probabilities of two potential outcome variables are identifiable if (i) two potential outcome variables are independent or (ii) one potential outcome variable can be deterministically formulated as a function of another. These existing researches have shown that the joint probabilities of potential outcome variables play an important role in solving various problems related to statistical causal inference. However, although a great deal of effort has been devoted to evaluating EoC over the past three decades, less emphasis has been placed on evaluating CoE, despite their critical importance in practical sciences.

In this paper, we provide novel identification conditions and a statistical estimation method for the joint probabilities of potential outcome variables, using an intermediate variable together with covariate information. The identification conditions are applicable to the identification conditions of PCs. Specifically, they have the following properties:

(1) Shingaki and Kuroki (2021) requires to observe two proxy covariates for a single set of two potential outcome variables that are not associated with each other in order to identify PCs, whereas our method allows for two observed proxy covariates that can be associated with each other.

(2) Shingaki and Kuroki (2021) requires that proxy covariates take four values or more to identify PCs. In contrast, we show that to identify PCs it is enough to observe proxy covariates that take two or three values.

Additionally, note that when PCs are identifiable through our proposed identification conditions, the estimation problem for them is reduced to that of singular models. Thus, they cannot be evaluated by standard statistical likelihood-based estimation methods. In contrast, our proposed identification conditions show that we can derive consistent estimators of the joint probabilities of potential outcome variables via the method of moments, which leads to the asymptotic normality of the proposed estimators through the delta method under regular conditions (Ferguson, 1996). In Supplementary Material B, we also propose a new method for estimating PCs as well as Shingaki and Kuroki (2023). Thus, the results of this paper extend the range of both solvable evaluation problems in statistical causal inference and their application to practical science.

2 PRELIMINARIES

In this section, we introduce the potential outcome variables used to discuss our problems. For the graph-theoretic terminology and the basic theory of structural causal models used in this paper, we refer readers to Pearl (2009). Additionally, we assume that readers are familiar with the basic theory of statistical causal inference (Imbens and Rubin, 2015; Pearl, 2009).

Letting N denote the sample size, we assume that X and Y represent the observed dichotomous treatment variable and the observed dichotomous outcome variable, respectively. For $x \in \{x_0, x_1\}$ (where x_1 is the occurrence of an event [the "experimental treatment"], x_0 is the non-occurrence of an event [the "controlled treatment"]) and $y \in \{y_0, y_1\}$ (where y_1 is the occurrence of an event [the "positive event"], y_0 is the non-occurrence of an event [the "negative event"]), let $p(X = x, Y = y) = p(x, y)$ be the joint probability of $(X, Y) = (x, y)$, $p(Y = y \,|\, X = x) = p(y \,|\, x)$ be the conditional probability of $Y = y$ given $X = x$, and $p(X = x) = p(x)$ be the marginal probability of $X = x$. Similar notation is used for other probabilities. Then, in principle, for $x \in \{x_0, x_1\}$, the i-th of the N subjects has a potential outcome variable $Y_x(i)$

that would have resulted if X had been x for the i-th subject. Here, $Y_x(i) = y$ denotes the counterfactual sentence that "Y would be y had X been x for the i-th subject." The potential outcome variable $Y_x(i)$ is observed only if X is x for the i-th subject, denoted as $X(i) = x$. This property is referred to as consistency (Pearl, 2009; Robins, 1989), and formulated as

$$X(i) = x \implies Y_x(i) = Y \tag{1}$$

for the i-th subject. Similarly, for $x \in \{x_0, x_1\}$ and $s \in \{s_0, s_1\}$, the i-th of the N subjects has a potential outcome variable $Y_{xs}(i)$ that would have resulted if X and S had been x and s for the i-th subject, respectively.

In this paper, we assume the stable unit treatment value assumption (Imbens and Rubin, 2015), which can be summarized as follows: (i) the treatment status of any subject does not affect the outcomes of the other subjects (no interference), and (ii) the treatments of all subjects are comparable (no variations in treatment). Then, when N subjects in the study are considered as a random sample from the population of interest, $Y_x(i)$ is referred to as the value of a random variable, Y_x.

The causal risk of $X = x$ on $Y = y$ is defined as $p(Y_x = y)$. According to Pearl (2009), this can be represented as

$$p(Y_x = y) = \sum_{\boldsymbol{u}} p(y \mid x, \boldsymbol{u}) p(\boldsymbol{u}) \tag{2}$$

based on a set of background variables $\boldsymbol{U}$. Here, summations $(\sum_{\boldsymbol{u}})$ are replaced by integrals $\left(\int_{\boldsymbol{u}}\right)$ whenever the summed variables are continuous. Equation (2) is identifiable and given by $p(Y_x = y) = p(y \mid x)$, if an ideal randomized experiment with X is feasible. Here, "identifiable" means that the causal quantities, such as $p(Y_x = y)$, can be estimated consistently from a joint probability of observed variables. In contrast, when it is difficult to conduct an experimental study, we can still evaluate the causal risks according to the conditionally ignorable treatment assignment condition (Rosenbaum and Rubin, 1983) or, graphically, the back-door criterion (Pearl, 2009). In other words, for a treatment variable X, if there exists a set $\boldsymbol{Z}$ of observed variables such that X is conditionally independent of (Y_{x_0}, Y_{x_1}) given $\boldsymbol{Z}$, then we say that treatment assignment is conditionally ignorable given $\boldsymbol{Z}$. In this case, the causal risks are identifiable and given by

$$p(Y_x = y) = \sum_{\boldsymbol{z}} p(y \mid x, \boldsymbol{z}) p(\boldsymbol{z}). \tag{3}$$

Although there are other identification conditions that can be used to evaluate causal risks (e.g., Pearl, 2009; Tian and Pearl, 2002), we do not cover them in this paper due to space constraints.

Following to Pearl (2009), we define three probabilities of causation (PCs), namely, the probability of necessity (PN), the probability of sufficiency (PS), and the probability of necessity and sufficiency (PNS). PN is defined as

$$\mathrm{PN} = p(Y_{x_0} = y_0 \mid x_1, y_1), \tag{4}$$

which stands for the probability that a negative event would have occurred ($Y = y_0$) in the presence of a controlled treatment ($X = x_0$), given that $X = x_1$ and $Y = y_1$ did in fact occur. PN has applications in epidemiology, legal reasoning, and artificial intelligence. Epidemiologists have long been concerned with estimating the probability that a certain case of a disease is attributable to a particular exposure, which is normally interpreted counterfactually as "the probability that the disease would not have occurred in the absence of exposure, given that disease and exposure did in fact occur." Such a probability is evaluated by PN.

PS is defined as

$$\mathrm{PS} = p(Y_{x_1} = y_1 \mid x_0, y_0), \tag{5}$$

which stands for the probability that a positive event would have occurred ($Y = y_1$) in the presence of an experimental treatment ($X = x_1$), given that $X = x_0$ and $Y = y_0$ did in fact occur. PS has applications in policy analysis, artificial intelligence, and psychology. Policy makers may well be interested in the dangers that a certain exposure may present to the healthy population (Khoury et al., 1989). Counterfactually, this notion is expressed as the "probability that a healthy unexposed subject would have gotten the disease had they been exposed." In artificial intelligence, PS plays a major role in the generation of explanations (Pearl, 2009).

PNS is defined as

$$\mathrm{PNS} = p(Y_{x_0} = y_0, Y_{x_1} = y_1), \tag{6}$$

which stands for the probability that $X = x_1$ is a necessary and sufficient cause for $Y = y_1$. Therefore, it measures both the sufficiency and the necessity of $X = x_1$ to produce $Y = y_1$.

Since PN, PS, and PNS involve joint probabilities of two potential outcome variables, they are not estimable from statistical data even under successful experimental studies without any information (Pearl, 2009).

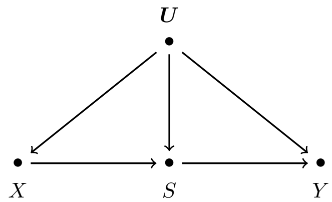


Figure 1: Problem setting.

3 PROBLEM SETTING

Regarding the dichotomous variables X, Y, and S, letting $\boldsymbol{U}$ be the set of all discrete and continuous variables that could affect X, Y and S, both observed and unobserved, we consider the problem of evaluating the joint probabilities of potential outcome variables based on the directed acyclic graph shown in Figure 1. Intuitively, for example, a directed edge from S to Y ($S \rightarrow Y$) indicates that S could have an effect on Y not mediated by other variables in the graph. A variable that is affected by X and has an effect on Y ($X \rightarrow S \rightarrow Y$) is called an intermediate variable, or a mediator. Additionally, the directed path from X to Y through S indicates that X could have an effect on Y mediated by S. The absence of a directed edge from X to Y ($X \rightarrow Y$) indicates that X cannot have an effect on Y without being mediated by other variables in the graph. A directed edge from $\boldsymbol{U}$ to Y ($\boldsymbol{U} \rightarrow Y$) indicates that some elements of $\boldsymbol{U}$ could have an effect on Y. Additionally, the absence of a directed edge from Y to $\boldsymbol{U}$ ($Y \rightarrow \boldsymbol{U}$) indicates that Y cannot be a cause of any element of $\boldsymbol{U}$. Here, $\boldsymbol{U}$ represents the set of all discrete and continuous variables, both observed and unobserved, that are not affected by X or Y. A variable that is not affected by X, such as the elements of $\boldsymbol{U}$, is called a covariate. Figure 1 also graphically represents the data-generating process

$$Y = g_y(S, \boldsymbol{U}, \epsilon_y), \quad S = g_s(X, \boldsymbol{U}, \epsilon_s), \quad X = g_x(\boldsymbol{U}, \epsilon_x), \tag{7}$$

where ϵ_x, ϵ_y, and ϵ_s are independent random disturbances that are also independent of $\boldsymbol{U}$. When structural equation models, such as equation (7), are used to represent the data-generating process, the corresponding graph, such as that shown in Figure 1, is called a causal diagram.

Note that since $\boldsymbol{U}$ can have an effect on both X and S in Figure 1, the causal risks are not identifiable without any additional information (Tian and Pearl, 2002). To solve this problem, we introduce a univariate proxy covariate for a set of two potential outcome variables, such as Z in Figure 3a.

4 IDENTIFICATION

In the situation shown in Figure 2a, the impact of $\boldsymbol{U} \cup \{\epsilon_x, \epsilon_y, \epsilon_s\}$ on Y remains restricted to the modification of the functional relationships between S and Y, irrespective of its complexity. This yields four functions for the two dichotomous variables S and Y; thus, the value taken by $\boldsymbol{U} \cup \{\epsilon_x, \epsilon_y, \epsilon_s\}$ selects one of these four functions (Pearl, 2009). Considering this, when S is formally considered as a treatment variable, the states of $\boldsymbol{U} \cup \{\epsilon_x, \epsilon_y, \epsilon_s\}$ are divided into the following four types:

$u_1 = (Y_{s_0} = y_0, Y_{s_1} = y_0)$ represents the "never-taker" situation, in which the treatment received is irrelevant because the negative event occurs with the experimental or controlled treatment.

$u_2 = (Y_{s_0} = y_0, Y_{s_1} = y_1)$ represents the "complier" situation, in which the negative event occurs if and only if the subjects receive the controlled treatment.

$u_3 = (Y_{s_0} = y_1, Y_{s_1} = y_0)$ represents the "defier" situation, in which the negative event occurs if and only if the subjects receive the experimental treatment.

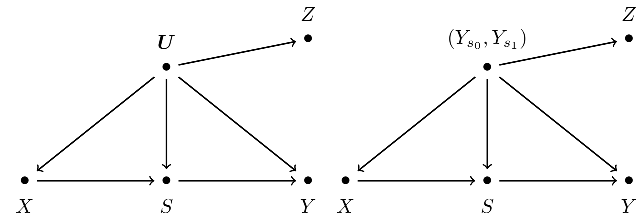
$u_4 = (Y_{s_0} = y_1, Y_{s_1} = y_1)$ represents the "always-taker" situation, in which the treatment received is again irrelevant because the positive event occurs, with the experimental or controlled treatment.

The names of these four groups come from instrumental variable literature in the context of randomized experiments with non-compliance (Angrist et al., 1996; Kawakami et al., 2023). This nomenclature will be repurposed here for the joint probabilities of potential outcome variables.

According to this partition of the states of $\boldsymbol{U} \cup \{\epsilon_x, \epsilon_y, \epsilon_s\}$, we redefine this as U taking a value of u ($u \in \{u_1, u_2, u_3, u_4\}$). Then, the corresponding probabilities that we wish to evaluate are the joint probabilities of potential outcome variables, namely, $p(u_1)$, $p(u_2)$, $p(u_3)$, and $p(u_4)$.

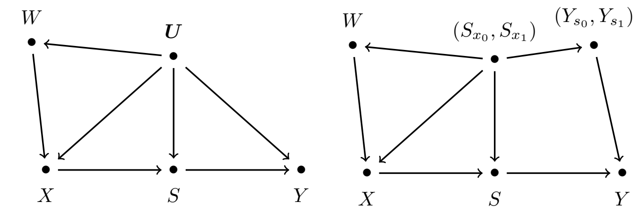
For any x, y, z, and s, we assume that Figure 2a can be redescribed as Figure 2b and that the corresponding recursive factorization of the joint probabilities of X, Y, and Z given S, $p(x, y, z \mid s)$, is given by

$$p(x, y, z \mid s) = \sum_{i=1}^{4} p(y \mid s, u_i) p(x \mid s, u_i) p(z \mid u_i) p(u_i \mid s). \tag{8}$$



(a) Unmeasured confounders (b) Potential outcomes

Figure 2: A graphical representation of Theorem 1.



(a) Unmeasured confounders (b) Potential outcomes

Figure 3: A graphical representation of Theorem 2.

Letting

$$P_{xyzs} = \begin{pmatrix} p(y \mid s) & p(x, y \mid s) \\ p(y, z \mid s) & p(x, y, z \mid s) \end{pmatrix}, \tag{9}$$

regarding the joint probabilities of potential outcome variables $p(x, y, z, s, u)$, we derive the following theorem:

Theorem 1. *Letting Z be a variable taking three values, say, $z \in \{z_1, z_2, z_3\}$, the joint probabilities of potential outcome variables $p(x, y, z, s, u)$ are identifiable for $x \in \{x_0, x_1\}$, $y \in \{y_0, y_1\}$, $s \in \{s_0, s_1\}$, $z \in \{z_1, z_2, z_3\}$, and $u \in \{u_1, u_2, u_3, u_4\}$ if the following conditions are satisfied:*

Condition 1. *The probabilities $p(x, y, z \mid s)$ are available for $x \in \{x_0, x_1\}$, $y \in \{y_0, y_1\}$, $s \in \{s_0, s_1\}$, and $z \in \{z_1, z_2, z_3\}$.*

Condition 2. *For positive probabilities $p(x, y, z, s, u)$, $p(z \mid x, s, u) = p(z \mid u)$ holds for $x \in \{x_0, x_1\}$, $s \in \{s_0, s_1\}$, and $z \in \{z_1, z_2, z_3\}$.*

Condition 3. *For $x \in \{x_0, x_1\}$, $s \in \{s_0, s_1\}$, and $z \in \{z_1, z_2, z_3\}$, the 2×2 matrices P_{xyzs} are invertible, and*

$$\begin{aligned} \frac{\det(P_{x_0 y_0 s_1 z})}{\det(P_{x_0 y_0 s_1 z'})} &\neq \frac{\det(P_{x_0 y_0 s_0 z})}{\det(P_{x_0 y_0 s_0 z'})}, \\ \frac{\det(P_{x_0 y_0 s_0 z})}{\det(P_{x_0 y_0 s_0 z'})} &\neq \frac{\det(P_{x_1 y_1 s_0 z})}{\det(P_{x_1 y_1 s_0 z'})}, \\ \frac{\det(P_{x_0 y_0 s_1 z})}{\det(P_{x_0 y_0 s_1 z'})} &\neq \frac{\det(P_{x_0 y_1 s_0 z})}{\det(P_{x_0 y_1 s_0 z'})}, \\ \frac{\det(P_{x_0 y_1 s_1 z})}{\det(P_{x_0 y_1 s_1 z'})} &\neq \frac{\det(P_{x_0 y_1 s_0 z})}{\det(P_{x_0 y_1 s_0 z'})} \end{aligned} \tag{10}$$

hold for $z \neq z' (z, z' \in \{z_1, z_2, z_3\})$, where $\det(\cdot)$ denotes the determinant.

A proof of Theorem 1 is given in the Supplementary Material A.1. Theorem 1 extends the result of Shingaki and Kuroki (2021) from two proxy covariates taking four values or more to two proxy covariates taking two values or more.

Next, similar to Theorem 1, when S is formally considered as an outcome variable, according to the functional relationships between two dichotomous variables X and S, the states of $\boldsymbol{U} \cup \{\epsilon_s, \epsilon_x\}$ are divided into the following four types:

$v_1 = (S_{x_0} = s_0, S_{x_1} = s_0)$ represents the "never-taker" situation, in which the treatment received is irrelevant because the negative event occurs with the experimental or controlled treatment.

$v_2 = (S_{x_0} = s_0, S_{x_1} = s_1)$ represents the "complier" situation, in which the negative event occurs if and only if the subjects receive the controlled treatment.

$v_3 = (S_{x_0} = s_1, S_{x_1} = s_0)$ represents the "defier" situation, in which the negative event occurs if and only if the subjects receive the experimental treatment.

$v_4 = (S_{x_0} = s_1, S_{x_1} = s_1)$ represents the "always-taker" situation, in which the treatment received is again irrelevant because the positive event occurs, with the experimental or controlled treatment.

For x, w, u, and s, we assume that Figure 3a can be redescribed as Figure 3b and that the corresponding recursive factorization of the joint probabilities of W, S, and U given X, $p(w, s, u \mid x)$, is given by

$$p(w, s, u \mid x) = \sum_{i=1}^{4} p(s \mid x, v_i) p(w \mid x, v_i) p(u \mid v_i) p(v_i \mid x). \tag{11}$$

Letting

$$P_{wsux} = \begin{pmatrix} p(s \mid x) & p(w, s \mid x) \\ p(s, u \mid x) & p(w, s, u \mid x) \end{pmatrix}, \tag{12}$$

regarding the joint probabilities of potential outcome variables $p(x, w, s, u, v)$, we derive the following theorem:

Theorem 2. *Letting W be a variable taking two values, say, $w \in \{w_1, w_2\}$, the joint probabilities of potential outcome variables $p(x, w, s, u, v)$ are identifiable for $x \in \{x_0, x_1\}$, $w \in \{w_0, w_1\}$, $s \in \{s_0, s_1\}$,*

$u \in \{u_1, u_2, u_3, u_4\}$, and $v \in \{v_1, v_2, v_3, v_4\}$ if the following conditions are satisfied:

Condition 4. *The probabilities $p(x, s, w, u)$ are available for $x \in \{x_0, x_1\}$, $s \in \{s_0, s_1\}$, $w \in \{w_0, w_1\}$, and $u \in \{u_1, u_2, u_3, u_4\}$.*

Condition 5. *For positive probabilities $p(x, s, w, u, v)$, $p(u \mid x, w, v) = p(u \mid v)$ holds for $x \in \{x_0, x_1\}$, $w \in \{w_0, w_1\}$, and $u \in \{u_1, u_2, u_3, u_4\}$, and $v \in \{v_1, v_2, v_3, v_4\}$.*

Condition 6. *For $x \in \{x_0, x_1\}$, $s \in \{s_0, s_1\}$, and $u \in \{u_1, u_2, u_3, u_4\}$, the 2×2 matrices P_{wsux} are invertible, and*

$$\begin{aligned} \frac{\det(P_{w_0 s_0 u x_1})}{\det(P_{w_0 s_0 u' x_1})} &\neq \frac{\det(P_{w_0 s_0 u x_0})}{\det(P_{w_0 s_0 u' x_0})}, \\ \frac{\det(P_{w_0 s_0 u x_0})}{\det(P_{w_0 s_0 u' x_0})} &\neq \frac{\det(P_{w_1 s_1 u x_0})}{\det(P_{w_1 s_1 u' x_0})}, \\ \frac{\det(P_{w_0 s_0 u x_1})}{\det(P_{w_0 s_0 u' x_1})} &\neq \frac{\det(P_{w_0 s_1 u x_0})}{\det(P_{w_0 s_1 u' x_0})}, \\ \frac{\det(P_{w_0 s_1 u x_1})}{\det(P_{w_0 s_1 u' x_1})} &\neq \frac{\det(P_{w_0 s_1 u x_0})}{\det(P_{w_0 s_1 u' x_0})} \end{aligned} \tag{13}$$

hold for $u \neq u'$ $(u, u' \in \{u_1, u_2, u_3, u_4\})$.

A proof of Theorem 2 is given in the Supplementary Material A.2.

Theorems 1 and 2 enable us to identify PCs, as shown in Algorithm 1. To emphasize the difference between our results and those in Shingaki and Kuroki (2021), consider the situations shown in Figure 4. Shingaki and Kuroki (2021) requires that two proxy covariates Z and W of (Y_{x_0}, Y_{x_1}) are not associated with each other and W is not associated with Y to identify PCs, as shown in Figure 4a. Thus, to identify PCs, even if W is associated with Z or Y, we use the information for an intermediate variable S, as in Figure 4b. The first step for identifying PCs is to consider the graph obtained by conditioning on W in graph G, as shown in Figure 4c. Here, the dashed, directed edges in Figure 4c indicate that there is no path through W (or such edges are removed) in Figure 4b because of conditioning on W. Then, since Figure 4c implies that $p(x, y, z, s, u \mid w)$ is identifiable through Theorem 1, $p(x, y, z, s, w, u) = p(x, y, z, s, u \mid w)p(w)$ is also identifiable. The second step is to consider the graph obtained by marginalizing Y and Z in Figure 4d, as shown in Figure 4d. Since $p(x, y, z, s, w, u)$ is now available by Theorem 1, Theorem 2 shows that $p(x, w, s, u, v)$ is identifiable. Then, $p(u, v) = \sum_{x,w,s} p(x, w, s, u, v)$ is also identifiable.

Noting the composition property

$$S_x(i) = s \implies Y_{x,s}(i) = Y_x(i) \tag{14}$$

Algorithm 1 Identification of the joint distribution of (Y_{x_0}, Y_{x_1}).

Input: Joint distribution $p(x, y, z, s, w)$ according to Figure 4b
Output: Joint distribution $p(Y_{x_0} = y, Y_{x_1} = y')$

1: Identify $p(x, y, z, s, u \mid w_0)$ and $p(x, y, z, s, u \mid w_1)$ applying Theorem 1 after conditioning $W = w_0$ and $W = w_1$, respectively.
2: Identify $p(x, y, z, s, w_0, u)$ as $p(x, y, z, s, u \mid w_0)p(w_0)$ and $p(x, y, z, s, w_1, u)$ as $p(x, y, z, s, u \mid w_1)p(w_1)$.
3: Identify $p(x, w, s, u, v)$ applying Theorem 2 after identifying $p(x, s, w, u) = \sum_{y,z} p(x, y, z, s, w, u)$.
4: Identify $p(Y_{x_0}, Y_{x_1})$ from equation (16) after identifying $p(u, v) = \sum_{x,w,s} p(x, w, s, u, v)$.

for the i-th subject (Pearl, 2009), we obtain

$$\begin{aligned} &p(Y_{x_0} = y, Y_{x_1} = y') \\ &= \sum_{s,s'} p(Y_{x_0, S_{x_0}} = y, Y_{x_1, S_{x_1}} = y', S_{x_0} = s, S_{x_1} = s') \\ &= \sum_{s,s'} p(Y_s = y, Y_{s'} = y', S_{x_0} = s, S_{x_1} = s'). \end{aligned} \tag{15}$$

Therefore, the joint probabilities of the potential outcomes is identifiable from $p(u, v)$ as follows:

$$\begin{aligned} p(Y_{x_0} = y_0, Y_{x_1} = y_0) &= p(u_1, v_1) + p(u_2, v_1) + p(u_1, v_2) \\ &\quad + p(u_1, v_3) + p(u_1, v_4) + p(u_3, v_4), \\ p(Y_{x_0} = y_0, Y_{x_1} = y_1) &= p(u_2, v_2) + p(u_3, v_3), \\ p(Y_{x_0} = y_1, Y_{x_1} = y_0) &= p(u_3, v_2) + p(u_2, v_3), \\ p(Y_{x_0} = y_1, Y_{x_1} = y_1) &= p(u_3, v_1) + p(u_4, v_1) + p(u_4, v_2) \\ &\quad + p(u_4, v_3) + p(u_2, v_4) + p(u_4, v_4). \end{aligned} \tag{16}$$

The proposed identification conditions show that we can derive consistent estimators of the joint probabilities of potential outcome variables via the method of moments. For further details of the estimation, see the Supplementary Material B.

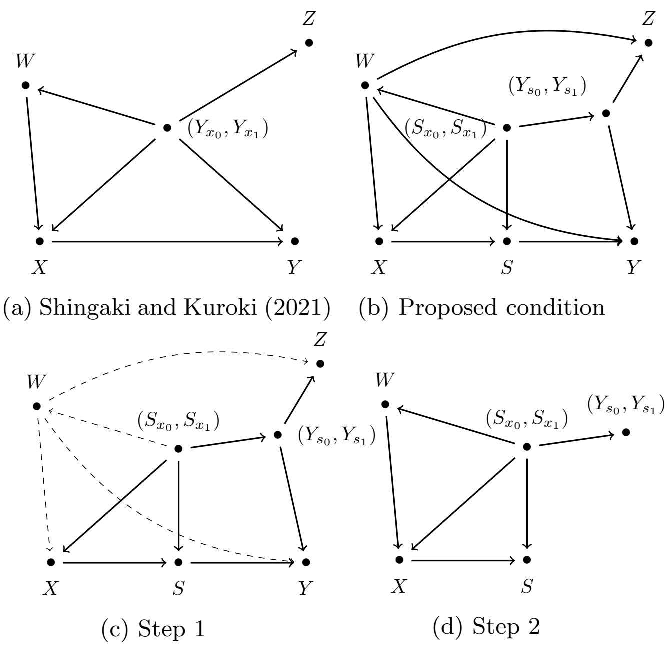
5 NUMERICAL EXPERIMENT

In this section, we present a numerical experiment to examine the properties of our proposed estimation method using the joint probabilities of the potential outcomes $p(Y_{x_0} = y_0, Y_{x_1} = y_0)$, $p(Y_{x_0} = y_0, Y_{x_1} = y_1)$, $p(Y_{x_0} = y_1, Y_{x_1} = y_0)$, and $p(Y_{x_0} = y_1, Y_{x_1} = y_1)$. For simplicity, letting X, Y, Z, S, W, U, and V be discrete variables, we consider the causal diagrams shown in Figure 4b, where the joint probabilities of (X, Y, Z, W, U) are given according to Table 1. Under the situation where (X, Y, Z, S, W) can be observed

Table 1: Conditional probability tables in the simulation.

	$p(W \mid V)$		$p(X=0 \mid W,V)$		$p(S=0 \mid X,V)$		$p(U,V)$			
	$W=0$	$W=1$	$W=0$	$W=1$	$X=0$	$X=1$	$U=1$	$U=2$	$U=3$	$U=4$
$V=1$	1/2	1/2	1/2	1/2	1	1	10/80	6/80	3/80	1/80
$V=2$	1/2	1/2	1/2	1/2	1	0	6/80	10/80	1/80	3/80
$V=3$	1/2	1/2	1/2	1/2	0	1	3/80	1/80	10/80	6/80
$V=4$	1/2	1/2	1/2	1/2	0	0	1/80	3/80	6/80	10/80

	$p(Y=0 \mid S,U)$		$p(Z=1 \mid W,U)$		$p(Z=2 \mid W,U)$		$p(Z=3 \mid W,U)$	
	$S=0$	$S=1$	$W=0$	$W=1$	$W=0$	$W=1$	$W=0$	$W=1$
$U=1$	1	1	8/10	8/10	1/20	1/20	3/20	3/20
$U=2$	1	0	3/20	3/10	8/10	8/20	1/20	1/20
$U=3$	0	1	7/10	7/10	1/10	1/10	2/10	2/10
$U=4$	0	0	2/10	2/10	7/10	7/10	1/10	1/10



(a) Shingaki and Kuroki (2021) (b) Proposed condition

(c) Step 1 (d) Step 2

Figure 4: Difference between our results and those from Shingaki and Kuroki (2021).

but U and V cannot, the properties of the proposed estimators $\widehat{p}(Y_{x_0}=y_0, Y_{x_1}=y_0)$, $\widehat{p}(Y_{x_0}=y_0, Y_{x_1}=y_1)$, $\widehat{p}(Y_{x_0}=y_1, Y_{x_1}=y_0)$, and $\widehat{p}(Y_{x_0}=y_1, Y_{x_1}=y_1)$ of $p(Y_{x_0}=y_0, Y_{x_1}=y_0)$, $p(Y_{x_0}=y_0, Y_{x_1}=y_1)$, $p(Y_{x_0}=y_1, Y_{x_1}=y_0)$, and $p(Y_{x_0}=y_1, Y_{x_1}=y_1)$, respectively, are verified in a numerical experiment using a setting with sample sizes $n=500$, 1000, 5000, and 10000. In this situation, since $p(Y_{x_0}=y_0, Y_{x_1}=y_0)=0.4$, $p(Y_{x_0}=y_0, Y_{x_1}=y_1)=0.25$, $p(Y_{x_0}=y_1, Y_{x_1}=y_0)=0.025$, and $p(Y_{x_0}=y_1, Y_{x_1}=y_1)=0.325$, the sample means of $\widehat{p}(Y_{x_0}=y_0, Y_{x_1}=y_0)$, $\widehat{p}(Y_{x_0}=y_0, Y_{x_1}=y_1)$, $\widehat{p}(Y_{x_0}=y_1, Y_{x_1}=y_0)$, and $\widehat{p}(Y_{x_0}=y_1, Y_{x_1}=y_1)$ are expected to be close to 0.4, 0.25, 0.025, and 0.325, respectively. Table 2 shows the basic statistics and the box plots of $\widehat{p}(Y_{x_0}=y_0, Y_{x_1}=y_0)$, $\widehat{p}(Y_{x_0}=y_0, Y_{x_1}=y_1)$, $\widehat{p}(Y_{x_0}=y_1, Y_{x_1}=y_0)$, and $\widehat{p}(Y_{x_0}=y_1, Y_{x_1}=y_1)$ for 1000 replications with the given sample size n, respectively.

From Table 2, the sample means of $\widehat{p}(Y_{x_0}=y_0, Y_{x_1}=y_0)$, $\widehat{p}(Y_{x_0}=y_0, Y_{x_1}=y_1)$, $\widehat{p}(Y_{x_0}=y_1, Y_{x_1}=y_0)$, and $\widehat{p}(Y_{x_0}=y_1, Y_{x_1}=y_1)$ tend to approach the true values, and the sample standard errors decrease as the sample size increases. Thus, it seems that the proposed estimation method provides the consistent estimators of $p(Y_{x_0}=y_0, Y_{x_1}=y_0)$ and $p(Y_{x_0}=y_1, Y_{x_1}=y_1)$. In contrast, the sample means of for $\widehat{p}(Y_{x_0}=y_0, Y_{x_1}=y_1)$ and $\widehat{p}(Y_{x_0}=y_1, Y_{x_1}=y_0)$, whose true values are relatively small, tend to approach true values very slowly. The proposed estimation method requires calculating small nonzero determinants of probability matrices to estimate the joint probabilities of the potential outcomes. When the true joint probabilities of the potential outcomes are small, it is difficult to calculate the small nonzero determinants. This may be a reason why $\widehat{p}(Y_{x_0}=y_0, Y_{x_1}=y_1)$ and $\widehat{p}(Y_{x_0}=y_1, Y_{x_1}=y_0)$ is difficult to estimate.

6 DISCUSSION

In addressing the challenges associated with identifying and estimating probabilities of causation, this study introduced novel identification conditions and a statistical estimation method. Our approach employed an intermediate variable in conjunction with covariate information to tackle the identification and estimation problems. By introducing the intermediate variable, the proposed method expands the scope of identifiable situations of Shingaki and Kuroki (2021),

Table 2: Basic statistics in the numerical experiment.

	(a) $\widehat{p}(Y_{x_0}=y_0, Y_{x_1}=y_0)$				(b) $\widehat{p}(Y_{x_0}=y_0, Y_{x_1}=y_1)$			
	$n=500$	$n=1000$	$n=5000$	$n=10000$	$n=500$	$n=1000$	$n=5000$	$n=10000$
Minimum	0.162	0.089	0.185	0.202	0.000	0.000	0.000	0.000
1st Quantile	0.403	0.399	0.402	0.402	0.115	0.114	0.123	0.124
Median	0.489	0.480	0.469	0.460	0.198	0.200	0.189	0.200
Mean	0.523	0.514	0.504	0.484	0.208	0.204	0.201	0.203
3rd Quantile	0.601	0.587	0.560	0.536	0.286	0.288	0.276	0.275
Maximum	1.459	1.488	1.318	1.323	0.576	0.869	0.564	0.754
s.e.	0.187	0.178	0.158	0.135	0.122	0.123	0.108	0.108
	(c) $\widehat{p}(Y_{x_0}=y_1, Y_{x_1}=y_0)$				(d) $\widehat{p}(Y_{x_0}=y_1, Y_{x_1}=y_1)$			
	$n=500$	$n=1000$	$n=5000$	$n=10000$	$n=500$	$n=1000$	$n=5000$	$n=10000$
Minimum	0.000	0.000	0.000	0.000	0.010	0.045	0.070	0.031
1st Quantile	0.055	0.057	0.062	0.062	0.274	0.277	0.270	0.258
Median	0.104	0.102	0.105	0.106	0.359	0.342	0.326	0.322
Mean	0.118	0.115	0.115	0.112	0.379	0.371	0.343	0.327
3rd Quantile	0.158	0.157	0.151	0.150	0.447	0.423	0.395	0.377
Maximum	0.637	0.595	0.709	0.615	1.274	1.106	1.000	0.957
s.e.	0.089	0.083	0.083	0.071	0.155	0.150	0.120	0.107

as it no longer necessitates the conditional independence $W \perp\!\!\!\perp Z \mid \{X, U\}$ between two proxy covariates to identify probabilities of causation.

Our estimation method may have difficulty obtaining reliable statistics of the recovered probabilities due to numerical difficulties in computing small nonzero determinants. To account for the standard errors of the recovered probabilities, we use standard bootstrap methods in Section 5. In practice, it is also important to select appropriate proxy variables from several candidate covariates that satisfy our proposed conditions so as not to suffer from such numerical difficulties. This is a problem that we are leaving for future work.

In this study, we specifically concentrated on a simple scenario involving a single treatment variable X and a single outcome variable Y. It is worth noting that our findings can be readily extended to settings featuring multiple treatment variables. However, such extensions necessitate the observation of proxy variables with more than two or three categories. This requirement stems from the increased complexity in the probabilities of causation, contingent upon the number of treatment variables of interest.

Finally, it is essential to recognize that the typical covariates present in a given observational study often serve as reliable proxies for potential outcome variables. One of the foundational assumptions in our proposed method is the presence of a proxy variable Z, that remains independent of both the treatment X and the outcome Y. Fortunately, identifying such a proxy is not arduous, especially in studies like a GWAS (genome-wide association study) where a vast array of covariates can be observed. Furthermore, in studies of this nature, variables that can satisfy the front-door criterion in our causal diagram can serve as prime candidates for an intermediate variable in our proposed methodology. Consequently, our approach is widely applicable for identifying and estimating "causes of effects" (CoE) across a multitude of observational studies.

Acknowledgements

We would like to acknowledge the helpful comments of the four anonymous reviewers. This research was partially supported by Grant-in-Aid for Scientific Research (C) from the Japan Society for the Promotion of Science, Grant Number 19K11856.

References

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455.

Beyea, J. and Greenland, S. (1999). The importance of specifying the underlying biologic model in estimating the probability of causation. *Health Physics*, 76(3):269–274.

Cai, Z. and Kuroki, M. (2005). Variance estimators for three "probabilities of causation". *Risk Analysis*, 25(6):1611–1620.

Cox, L. A. (1984). Probability of causation and the attributable proportion risk. *Risk Analysis*, 4(3):221–230.

Dammann, O. (2020). *Etiological Explanations: Illness Causation Theory*. CRC Press.

Dawid, A. P. and Musio, M. (2022). What can group-level data tell us about individual causality? In *Statistics in the Public Interest*, pages 235–256. Springer.

Dawid, A. P., Musio, M., and Murtas, R. (2017). The probability of causation. *Law, Probability and Risk*, 16(4):163–179.

Dawid, P., Humphreys, M., and Musio, M. (2022). Bounding causes of effects with mediators. *Sociological Methods and Research*, (Online First Articles).

Dehejia, R. H. and Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94(448):1053–1062.

Ferguson, T. S. (1996). *A Course in Large Sample Theory*. Routledge.

Galhotra, S., Pradhan, R., and Salimi, B. (2021). Explaining black-box algorithms using probabilistic contrastive counterfactuals. In *Proceedings of the 2021 International Conference on Management of Data*, SIGMOD '21, page 577–590, New York, NY, USA. Association for Computing Machinery.

Greenland, S. (1987). Variance estimators for attributable fraction estimates consistent in both large strata and sparse data. *Statist. Med.*, 6(6):701–708.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Kawakami, Y., Shingaki, R., and Kuroki, M. (2023). Identification and estimation of the probabilities of potential outcome types using covariate information in studies with non-compliance. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(10):12234–12242.

Khoury, M. J., Flanders, W. D., Greenland, S., and Adams, M. J. (1989). On the measurement of susceptibility in epidemiologic studies. *American Journal of Epidemiology*, 129(1):183–190.

Kommiya Mothilal, R., Mahajan, D., Tan, C., and Sharma, A. (2021). Towards unifying feature attribution and counterfactual explanations: Different means to the same end. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '21, pages 652–663, New York, NY, USA. Association for Computing Machinery.

Kuroki, M. and Cai, Z. (2011). Statistical analysis of 'probabilities of causation' using co-variate information. *Scand. J. Statist.*, 38(3):564–577.

Kuroki, M. and Pearl, J. (2014). Measurement bias and effect restoration in causal inference. *Biometrika*, 101(2):423–437.

LaLonde, R. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review*, 76(4):604–20.

Lu, Z., Geng, Z., Li, W., Zhu, S., and Jia, J. (2023). Evaluating causes of effects by posterior effects of causes. *Biometrika*, 110(2):449–465.

Mueller, S., Li, A., and Pearl, J. (2022). Causes of effects: Learning individual responses from population data. *International Joint Conference on Artificial Intelligence*, 31.

Murtas, R., Dawid, A. P., and Musio, M. (2017). New bounds for the probability of causation in mediation analysis. *arXiv preprint arXiv:1706.04857*.

Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pages 411–420, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2nd edition.

Robins, J. (1989). The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies. In Sechrest, L., Freeman, H., and Mulley, A., editors, *Health Service Research Methodology: A Focus on AIDS*, pages 113–159. U.S. Public Health Service, Washington, DC.

Robins, J. and Greenland, S. (1989). The probability of causation under a stochastic model for individual risk. *Biometrics*, 45(4):1125–1138.

Robins, J. and Richardson, T. S. (2011). Alternative graphical causal models and the identification of direct effects. In Shrout, P., Keyes, K., and Ornstein, K., editors, *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures*, pages 103–158. Oxford University Press.

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

Shingaki, R. and Kuroki, M. (2021). Identification and estimation of joint probabilities of potential outcomes in observational studies with covariate information. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*, volume 34, pages 26475–26486. Curran Associates, Inc.

Shingaki, R. and Kuroki, M. (2023). Probabilities of potential outcome types in experimental studies: Identification and estimation based on proxy covariate information. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(10):12287–12294.

Tian, J. and Pearl, J. (2000). Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28(1):287–313.

Tian, J. and Pearl, J. (2002). A general identification condition for causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pages 567–573, USA. American Association for Artificial Intelligence.

VanderWeele, T. J. (2012). The sufficient cause framework in statistics, philosophy and the biomedical and social sciences. In Berzuini, C., Dawid, P., and Bernardinell, L., editors, *Causality: Statistical Perspectives and Applications*, chapter 13, pages 180–191. Wiley.

Watson, D. S., Gultchin, L., Taly, A., and Floridi, L. (2021). Local explanations via necessity and sufficiency: unifying theory and practice. In de Campos, C. and Maathuis, M. H., editors, *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pages 1382–1392. PMLR.

Wright, S. (1923). The theory of path coefficients a reply to niles's criticism. *Genetics*, 8(3):239.

Wright, S. (1934). The method of path coefficients. *The Annals of Mathematical Statistics*, 5(3):161–215.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes] We describe the analysis changing the sample size in the numerical experiment section.
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Yes]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Identification and Estimation of "Causes of Effects" using Covariate-Mediator Information (Supplementary Material)

A PROOFS OF THEOREMS

A.1 Proof of Theorem 1

From Conditions 1 and 2 in Theorem 1, and by the consistency property, we have

$$\begin{aligned} p(x,y_0,z,s_0) &= \sum_{i=1}^{4} p(y_0 \,|\, x,z,s_0,u_i)p(z \,|\, s_0,x,u_i)p(x,u_i,s_0) \\ &= \sum_{i=1}^{4} p(y_0 \,|\, x,z,s_0,u_i)p(z \,|\, u_i)p(x \,|\, s_0,u_i)p(u_i,s_0) = \sum_{i=1,2} p(x \,|\, s_0,u_i)p(z \,|\, u_i)p(u_i,s_0), \qquad \text{(A.1)} \\ p(x,y_1,z,s_0) &= \sum_{i=1}^{4} p(y_1 \,|\, x,z,s_0,u_i)p(z \,|\, u_i)p(x \,|\, s_0,u_i)p(u_i,s_0) = \sum_{i=3,4} p(x \,|\, s_0,u_i)p(z \,|\, u_i)p(u_i,s_0), \qquad \text{(A.2)} \\ p(x,y_0,z,s_1) &= \sum_{i=1}^{4} p(y_0 \,|\, x,z,s_1,u_i)p(z \,|\, u_i)p(x \,|\, s_1,u_i)p(u_i,s_1) = \sum_{i=1,3} p(x \,|\, s_1,u_i)p(z \,|\, u_i)p(u_i,s_1), \qquad \text{(A.3)} \\ p(x,y_1,z,s_1) &= \sum_{i=1}^{4} p(y_1 \,|\, x,z,s_1,u_i)p(z \,|\, u_i)p(x \,|\, s_1,u_i)p(u_i,s_1) = \sum_{i=2,4} p(x \,|\, s_1,u_i)p(z \,|\, u_i)p(u_i,s_1), \qquad \text{(A.4)} \end{aligned}$$

for $x \in \{x_0, x_1\}$ and $z \in \{z_1, z_2, z_3\}$, since

$$\begin{array}{llll} p(y_0 \,|\, x,z,s_0,u_i) = 1 & \text{for } i = 1,2, & p(y_0 \,|\, x,z,s_0,u_i) = 0 & \text{for } i = 3,4, \\ p(y_1 \,|\, x,z,s_0,u_i) = 1 & \text{for } i = 3,4, & p(y_1 \,|\, x,z,s_0,u_i) = 0 & \text{for } i = 1,2, \\ p(y_0 \,|\, x,z,s_1,u_i) = 1 & \text{for } i = 1,3, & p(y_0 \,|\, x,z,s_1,u_i) = 0 & \text{for } i = 2,4, \\ p(y_1 \,|\, x,z,s_1,u_i) = 1 & \text{for } i = 2,4, & p(y_1 \,|\, x,z,s_1,u_i) = 0 & \text{for } i = 1,3 \end{array} \qquad \text{(A.5)}$$

hold. Thus, letting

$$P_{xyzs_0} = \begin{pmatrix} p(y,s_0) & p(x,y,s_0) \\ p(y,z,s_0) & p(x,y,z,s_0) \end{pmatrix}, \qquad \text{(A.6)}$$

$$Q_{y_0zs_0} = \begin{pmatrix} 1 & p(z \,|\, u_1) \\ 1 & p(z \,|\, u_2) \end{pmatrix}, \quad \Delta_{y_0s_0} = \begin{pmatrix} p(u_1,s_0) & 0 \\ 0 & p(u_2,s_0) \end{pmatrix}, \quad R_{xy_0s_0} = \begin{pmatrix} 1 & p(x \,|\, s_0,u_1) \\ 1 & p(x \,|\, s_0,u_2) \end{pmatrix}, \qquad \text{(A.7)}$$

$$Q_{y_1zs_0} = \begin{pmatrix} 1 & p(z \,|\, u_3) \\ 1 & p(z \,|\, u_4) \end{pmatrix}, \quad \Delta_{y_1s_0} = \begin{pmatrix} p(u_3,s_0) & 0 \\ 0 & p(u_4,s_0) \end{pmatrix}, \quad R_{xy_1s_0} = \begin{pmatrix} 1 & p(x \,|\, s_0,u_3) \\ 1 & p(x \,|\, s_0,u_4) \end{pmatrix}, \qquad \text{(A.8)}$$

since P_{xyzs_0} is invertible from Condition 3 in Theorem 1, from equations (A.1) and (A.2) we obtain

$$P_{xyzs_0} = Q^{\top}_{yzs_0}\Delta_{ys_0}R_{xys_0}, \qquad \text{(A.9)}$$

where $\top$ indicates for a transposed vector/matrix. Thus, we have

$$P^{-1}_{xyzs_0}Q^{\top}_{yzs_0} = R^{-1}_{xys_0}\Delta^{-1}_{ys_0} \qquad \text{(A.10)}$$

for $z \in \{z_1, z_2, z_3\}$, that is,

$$P^{-1}_{xyz_1s_0} Q^{\top}_{yz_1s_0} = P^{-1}_{xyz_2s_0} Q^{\top}_{yz_2s_0}, \tag{A.11}$$

specifically,

$$\begin{aligned} &\frac{1}{\det(P_{xy_0z_1s_0})} \begin{pmatrix} p(x,y_0,z_1,s_0) & -p(x,y_0,s_0) \\ -p(y_0,z_1,s_0) & p(y_0,s_0) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ p(z_1 \mid u_1) & p(z_1 \mid u_2) \end{pmatrix} \\ &= \frac{1}{\det(P_{xy_0z_2s_0})} \begin{pmatrix} p(x,y_0,z_2,s_0) & -p(x,y_0,s_0) \\ -p(y_0,z_2,s_0) & p(y_0,s_0) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ p(z_2 \mid u_1) & p(z_2 \mid u_2) \end{pmatrix}, \end{aligned} \tag{A.12}$$

$$\begin{aligned} &\frac{1}{\det(P_{xy_1z_1s_0})} \begin{pmatrix} p(x,y_1,z_1,s_0) & -p(x,y_1,s_0) \\ -p(y_1,z_1,s_0) & p(y_1,s_0) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ p(z_1 \mid u_3) & p(z_1 \mid u_4) \end{pmatrix} \\ &= \frac{1}{\det(P_{xy_1z_2s_0})} \begin{pmatrix} p(x,y_1,z_2,s_0) & -p(x,y_1,s_0) \\ -p(y_1,z_2,s_0) & p(y_1,s_0) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ p(z_2 \mid u_3) & p(z_2 \mid u_4) \end{pmatrix} \end{aligned} \tag{A.13}$$

hold for $x \in \{x_0, x_1\}$. Thus, we derive

$$p(y_0,s_0)p(z_1 \mid u_1) - p(y_0,z_1,s_0) = \frac{\det(P_{xy_0z_1s_0})}{\det(P_{xy_0z_2s_0})}(p(y_0,s_0)p(z_2 \mid u_1) - p(y_0,z_2,s_0)), \tag{A.14}$$

$$p(y_0,s_0)p(z_1 \mid u_2) - p(y_0,z_1,s_0) = \frac{\det(P_{xy_0z_1s_0})}{\det(P_{xy_0z_2s_0})}(p(y_0,s_0)p(z_2 \mid u_2) - p(y_0,z_2,s_0)), \tag{A.15}$$

$$p(y_1,s_0)p(z_1 \mid u_3) - p(y_1,z_1,s_0) = \frac{\det(P_{xy_1z_1s_0})}{\det(P_{xy_1z_2s_0})}(p(y_1,s_0)p(z_2 \mid u_3) - p(y_1,z_2,s_0)), \tag{A.16}$$

$$p(y_1,s_0)p(z_1 \mid u_4) - p(y_1,z_1,s_0) = \frac{\det(P_{xy_1z_1s_0})}{\det(P_{xy_1z_2s_0})}(p(y_1,s_0)p(z_2 \mid u_4) - p(y_1,z_2,s_0)). \tag{A.17}$$

Similarly, letting

$$P_{xyzs_1} = \begin{pmatrix} p(y,s_1) & p(x,y,s_1) \\ p(y,z,s_1) & p(x,y,z,s_1) \end{pmatrix}, \tag{A.18}$$

$$Q_{y_0zs_1} = \begin{pmatrix} 1 & p(z \mid u_1) \\ 1 & p(z \mid u_3) \end{pmatrix}, \quad \Delta_{y_0s_1} = \begin{pmatrix} p(u_1,s_1) & 0 \\ 0 & p(u_3,s_1) \end{pmatrix}, \quad R_{xy_0s_1} = \begin{pmatrix} 1 & p(x \mid s_1,u_1) \\ 1 & p(x \mid s_1,u_3) \end{pmatrix}, \tag{A.19}$$

$$Q_{y_1zs_1} = \begin{pmatrix} 1 & p(z \mid u_2) \\ 1 & p(z \mid u_4) \end{pmatrix}, \quad \Delta_{y_1s_1} = \begin{pmatrix} p(u_2,s_1) & 0 \\ 0 & p(u_4,s_1) \end{pmatrix}, \quad R_{xy_1s_1} = \begin{pmatrix} 1 & p(x \mid s_1,u_2) \\ 1 & p(x \mid s_1,u_4) \end{pmatrix}, \tag{A.20}$$

since P_{xyzs_1} is invertible from Condition 3 in Theorem 1, from equations (A.3) and (A.4) we obtain

$$P_{xyzs_1} = Q^{\top}_{yzs_1} \Delta_{ys_1} R_{xys_1} \tag{A.21}$$

for $x \in \{x_0, x_1\}$, $y \in \{y_0, y_1\}$ and $z \in \{z_1, z_2, z_3\}$. Thus, we have

$$P^{-1}_{xyzs_1} Q^{\top}_{yzs_1} = R^{-1}_{xys_1} \Delta^{-1}_{ys_1} \tag{A.22}$$

for $z \in \{z_1, z_2, z_3\}$, that is,

$$P^{-1}_{xyz_1s_1} Q^{\top}_{yz_1s_1} = P^{-1}_{xyz_2s_1} Q^{\top}_{yz_2s_1}, \tag{A.23}$$

specifically,

$$\begin{aligned} &\frac{1}{\det(P_{xy_0z_1s_1})} \begin{pmatrix} p(x,y_0,z_1,s_1) & -p(x,y_0,s_1) \\ -p(y_0,z_1,s_1) & p(y_0,s_1) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ p(z_1 \mid u_1) & p(z_1 \mid u_3) \end{pmatrix} \\ &= \frac{1}{\det(P_{xy_0z_2s_1})} \begin{pmatrix} p(x,y_0,z_2,s_1) & -p(x,y_0,s_1) \\ -p(y_0,z_2,s_1) & p(y_0,s_1) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ p(z_2 \mid u_1) & p(z_2 \mid u_3) \end{pmatrix}, \end{aligned} \tag{A.24}$$

$$\begin{aligned} &\frac{1}{\det(P_{xy_1z_1s_1})} \begin{pmatrix} p(x,y_1,z_1,s_1) & -p(x,y_1,s_1) \\ -p(y_1,z_1,s_1) & p(y_1,s_1) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ p(z_1 \mid u_2) & p(z_1 \mid u_4) \end{pmatrix} \\ &= \frac{1}{\det(P_{xy_1z_2s_1})} \begin{pmatrix} p(x,y_1,z_2,s_1) & -p(x,y_1,s_1) \\ -p(y_1,z_2,s_1) & p(y_1,s_1) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ p(z_2 \mid u_2) & p(z_2 \mid u_4) \end{pmatrix} \end{aligned} \tag{A.25}$$

hold for $x \in \{x_0, x_1\}$. Thus, we derive

$$p(y_0, s_1)p(z_1 \mid u_1) - p(y_0, z_1, s_1) = \frac{\det(P_{xy_0z_1s_1})}{\det(P_{xy_0z_2s_1})}(p(y_0, s_1)p(z_2 \mid u_1) - p(y_0, z_2, s_1)), \tag{A.26}$$

$$p(y_1, s_1)p(z_1 \mid u_2) - p(y_1, z_1, s_1) = \frac{\det(P_{xy_1z_1s_1})}{\det(P_{xy_1z_2s_1})}(p(y_1, s_1)p(z_2 \mid u_2) - p(y_1, z_2, s_1)), \tag{A.27}$$

$$p(y_0, s_1)p(z_1 \mid u_3) - p(y_0, z_1, s_1) = \frac{\det(P_{xy_0z_1s_1})}{\det(P_{xy_0z_2s_1})}(p(y_0, s_1)p(z_2 \mid u_3) - p(y_0, z_2, s_1)), \tag{A.28}$$

$$p(y_1, s_1)p(z_1 \mid u_4) - p(y_1, z_1, s_1) = \frac{\det(P_{xy_1z_1s_1})}{\det(P_{xy_1z_2s_1})}(p(y_1, s_1)p(z_2 \mid u_4) - p(y_1, z_2, s_1)). \tag{A.29}$$

From Condition 3 in Theorem 1, a system of eight linear equations (A.14)-(A.17) and (A.26)-(A.29) provides a unique solution regarding $p(z|u)$ for $z \in \{z_1, z_2\}$. Indeed, if Condition 3 in Theorem 1 holds, we have

$$\begin{pmatrix} p(z_1 \mid u_1) \\ p(z_2 \mid u_1) \end{pmatrix} = \begin{pmatrix} 1 & -\frac{\det(P_{xy_0z_1s_0})}{\det(P_{xy_0z_2s_0})} \\ 1 & -\frac{\det(P_{xy_0z_1s_1})}{\det(P_{xy_0z_2s_1})} \end{pmatrix}^{-1} \begin{pmatrix} p(z_1 \mid y_0, s_0) - \frac{\det(P_{xy_0z_1s_0})}{\det(P_{xy_0z_2s_0})}p(z_2 \mid y_0, s_0) \\ p(z_1 \mid y_0, s_1) - \frac{\det(P_{xy_0z_1s_1})}{\det(P_{xy_0z_2s_1})}p(z_2 \mid y_0, s_1) \end{pmatrix}, \tag{A.30}$$

$$\begin{pmatrix} p(z_1 \mid u_2) \\ p(z_2 \mid u_2) \end{pmatrix} = \begin{pmatrix} 1 & -\frac{\det(P_{xy_0z_1s_0})}{\det(P_{xy_0z_2s_0})} \\ 1 & -\frac{\det(P_{xy_1z_1s_1})}{\det(P_{xy_1z_2s_1})} \end{pmatrix}^{-1} \begin{pmatrix} p(z_1 \mid y_0, s_0) - \frac{\det(P_{xy_0z_1s_0})}{\det(P_{xy_0z_2s_0})}p(z_2 \mid y_0, s_0) \\ p(z_1 \mid y_1, s_1) - \frac{\det(P_{xy_1z_1s_1})}{\det(P_{xy_1z_2s_1})}p(z_2 \mid y_1, s_1) \end{pmatrix}, \tag{A.31}$$

$$\begin{pmatrix} p(z_1 \mid u_3) \\ p(z_2 \mid u_3) \end{pmatrix} = \begin{pmatrix} 1 & -\frac{\det(P_{xy_1z_1s_0})}{\det(P_{xy_1z_2s_0})} \\ 1 & -\frac{\det(P_{xy_0z_1s_1})}{\det(P_{xy_0z_2s_1})} \end{pmatrix}^{-1} \begin{pmatrix} p(z_1 \mid y_1, s_0) - \frac{\det(P_{xy_1z_1s_0})}{\det(P_{xy_1z_2s_0})}p(z_2 \mid y_1, s_0) \\ p(z_1 \mid y_0, s_1) - \frac{\det(P_{xy_0z_1s_1})}{\det(P_{xy_0z_2s_1})}p(z_2 \mid y_0, s_1) \end{pmatrix}, \tag{A.32}$$

$$\begin{pmatrix} p(z_1 \mid u_4) \\ p(z_2 \mid u_4) \end{pmatrix} = \begin{pmatrix} 1 & -\frac{\det(P_{xy_1z_1s_0})}{\det(P_{xy_1z_2s_0})} \\ 1 & -\frac{\det(P_{xy_1z_1s_1})}{\det(P_{xy_1z_2s_1})} \end{pmatrix}^{-1} \begin{pmatrix} p(z_1 \mid y_1, s_0) - \frac{\det(P_{xy_1z_1s_0})}{\det(P_{xy_1z_2s_0})}p(z_2 \mid y_1, s_0) \\ p(z_1 \mid y_1, s_1) - \frac{\det(P_{xy_1z_1s_1})}{\det(P_{xy_1z_2s_1})}p(z_2 \mid y_1, s_1) \end{pmatrix}. \tag{A.33}$$

Therefore, noting that $Q_{yzs}^{\top}$ is identifiable and invertible, and that P_{xyzs} is available for $x \in \{x_0, x_1\}$, $y \in \{y_0, y_1\}$, $s \in \{s_0, s_1\}$, and $z \in \{z_1, z_2, z_3\}$, we derive

$$Q_{y_0zs_0}^{-\top}P_{xy_0zs_0} = \Delta_{y_0s_0}R_{xy_0s_0} = \begin{pmatrix} p(u_1, s_0) & p(x, u_1, s_0) \\ p(u_2, s_0) & p(x, u_2, s_0) \end{pmatrix}, \tag{A.34}$$

$$Q_{y_1zs_0}^{-\top}P_{xy_1zs_0} = \Delta_{y_1s_0}R_{xy_1s_0} = \begin{pmatrix} p(u_3, s_0) & p(x, u_3, s_0) \\ p(u_4, s_0) & p(x, u_4, s_0) \end{pmatrix}, \tag{A.35}$$

$$Q_{y_0zs_1}^{-\top}P_{xy_0zs_1} = \Delta_{y_0s_1}R_{xy_0s_1} = \begin{pmatrix} p(u_1, s_1) & p(x, u_1, s_1) \\ p(u_3, s_1) & p(x, u_3, s_1) \end{pmatrix}, \tag{A.36}$$

$$Q_{y_1zs_1}^{-\top}P_{xy_1zs_1} = \Delta_{y_1s_1}R_{xy_1s_1} = \begin{pmatrix} p(u_2, s_1) & p(x, u_2, s_1) \\ p(u_4, s_1) & p(x, u_4, s_1) \end{pmatrix}. \tag{A.37}$$

Hence, since $p(z \mid x, s, u) = p(z \mid u)$ by Condition 2 in Theorem 1,

$$\begin{aligned} p(x, y, z, s, u) &= p(y \mid x, z, s, u)p(x, z, s, u) \\ &= p(y \mid x, z, s, u)p(z \mid x, s, u)p(x, s, u) \\ &= p(y \mid x, z, s, u)p(z \mid u)p(x, s, u) \end{aligned}$$

are identifiable for $x \in \{x_0, x_1\}$, $y \in \{y_0, y_1\}$, $s \in \{s_0, s_1\}$, $z \in \{z_1, z_2, z_3\}$, and $u \in \{u_1, u_2, u_3, u_4\}$. Equivalently, by equations (A.5), we have

$$p(x, y_0, z, s_0, u_i) = \begin{cases} p(z \mid u_i)p(x, u_i, s_0) & \text{for } i = 1, 2, \\ 0 & \text{for } i = 3, 4, \end{cases}$$

$$p(x, y_1, z, s_0, u_i) = \begin{cases} 0 & \text{for } i = 1, 2, \\ p(z \mid u_i) p(x, u_i, s_0) & \text{for } i = 3, 4, \end{cases}$$

$$p(x, y_0, z, s_1, u_i) = \begin{cases} p(z \mid u_i) p(x, u_i, s_1) & \text{for } i = 1, 3, \\ 0 & \text{for } i = 2, 4, \end{cases}$$

$$p(x, y_1, z, s_1, u_i) = \begin{cases} 0 & \text{for } i = 1, 3, \\ p(z \mid u_i) p(x, u_i, s_1) & \text{for } i = 2, 4 \end{cases}$$

for $x \in \{x_0, x_1\}$ and $z \in \{z_1, z_2, z_3\}$. □

A.2 Proof of Theorem 2

From Conditions 1 and 2 in Theorem 2, and by the consistency property, we have

$$\begin{aligned} p(w, s_0, u, x_0) &= \sum_{i=1}^{4} p(s_0 \mid w, u, x_0, v_i) p(u \mid x_0, w, v_i) p(x_0 \ w, v_i) \\ &= \sum_{i=1}^{4} p(s_0 \mid w, u, x_0, v_i) p(x_0, w, v_i) p(u \mid v_i) \\ &= \sum_{i=1}^{4} p(s_0 \mid w, u, x_0, v_i) p(w \mid x_0, v_i) p(u \mid v_i) p(v_i, x_0) = \sum_{i=1,2} p(w \mid x_0, v_i) p(u \mid v_i) p(v_i, x_0), \end{aligned} \tag{A.38}$$

$$p(w, s_1, u, x_0) = \sum_{i=1}^{4} p(s_1 | w, u, x_0, v_i) p(w \mid x_0, v_i) p(u \mid v_i) p(v_i, x_0) = \sum_{i=3,4} p(w \mid x_0, v_i) p(u \mid v_i) p(v_i, x_0), \tag{A.39}$$

$$p(w, s_0, u, x_1) = \sum_{i=1}^{4} p(s_0 | w, u, x_1, v_i) p(w \mid x_1, v_i) p(u \mid v_i) p(v_i, x_1) = \sum_{i=1,3} p(w \mid x_1, v_i) p(u \mid v_i) p(v_i, x_1), \tag{A.40}$$

$$p(w, s_1, u, x_1) = \sum_{i=1}^{4} p(s_1 \mid w, u, x_1, v_i) p(w \mid x_1, v_i) p(u \mid v_i) p(v_i, x_1) = \sum_{i=2,4} p(w \mid x_1, v_i) p(u \mid v_i) p(v_i, x_1) \tag{A.41}$$

for $w \in \{w_0, w_1\}$ and $u \in \{u_1, u_2, u_3, u_4\}$, since

$$\begin{aligned} &p(s_0 \mid w, u, x_0, v_i) = 1 \quad \text{for } i = 1, 2, \quad p(s_0 \mid w, u, x_0, v_i) = 0 \quad \text{for } i = 3, 4, \\ &p(s_1 \mid w, u, x_0, v_i) = 1 \quad \text{for } i = 3, 4, \quad p(s_1 \mid w, u, x_0, v_i) = 0 \quad \text{for } i = 1, 2, \\ &p(s_0 \mid w, u, x_1, v_i) = 1 \quad \text{for } i = 1, 3, \quad p(s_0 \mid w, u, x_1, v_i) = 0 \quad \text{for } i = 2, 4, \\ &p(s_1 \mid w, u, x_1, v_i) = 1 \quad \text{for } i = 2, 4, \quad p(s_1 \mid w, u, x_1, v_i) = 0 \quad \text{for } i = 1, 3 \end{aligned} \tag{A.42}$$

hold. Thus, letting

$$P_{wsux_0} = \begin{pmatrix} p(s, x_0) & p(w, s, x_0) \\ p(s, u, x_0) & p(w, s, u, x_0) \end{pmatrix}, \tag{A.43}$$

$$Q_{s_0ux_0} = \begin{pmatrix} 1 & p(u \mid v_1) \\ 1 & p(u \mid v_2) \end{pmatrix}, \quad \Delta_{s_0x_0} = \begin{pmatrix} p(v_1, x_0) & 0 \\ 0 & p(v_2, x_0) \end{pmatrix}, \quad R_{ws_0x_0} = \begin{pmatrix} 1 & p(w \mid x_0, v_1) \\ 1 & p(w \mid x_0, v_2) \end{pmatrix}, \tag{A.44}$$

$$Q_{s_1ux_0} = \begin{pmatrix} 1 & p(u \mid v_3) \\ 1 & p(u \mid v_4) \end{pmatrix}, \quad \Delta_{s_1x_0} = \begin{pmatrix} p(v_3, x_0) & 0 \\ 0 & p(v_4, x_0) \end{pmatrix}, \quad R_{ws_1x_0} = \begin{pmatrix} 1 & p(w \mid x_0, v_3) \\ 1 & p(w \mid x_0, v_4) \end{pmatrix}, \tag{A.45}$$

since P_{ws_0ux} is invertible from Condition 3 in Theorem 1, from equations (A.38) and (A.39) we obtain

$$P_{wsux_0} = Q_{sux_0}^{\top} \Delta_{sx_0} R_{wsx_0}, \tag{A.46}$$

Thus, we have

$$P_{wsux_0}^{-1} Q_{yux_0}^{\top} = R_{wsx_0}^{-1} \Delta_{yx_0}^{-1} \tag{A.47}$$

for $u \in \{u_1, u_2, u_3, u_4\}$, that is,

$$P^{-1}_{wsux_0} Q^{\top}_{yux_0} = P^{-1}_{wsu'x_0} Q^{\top}_{yu'x_0}, \tag{A.48}$$

specifically,

$$\frac{1}{\det(P_{ws_0ux_0})} \begin{pmatrix} p(w, s_0, u, x_0) & -p(w, s_0, x_0) \\ -p(s_0, u, x_0) & p(s_0, x_0) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ p(u \mid v_1) & p(u \mid v_2) \end{pmatrix} = \frac{1}{\det(P_{ws_0u'x_0})} \begin{pmatrix} p(w, s_0, u', x_0) & -p(w, s_0, x_0) \\ -p(s_0, u', x_0) & p(s_0, x_0) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ p(u' \mid v_1) & p(u' \mid v_2) \end{pmatrix}, \tag{A.49}$$

$$\frac{1}{\det(P_{ws_1ux_0})} \begin{pmatrix} p(w, s_1, u, x_0) & -p(w, s_1, x_0) \\ -p(s_1, u, x_0) & p(s_1, x_0) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ p(u \mid v_3) & p(u \mid v_4) \end{pmatrix} = \frac{1}{\det(P_{ws_1u'x_0})} \begin{pmatrix} p(w, s_1, u', x_0) & -p(w, s_1, x_0) \\ -p(s_1, u', x_0) & p(s_1, x_0) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ p(u' \mid v_3) & p(u' \mid v_4) \end{pmatrix} \tag{A.50}$$

hold for $u \neq u'$ $(u, u' \in \{u_1, u_2, u_3, u_4\})$. Thus, we derive

$$p(s_0, x_0)p(u \mid v_1) - p(s_0, u, x_0) = \frac{\det(P_{ws_0ux_0})}{\det(P_{ws_0u'x_0})}(p(s_0, x_0)p(u' \mid v_1) - p(s_0, u', x_0)), \tag{A.51}$$

$$p(s_0, x_0)p(u \mid v_2) - p(s_0, u, x_0) = \frac{\det(P_{ws_0ux_0})}{\det(P_{ws_0u'x_0})}(p(s_1, x_0)p(u' \mid v_2) - p(s_0, u', x_0)), \tag{A.52}$$

$$p(s_1, x_0)p(u \mid v_3) - p(s_1, u, x_0) = \frac{\det(P_{ws_1ux_0})}{\det(P_{ws_1u'x_0})}(p(s_1, x_0)p(u' \mid v_3) - p(s_1, u', x_0)), \tag{A.53}$$

$$p(s_1, x_0)p(u \mid v_4) - p(s_1, u, x_0) = \frac{\det(P_{ws_1ux_0})}{\det(P_{ws_1u'x_0})}(p(s_1, x_0)p(u' \mid v_4) - p(s_1, u', x_0)). \tag{A.54}$$

Similarly, letting

$$P_{wsux_1} = \begin{pmatrix} p(s, x_1) & p(w, s, x_1) \\ p(s, u, x_1) & p(w, s, u, x_1) \end{pmatrix}, \tag{A.55}$$

$$Q_{s_0ux_1} = \begin{pmatrix} 1 & p(u \mid v_1) \\ 1 & p(u \mid v_3) \end{pmatrix}, \quad \Delta_{s_0x_1} = \begin{pmatrix} p(v_1, x_1) & 0 \\ 0 & p(v_3, x_1) \end{pmatrix}, \quad R_{ws_0x_1} = \begin{pmatrix} 1 & p(w \mid x_1, v_1) \\ 1 & p(w \mid x_1, v_3) \end{pmatrix}, \tag{A.56}$$

$$Q_{s_1ux_1} = \begin{pmatrix} 1 & p(u \mid v_2) \\ 1 & p(u \mid v_4) \end{pmatrix}, \quad \Delta_{s_1x_1} = \begin{pmatrix} p(v_2, x_1) & 0 \\ 0 & p(v_4, x_1) \end{pmatrix}, \quad R_{ws_1x_1} = \begin{pmatrix} 1 & p(w \mid x_1, v_2) \\ 1 & p(w \mid x_1, v_4) \end{pmatrix}, \tag{A.57}$$

since P_{wsus} is invertible from Condition 3 in Theorem 2, from equations (A.40) and (A.41) we obtain

$$P_{wsux_1} = Q^{\top}_{sux_1} \Delta_{sx_1} R_{wsx_1}. \tag{A.58}$$

Thus, we have

$$P^{-1}_{wsux_1} Q^{\top}_{sux_1} = R^{-1}_{wsx_1} \Delta^{-1}_{sx_1} \tag{A.59}$$

for $z \in \{z_1, z_2, z_3\}$, that is,

$$P^{-1}_{wsux_1} Q^{\top}_{sux_1} = P^{-1}_{wsu'x_1} Q^{\top}_{su'x_1}, \tag{A.60}$$

specifically,

$$\frac{1}{\det(P_{ws_0ux_1})} \begin{pmatrix} p(w, s_0, u, x_1) & -p(w, s_0, x_1) \\ -p(s_0, u, x_1) & p(s_0, x_1) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ p(u \mid v_1) & p(u \mid v_3) \end{pmatrix} = \frac{1}{\det(P_{ws_0u'x_1})} \begin{pmatrix} p(w, s_0, u', x_1) & -p(w, s_0, x_1) \\ -p(s_0, u', x_1) & p(s_0, x_1) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ p(u' \mid v_1) & p(u' \mid v_3) \end{pmatrix}, \tag{A.61}$$

$$\frac{1}{\det(P_{ws_1ux_1})} \begin{pmatrix} p(w, s_1, u, x_1) & -p(w, s_1, x_1) \\ -p(s_1, u, x_1) & p(s_1, x_1) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ p(u \mid v_2) & p(u \mid v_4) \end{pmatrix} = \frac{1}{\det(P_{ws_1u'x_1})} \begin{pmatrix} p(w, s_1, u', x_1) & -p(w, s_1, x_1) \\ -p(s_1, u', x_1) & p(s_1, x_1) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ p(u' \mid v_2) & p(u' \mid v_4) \end{pmatrix} \tag{A.62}$$

hold for $u \neq u'$ $(u, u' \in \{u_1, u_2, u_3, u_4\})$. Thus, we derive

$$p(s_0, x_1)p(u \,|\, v_1) - p(s_0, u, x_1) = \frac{\det(P_{ws_0ux_1})}{\det(P_{ws_0u'x_1})}(p(s_0, x_1)p(u' \,|\, v_1) - p(s_0, u', x_1)), \tag{A.63}$$

$$p(s_1, x_1)p(u \,|\, v_2) - p(s_1, u, x_1) = \frac{\det(P_{ws_1ux_1})}{\det(P_{ws_1u'x_1})}(p(s_1, x_1)p(u' \,|\, v_2) - p(s_1, u', x_1)), \tag{A.64}$$

$$p(s_0, x_1)p(u \,|\, v_3) - p(s_0, u, x_1) = \frac{\det(P_{ws_0ux_1})}{\det(P_{ws_0u'x_1})}(p(s_0, x_1)p(u' \,|\, v_3) - p(s_0, u', x_1)), \tag{A.65}$$

$$p(s_1, x_1)p(u \,|\, v_4) - p(s_1, u, x_1) = \frac{\det(P_{ws_1ux_1})}{\det(P_{ws_1u'x_1})}(p(s_1, x_1)p(u' \,|\, v_4) - p(s_1, u', x_1)). \tag{A.66}$$

From Condition 3 in Theorem 2, a system of eight linear equations (A.51)-(A.54) and (A.63)-(A.66) provides a unique solution regarding $p(u \,|\, v)$ for $u, u' \in \{u_1, u_2, u_3, u_4\}$ $(u \neq u')$. Indeed, if Condition 3 in Theorem 2 holds, we have

$$\begin{pmatrix} p(u \,|\, v_1) \\ p(u' \,|\, v_1) \end{pmatrix} = \begin{pmatrix} 1 & -\dfrac{\det(P_{ws_0ux_0})}{\det(P_{ws_0u'x_0})} \\ 1 & -\dfrac{\det(P_{ws_0ux_1})}{\det(P_{ws_0u'x_1})} \end{pmatrix}^{-1} \begin{pmatrix} p(u \,|\, s_0, x_0) - \dfrac{\det(P_{ws_0ux_0})}{\det(P_{ws_0u'x_0})}p(u' \,|\, s_0, x_0) \\ p(u \,|\, s_0, x_1) - \dfrac{\det(P_{ws_0ux_1})}{\det(P_{ws_0u'x_1})}p(u' \,|\, s_0, x_1) \end{pmatrix}, \tag{A.67}$$

$$\begin{pmatrix} p(u \,|\, v_2) \\ p(u' \,|\, v_2) \end{pmatrix} = \begin{pmatrix} 1 & -\dfrac{\det(P_{ws_0ux_0})}{\det(P_{ws_0u'x_0})} \\ 1 & -\dfrac{\det(P_{ws_1ux_1})}{\det(P_{ws_1u'x_1})} \end{pmatrix}^{-1} \begin{pmatrix} p(u \,|\, s_0, x_0) - \dfrac{\det(P_{ws_0ux_0})}{\det(P_{ws_0u'x_0})}p(u' \,|\, s_0, x_0) \\ p(u \,|\, s_1, x_1) - \dfrac{\det(P_{ws_1ux_1})}{\det(P_{ws_1u'x_1})}p(u' \,|\, s_1, x_1) \end{pmatrix}, \tag{A.68}$$

$$\begin{pmatrix} p(u \,|\, v_3) \\ p(u' \,|\, v_3) \end{pmatrix} = \begin{pmatrix} 1 & -\dfrac{\det(P_{ws_1ux_0})}{\det(P_{ws_1u'x_0})} \\ 1 & -\dfrac{\det(P_{ws_0ux_1})}{\det(P_{ws_0u'x_1})} \end{pmatrix}^{-1} \begin{pmatrix} p(u \,|\, s_1, x_0) - \dfrac{\det(P_{ws_1ux_0})}{\det(P_{ws_1u'x_0})}p(u' \,|\, s_1, x_0) \\ p(u \,|\, s_0, x_1) - \dfrac{\det(P_{ws_0ux_1})}{\det(P_{ws_0u'x_1})}p(u' \,|\, s_0, x_1) \end{pmatrix}, \tag{A.69}$$

$$\begin{pmatrix} p(u \,|\, v_4) \\ p(u' \,|\, v_4) \end{pmatrix} = \begin{pmatrix} 1 & -\dfrac{\det(P_{ws_1ux_0})}{\det(P_{ws_1u'x_0})} \\ 1 & -\dfrac{\det(P_{ws_1ux_1})}{\det(P_{ws_1u'x_1})} \end{pmatrix}^{-1} \begin{pmatrix} p(u \,|\, s_1, x_0) - \dfrac{\det(P_{ws_1ux_0})}{\det(P_{ws_1u'x_0})}p(u' \,|\, s_1, x_0) \\ p(u \,|\, s_1, x_1) - \dfrac{\det(P_{ws_1ux_1})}{\det(P_{ws_1u'x_1})}p(u' \,|\, s_1, x_1) \end{pmatrix}. \tag{A.70}$$

Therefore, noting that $Q_{sux}^{\top}$ is identifiable and invertible and P_{wsux} is available for $x \in \{x_0, x_1\}$, $s \in \{s_0, s_1\}$, $s \in \{x_0, x_1\}$, and $u \in \{u_1, u_2, u_3, u_4\}$, we derive

$$Q_{s_0ux_0}^{-\top}P_{ws_0ux_0} = \Delta_{s_0x_0}R_{ws_0x_0} = \begin{pmatrix} p(v_1, x_0) & p(w, x_0, v_1) \\ p(v_2, x_0) & p(w, x_0, v_2) \end{pmatrix}, \tag{A.71}$$

$$Q_{s_1ux_0}^{-\top}P_{ws_1ux_0} = \Delta_{s_1x_0}R_{ws_1x_0} = \begin{pmatrix} p(v_3, x_0) & p(w, x_0, v_3) \\ p(v_4, x_0) & p(w, x_0, v_4) \end{pmatrix}, \tag{A.72}$$

$$Q_{s_0ux_1}^{-\top}P_{ws_0ux_1} = \Delta_{s_0x_1}R_{ws_0x_1} = \begin{pmatrix} p(v_1, x_1) & p(w, x_1, v_1) \\ p(v_3, x_1) & p(w, x_1, v_3) \end{pmatrix}, \tag{A.73}$$

$$Q_{s_1ux_1}^{-\top}P_{ws_1ux_1} = \Delta_{s_1x_1}R_{ws_1x_1} = \begin{pmatrix} p(v_2, x_1) & p(w, x_1, v_2) \\ p(v_4, x_1) & p(w, x_1, v_4) \end{pmatrix}. \tag{A.74}$$

Hence, since $p(u \,|\, w, x, v) = p(u \,|\, v)$ by Condition 2 in Theorem 2,

$$\begin{aligned} p(w, s, x, u, v) &= p(s \,|\, w, u, x, v)p(w, u, x, v) \\ &= p(s \,|\, w, u, x, v)p(u \,|\, w, x, v)p(w, x, v) = p(s \,|\, w, u, x, v)p(u \,|\, v)p(w, x, v) \end{aligned} \tag{A.75}$$

are identifiable for $x \in \{x_0, x_1\}$, $s \in \{s_0, s_1\}$, $w \in \{w_1, w_2\}$, $u \in \{v_1, v_2, v_3, v_4\}$ and $v \in \{v_1, v_2, v_3, v_4\}$. Equivalently, by equations (A.42), we have

$$p(w, s_0, x_0, u, v_i) = \begin{cases} p(u \,|\, v_i)p(w, x_0, v_i) & \text{for } i = 1, 2, \\ 0 & \text{for } i = 3, 4, \end{cases}$$

$$p(w,s_1,x_0,u,v_i)=\begin{cases}0 & \text{for } i=1,2,\\ p(u\,|\,v_i)p(w,x_0,v_i) & \text{for } i=3,4,\end{cases}$$
$$p(w,s_0,x_1,u,v_i)=\begin{cases}p(u\,|\,v_i)p(w,x_1,v_i) & \text{for } i=1,3,\\ 0 & \text{for } i=2,4,\end{cases}$$
$$p(w,s_1,x_1,u,v_i)=\begin{cases}0 & \text{for } i=1,3,\\ p(u\,|\,v_i)p(w,x_1,v_i) & \text{for } i=2,4\end{cases}$$

for $w\in\{w_0,w_1\}$ and $u\in\{u_1,u_2,u_3,u_4\}$.

□

B ESTIMATION

When PCs are identifiable through the proposed condition, as seen from the proofs of Theorems 1 and 2 above, the estimation problem is reduced to that of singular models; thus, these probabilities cannot be evaluated by standard statistical estimation methods, such as the maximum likelihood estimation method. To solve this problem, we propose new estimators of PCs according to Algorithm 1.

Let

$$P_{xyzs\cdot w}=\begin{pmatrix} p(y,s\,|\,w) & p(x,y,s\,|\,w)\\ p(y,z,s\,|\,w) & p(x,y,z,s\,|\,w)\end{pmatrix}=\begin{pmatrix} \sum_{x,z}p(x,y,z,s,w)/p(w) & \sum_{z}p(x,y,z,s,w)/p(w)\\ \sum_{x}p(x,y,z,s,w)/p(w) & p(x,y,z,s,w)/p(w)\end{pmatrix},$$

$$A_{1x\cdot w}=\begin{pmatrix} 1 & -\dfrac{\det(P_{xy_0z_1s_0\cdot w})}{\det(P_{xy_0z_2s_0\cdot w})}\\ 1 & -\dfrac{\det(P_{xy_0z_1s_1\cdot w})}{\det(P_{xy_0z_2s_1\cdot w})}\end{pmatrix},\quad A_{2x\cdot w}=\begin{pmatrix} 1 & -\dfrac{\det(P_{xy_0z_1s_0\cdot w})}{\det(P_{xy_0z_2s_0\cdot w})}\\ 1 & -\dfrac{\det(P_{xy_1z_1s_1\cdot w})}{\det(P_{xy_1z_2s_1\cdot w})}\end{pmatrix},$$

$$A_{3x\cdot w}=\begin{pmatrix} 1 & -\dfrac{\det(P_{xy_1z_1s_0\cdot w})}{\det(P_{xy_1z_2s_0\cdot w})}\\ 1 & -\dfrac{\det(P_{xy_0z_1s_1\cdot w})}{\det(P_{xy_0z_2s_1\cdot w})}\end{pmatrix},\quad A_{4x\cdot w}=\begin{pmatrix} 1 & -\dfrac{\det(P_{xy_1z_1s_0\cdot w})}{\det(P_{xy_1z_2s_0\cdot w})}\\ 1 & -\dfrac{\det(P_{xy_1z_1s_1\cdot w})}{\det(P_{xy_1z_2s_1\cdot w})}\end{pmatrix},$$

$$\boldsymbol{b}_{1x\cdot w}=\begin{pmatrix} \dfrac{p(y_0,z_1,s_0\,|\,w)}{p(y_0,s_0\,|\,w)}-\dfrac{\det(P_{xy_0z_1s_0\cdot w})}{\det(P_{xy_0z_2s_0\cdot w})}\dfrac{p(y_0,z_2,s_0\,|\,w)}{p(y_0,s_0\,|\,w)}\\ \dfrac{p(y_0,z_1,s_1\,|\,w)}{p(y_0,s_1\,|\,w)}-\dfrac{\det(P_{xy_0z_1s_1\cdot w})}{\det(P_{xy_0z_2s_1\cdot w})}\dfrac{p(y_0,z_2,s_1\,|\,w)}{p(y_0,s_1\,|\,w)}\end{pmatrix},$$

$$\boldsymbol{b}_{2x\cdot w}=\begin{pmatrix} \dfrac{p(y_0,z_1,s_0\,|\,w)}{p(y_0,s_0\,|\,w)}-\dfrac{\det(P_{xy_0z_1s_0\cdot w})}{\det(P_{xy_0z_2s_0\cdot w})}\dfrac{p(y_0,z_2,s_0\,|\,w)}{p(y_0,s_0\,|\,w)}\\ \dfrac{p(y_1,z_1,s_1\,|\,w)}{p(y_1,s_1\,|\,w)}-\dfrac{\det(P_{xy_1z_1s_1\cdot w})}{\det(P_{xy_1z_2s_1\cdot w})}\dfrac{p(y_1,z_2,s_1\,|\,w)}{p(y_1,s_1\,|\,w)}\end{pmatrix},$$

$$\boldsymbol{b}_{3x\cdot w}=\begin{pmatrix} \dfrac{p(y_1,z_1,s_0\,|\,w)}{p(y_1,s_0\,|\,w)}-\dfrac{\det(P_{xy_1z_1s_0\cdot w})}{\det(P_{xy_1z_2s_0\cdot w})}\dfrac{p(y_1,z_2,s_0\,|\,w)}{p(y_1,s_0\,|\,w)}\\ \dfrac{p(y_0,z_1,s_1\,|\,w)}{p(y_0,s_1\,|\,w)}-\dfrac{\det(P_{xy_0z_1s_1\cdot w})}{\det(P_{xy_0z_2s_1\cdot w})}\dfrac{p(y_0,z_2,s_1\,|\,w)}{p(y_0,s_1\,|\,w)}\end{pmatrix},$$

$$\boldsymbol{b}_{4x\cdot w}=\begin{pmatrix} \dfrac{p(y_1,z_1,s_0\,|\,w)}{p(y_1,s_0\,|\,w)}-\dfrac{\det(P_{xy_1z_1s_0\cdot w})}{\det(P_{xy_1z_2s_0\cdot w})}\dfrac{p(y_1,z_2,s_0\,|\,w)}{p(y_1,s_0\,|\,w)}\\ \dfrac{p(y_1,z_1,s_1\,|\,w)}{p(y_1,s_1\,|\,w)}-\dfrac{\det(P_{xy_1z_1s_1\cdot w})}{\det(P_{xy_1z_2s_1\cdot w})}\dfrac{p(y_1,z_2,s_1\,|\,w)}{p(y_1,s_1\,|\,w)}\end{pmatrix}$$

for $x\in\{x_0,x_1\}$ and $w\in\{w_0,w_1\}$. Then, from equations (A.30)-(A.33) in Section A.1,

$$\begin{pmatrix} p(z_1\,|\,u_1,w)\\ p(z_2\,|\,u_1,w)\end{pmatrix}=A^{-1}_{1x\cdot w}\boldsymbol{b}_{1x\cdot w},\quad \begin{pmatrix} p(z_1\,|\,u_2,w)\\ p(z_2\,|\,u_2,w)\end{pmatrix}=A^{-1}_{2x\cdot w}\boldsymbol{b}_{2x\cdot w},$$
$$\begin{pmatrix} p(z_1\,|\,u_3,w)\\ p(z_2\,|\,u_3,w)\end{pmatrix}=A^{-1}_{3x\cdot w}\boldsymbol{b}_{3x\cdot w},\quad \begin{pmatrix} p(z_1\,|\,u_4,w)\\ p(z_2\,|\,u_4,w)\end{pmatrix}=A^{-1}_{4x\cdot w}\boldsymbol{b}_{4x\cdot w}. \tag{B.76}$$

Consider the matrices $\widehat{P}_{xyzs\cdot w}$ that are derived by replacing $p(y,s\,|\,w)$, $p(x,y,s\,|\,w)$, $p(y,z,s\,|\,w)$, and $p(x,y,z,s\,|\,w)$ of $P_{xyzs\cdot w}$ with the sample probabilities $\widehat{p}(y,s\,|\,w)$, $\widehat{p}(x,y,s\,|\,w)$, $\widehat{p}(y,z,s\,|\,w)$ and $\widehat{p}(x,y,z,s\,|\,w)$, respectively, for $x\in\{x_0,x_1\}$, $y\in\{y_0,y_1\}$, $z\in\{z_1,z_2\}$, $s\in\{s_0,s_1\}$, and $w\in\{w_0,w_1\}$. Additionally, consider $\widehat{A}_{ix\cdot w}$ and $\widehat{\boldsymbol{b}}_{ix\cdot w}$, which are derived by replacing $P_{xyzs\cdot w}$, $p(y,s\,|\,w)$, and $p(y,z,s\,|\,w)$ of $A_{ix\cdot w}$ and $b_{ix\cdot w}$ with the sample probabilities $\widehat{P}_{xyzs\cdot w}$, $\widehat{p}(y,s\,|\,w)$, and $\widehat{p}(y,z,s\,|\,w)$, respectively, for $i=1,2,3,4$. To estimate the probabilities $p(z\,|\,u,w)$, it is natural to replace $A_{ix\cdot w}$ and $b_{ix\cdot w}$ in (B.76) with $\widehat{A}_{ix\cdot w}$ and $\widehat{\boldsymbol{b}}_{ix\cdot w}$, respectively, for $w\in\{w_0,w_1\}$ and $i=1,2,3,4$. However, such estimates of $p(z\,|\,u,w)$ might not lie between 0 and 1 due to sampling error. Thus, using the solutions of the following optimization problems

$$\widehat{\theta}_{z_1u_1w},\widehat{\theta}_{z_2u_1w}=\underset{\theta_{z_1u_1w},\theta_{z_2u_1w}}{\operatorname{argmin}}\sum_{x\in\{x_0,x_1\}}\left\|\widehat{A}_{1x\cdot w}\begin{pmatrix}\dfrac{1}{1+\exp(-\theta_{z_1u_1w})}\\ \dfrac{1}{1+\exp(-\theta_{z_2u_1w})}\end{pmatrix}-\widehat{\boldsymbol{b}}_{1x\cdot w}\right\|_2^2,$$

$$\widehat{\theta}_{z_1u_2w},\widehat{\theta}_{z_2u_2w}=\underset{\theta_{z_1u_2w},\theta_{z_2u_2w}}{\operatorname{argmin}}\sum_{x\in\{x_0,x_1\}}\left\|\widehat{A}_{2x\cdot w}\begin{pmatrix}\dfrac{1}{1+\exp(-\theta_{z_1u_2w})}\\ \dfrac{1}{1+\exp(-\theta_{z_2u_2w})}\end{pmatrix}-\widehat{\boldsymbol{b}}_{2x\cdot w}\right\|_2^2,$$

$$\widehat{\theta}_{z_1u_3w},\widehat{\theta}_{z_2u_3w}=\underset{\theta_{z_1u_3w},\theta_{z_2u_3w}}{\operatorname{argmin}}\sum_{x\in\{x_0,x_1\}}\left\|\widehat{A}_{3x\cdot w}\begin{pmatrix}\dfrac{1}{1+\exp(-\theta_{z_1u_3w})}\\ \dfrac{1}{1+\exp(-\theta_{z_2u_3w})}\end{pmatrix}-\widehat{\boldsymbol{b}}_{3x\cdot w}\right\|_2^2,$$

$$\widehat{\theta}_{z_1u_4w},\widehat{\theta}_{z_2u_4w}=\underset{\theta_{z_1u_4w},\theta_{z_2u_4w}}{\operatorname{argmin}}\sum_{x\in\{x_0,x_1\}}\left\|\widehat{A}_{4x\cdot w}\begin{pmatrix}\dfrac{1}{1+\exp(-\theta_{z_1u_4w})}\\ \dfrac{1}{1+\exp(-\theta_{z_2u_4w})}\end{pmatrix}-\widehat{\boldsymbol{b}}_{4x\cdot w}\right\|_2^2,$$

we obtain consistent estimators $\widehat{p}(z\,|\,u,w)=1/\left(1+\exp(-\widehat{\theta}_{zuw})\right)$ of the probabilities $p(z\,|\,u,w)=1/\left(1+\exp(-\theta_{zuw})\right)$ for $z\in\{z_1,z_2\}$, $u\in\{u_1,u_2,u_3,u_4\}$, and $w\in\{w_0,w_1\}$.

Then, let

$$\widehat{Q}_{y_0z_1s_0\cdot w}=\begin{pmatrix}1 & \widehat{p}(z_1\,|\,u_1,w)\\ 1 & \widehat{p}(z_1\,|\,u_2,w)\end{pmatrix},\quad \widehat{Q}_{y_1z_1s_0\cdot w}=\begin{pmatrix}1 & \widehat{p}(z_1\,|\,u_3,w)\\ 1 & \widehat{p}(z_1\,|\,u_4,w)\end{pmatrix},$$
$$\widehat{Q}_{y_0z_1s_1\cdot w}=\begin{pmatrix}1 & \widehat{p}(z_1\,|\,u_1,w)\\ 1 & \widehat{p}(z_1\,|\,u_3,w)\end{pmatrix},\quad \widehat{Q}_{y_1z_1s_1\cdot w}=\begin{pmatrix}1 & \widehat{p}(z_1\,|\,u_2,w)\\ 1 & \widehat{p}(z_1\,|\,u_4,w)\end{pmatrix},$$
$$\widehat{P}_{xy_0z_1s_0\cdot w}=\begin{pmatrix}\widehat{p}(y_0,s_0\,|\,w) & \widehat{p}(x,y_0,s_0\,|\,w)\\ \widehat{p}(y_0,z_1,s_0\,|\,w) & \widehat{p}(x,y_0,z_1,s_0\,|\,w)\end{pmatrix},\quad \widehat{P}_{xy_1z_1s_0\cdot w}=\begin{pmatrix}\widehat{p}(y_1,s_0\,|\,w) & \widehat{p}(x,y_1,s_0\,|\,w)\\ \widehat{p}(y_1,z_1,s_0\,|\,w) & \widehat{p}(x,y_1,z_1,s_0\,|\,w)\end{pmatrix},$$
$$\widehat{P}_{xy_0z_1s_1\cdot w}=\begin{pmatrix}\widehat{p}(y_0,s_1\,|\,w) & \widehat{p}(x,y_0,s_1\,|\,w)\\ \widehat{p}(y_0,z_1,s_1\,|\,w) & \widehat{p}(x,y_0,z_1,s_1\,|\,w)\end{pmatrix},\quad \widehat{P}_{xy_1z_1s_1\cdot w}=\begin{pmatrix}\widehat{p}(y_1,s_1\,|\,w) & \widehat{p}(x,y_1,s_1\,|\,w)\\ \widehat{p}(y_1,z_1,s_1\,|\,w) & \widehat{p}(x,y_1,z_1,s_1\,|\,w)\end{pmatrix}$$

for $w\in\{w_0,w_1\}$. Using the solutions of the following optimization problems

$$\widehat{\theta}_{xs_0u_1\cdot w},\widehat{\theta}_{xs_0u_2\cdot w}=\underset{\theta_{xs_0u_1\cdot w},\theta_{xs_0u_2\cdot w}}{\operatorname{argmin}}\sum_{z\in\{z_1,z_2\}}\left\|\widehat{Q}^{\top}_{y_0zs_0\cdot w}\begin{pmatrix}\dfrac{1}{1+\exp(-\theta_{xs_0u_1\cdot w})}\\ \dfrac{1}{1+\exp(-\theta_{xs_0u_2\cdot w})}\end{pmatrix}-\widehat{P}_{xy_0zs_0\cdot w}\boldsymbol{e}_2\right\|_2^2,$$

$$\widehat{\theta}_{xs_0u_3\cdot w},\widehat{\theta}_{xs_0u_4\cdot w}=\underset{\theta_{xs_0u_3\cdot w},\theta_{xs_0u_4\cdot w}}{\operatorname{argmin}}\sum_{z\in\{z_1,z_2\}}\left\|\widehat{Q}^{\top}_{y_1zs_0\cdot w}\begin{pmatrix}\dfrac{1}{1+\exp(-\theta_{xs_0u_3\cdot w})}\\ \dfrac{1}{1+\exp(-\theta_{xs_0u_4\cdot w})}\end{pmatrix}-\widehat{P}_{xy_1zs_0\cdot w}\boldsymbol{e}_2\right\|_2^2,$$

$$\widehat{\theta}_{xs_1u_1\cdot w},\widehat{\theta}_{xs_1u_3\cdot w}=\underset{\theta_{xs_1u_1\cdot w},\theta_{xs_1u_3\cdot w}}{\operatorname{argmin}}\sum_{z\in\{z_1,z_2\}}\left\|\widehat{Q}^{\top}_{y_0zs_1\cdot w}\begin{pmatrix}\dfrac{1}{1+\exp(-\theta_{xs_1u_1\cdot w})}\\ \dfrac{1}{1+\exp(-\theta_{xs_1u_3\cdot w})}\end{pmatrix}-\widehat{P}_{xy_0zs_1\cdot w}\boldsymbol{e}_2\right\|_2^2,$$

$$\widehat{\theta}_{xs_1u_2\cdot w}, \widehat{\theta}_{xs_1u_4\cdot w} = \underset{\theta_{xs_1u_2\cdot w},\theta_{xs_1u_4\cdot w}}{\text{argmin}} \sum_{z\in\{z_1,z_2\}} \left\| \widehat{Q}^{\top}_{y_1zs_1\cdot w} \begin{pmatrix} \dfrac{1}{1+\exp(-\theta_{xs_1u_2\cdot w})} \\ \dfrac{1}{1+\exp(-\theta_{xs_1u_4\cdot w})} \end{pmatrix} - \widehat{P}_{xy_1zs_1\cdot w}\boldsymbol{e}_2 \right\|_2^2,$$

we obtain consistent estimators $\widehat{p}(x,s,u\,|\,w) = 1/\left(1+\exp(-\widehat{\theta}_{xsu\cdot w})\right)$ of the probabilities $p(x,s,u\,|\,w) = 1/\left(1+\exp(-\theta_{xsu\cdot w})\right)$ for $x\in\{x_0,x_1\}$, $s\in\{s_0,s_1\}$, $u\in\{u_1,u_2,u_3,u_4\}$, and $w\in\{w_0,w_1\}$.

Next, consider the matrices $\widehat{P}_{wsux}$ that are derived by replacing $p(s,x)$ and $p(w,s,x)$ of P_{wsux} with the sample probabilities $\widehat{p}(s,x)$ and $\widehat{p}(w,s,x)$, and $p(x,s,u) = \sum_w p(x,s,u\,|\,w)p(w)$ and $p(x,s,u,w) = p(x,s,u\,|\,w)p(w)$ of P_{wsux} with $\widehat{p}(x,s,u) = \sum_w \widehat{p}(x,s,u\,|\,w)\widehat{p}(w)$ and $\widehat{p}(x,s,u,w) = \widehat{p}(x,s,u\,|\,w)\widehat{p}(w)$ for $w\in\{w_0,w_1\}$, $s\in\{s_0,s_1\}$, $u\in\{u_1,u_2,u_3,u_4\}$, and $x\in\{x_0,x_1\}$. Let

$$P_{wsux} = \begin{pmatrix} p(s,x) & p(w,s,x) \\ p(s,u,x) & p(w,s,u,x) \end{pmatrix} = \begin{pmatrix} \sum_{y,z,u,w} p(x,y,z,s,u\,|\,w)p(w) & \sum_{y,z,u} p(x,y,z,s,u\,|\,w)p(w) \\ \sum_{y,z,w} p(x,y,z,s,u\,|\,w)p(w) & \sum_{y,z} p(x,y,z,s,u\,|\,w)p(w) \end{pmatrix},$$

$$A_{11w} = \begin{pmatrix} 1 & -\dfrac{\det(P_{ws_0u_1x_0})}{\det(P_{ws_0u_2x_0})} \\ 1 & -\dfrac{\det(P_{ws_0u_1x_1})}{\det(P_{ws_0u_2x_1})} \end{pmatrix}, \quad A_{12w} = \begin{pmatrix} 1 & -\dfrac{\det(P_{ws_0u_3x_0})}{\det(P_{ws_0u_4x_0})} \\ 1 & -\dfrac{\det(P_{ws_0u_3x_1})}{\det(P_{ws_0u_4x_1})} \end{pmatrix},$$

$$A_{21w} = \begin{pmatrix} 1 & -\dfrac{\det(P_{ws_0u_1x_0})}{\det(P_{ws_0u_2x_0})} \\ 1 & -\dfrac{\det(P_{ws_1u_1x_1})}{\det(P_{ws_1u_2x_1})} \end{pmatrix}, \quad A_{22w} = \begin{pmatrix} 1 & -\dfrac{\det(P_{ws_0u_3x_0})}{\det(P_{ws_0u_4x_0})} \\ 1 & -\dfrac{\det(P_{ws_1u_3x_1})}{\det(P_{ws_1u_4x_1})} \end{pmatrix},$$

$$A_{31w} = \begin{pmatrix} 1 & -\dfrac{\det(P_{ws_1u_1x_0})}{\det(P_{ws_1u_2x_0})} \\ 1 & -\dfrac{\det(P_{ws_0u_1x_1})}{\det(P_{ws_0u_2x_1})} \end{pmatrix}, \quad A_{32w} = \begin{pmatrix} 1 & -\dfrac{\det(P_{ws_1u_3x_0})}{\det(P_{ws_1u_4x_0})} \\ 1 & -\dfrac{\det(P_{ws_0u_3x_1})}{\det(P_{ws_0u_4x_1})} \end{pmatrix},$$

$$A_{41w} = \begin{pmatrix} 1 & -\dfrac{\det(P_{ws_1u_1x_0})}{\det(P_{ws_1u_2x_0})} \\ 1 & -\dfrac{\det(P_{ws_1u_1x_1})}{\det(P_{ws_1u_2x_1})} \end{pmatrix}, \quad A_{42w} = \begin{pmatrix} 1 & -\dfrac{\det(P_{ws_1u_3x_0})}{\det(P_{ws_1u_4x_0})} \\ 1 & -\dfrac{\det(P_{ws_1u_3x_1})}{\det(P_{ws_1u_4x_1})} \end{pmatrix},$$

$$\boldsymbol{b}_{11w} = \begin{pmatrix} \dfrac{p(s_0,u_1,x_0)}{p(s_0,x_0)} - \dfrac{\det(P_{ws_0u_1x_0})}{\det(P_{ws_0u_2x_0})}\dfrac{p(s_0,u_2,x_0)}{p(s_0,x_0)} \\ \dfrac{p(s_0,u_1,x_1)}{p(s_0,x_1)} - \dfrac{\det(P_{ws_0u_1x_1})}{\det(P_{ws_0u_2x_1})}\dfrac{p(s_0,u_2,x_1)}{p(s_0,x_1)} \end{pmatrix}, \quad \boldsymbol{b}_{12w} = \begin{pmatrix} \dfrac{p(s_0,u_3,x_0)}{p(s_0,x_0)} - \dfrac{\det(P_{ws_0u_3x_0})}{\det(P_{ws_0u_4x_0})}\dfrac{p(s_0,u_4,x_0)}{p(s_0,x_0)} \\ \dfrac{p(s_0,u_3,x_1)}{p(s_0,x_1)} - \dfrac{\det(P_{ws_0u_3x_1})}{\det(P_{ws_0u_4x_1})}\dfrac{p(s_0,u_4,x_1)}{p(s_0,x_1)} \end{pmatrix},$$

$$\boldsymbol{b}_{21w} = \begin{pmatrix} \dfrac{p(s_0,u_1,x_0)}{p(s_0,x_0)} - \dfrac{\det(P_{ws_0u_1x_0})}{\det(P_{ws_0u_2x_0})}\dfrac{p(s_0,u_2,x_0)}{p(s_0,x_0)} \\ \dfrac{p(s_1,u_1,x_1)}{p(s_1,x_1)} - \dfrac{\det(P_{ws_1u_1x_1})}{\det(P_{ws_1u_2x_1})}\dfrac{p(s_1,u_2,x_1)}{p(s_1,x_1)} \end{pmatrix}, \quad \boldsymbol{b}_{22w} = \begin{pmatrix} \dfrac{p(s_0,u_3,x_0)}{p(s_0,x_0)} - \dfrac{\det(P_{ws_0u_3x_0})}{\det(P_{ws_0u_4x_0})}\dfrac{p(s_0,u_4,x_0)}{p(s_0,x_0)} \\ \dfrac{p(s_1,u_3,x_1)}{p(s_1,x_1)} - \dfrac{\det(P_{ws_1u_3x_1})}{\det(P_{ws_1u_4x_1})}\dfrac{p(s_1,u_4,x_1)}{p(s_1,x_1)} \end{pmatrix},$$

$$\boldsymbol{b}_{31w} = \begin{pmatrix} \dfrac{p(s_1,u_1,x_0)}{p(s_1,x_0)} - \dfrac{\det(P_{ws_1u_1x_0})}{\det(P_{ws_1u_2x_0})}\dfrac{p(s_1,u_2,x_0)}{p(s_1,x_0)} \\ \dfrac{p(s_0,u_1,x_1)}{p(s_0,x_1)} - \dfrac{\det(P_{ws_0u_1x_1})}{\det(P_{ws_0u_2x_1})}\dfrac{p(s_0,u_2,x_1)}{p(s_0,x_1)} \end{pmatrix}, \quad \boldsymbol{b}_{32w} = \begin{pmatrix} \dfrac{p(s_1,u_3,x_0)}{p(s_1,x_0)} - \dfrac{\det(P_{ws_1u_3x_0})}{\det(P_{ws_1u_4x_0})}\dfrac{p(s_1,u_4,x_0)}{p(s_1,x_0)} \\ \dfrac{p(s_0,u_3,x_1)}{p(s_0,x_1)} - \dfrac{\det(P_{ws_0u_3x_1})}{\det(P_{ws_0u_4x_1})}\dfrac{p(s_0,u_4,x_1)}{p(s_0,x_1)} \end{pmatrix},$$

$$\boldsymbol{b}_{41w} = \begin{pmatrix} \dfrac{p(s_1,u_1,x_0)}{p(s_1,x_0)} - \dfrac{\det(P_{ws_1u_1x_0})}{\det(P_{ws_1u_2x_0})}\dfrac{p(s_1,u_2,x_0)}{p(s_1,x_0)} \\ \dfrac{p(s_1,u_1,x_1)}{p(s_1,x_1)} - \dfrac{\det(P_{ws_1u_1x_1})}{\det(P_{ws_1u_2x_1})}\dfrac{p(s_1,u_2,x_1)}{p(s_1,x_1)} \end{pmatrix}, \quad \boldsymbol{b}_{42w} = \begin{pmatrix} \dfrac{p(s_1,u_3,x_0)}{p(s_1,x_0)} - \dfrac{\det(P_{ws_1u_3x_0})}{\det(P_{ws_1u_4x_0})}\dfrac{p(s_1,u_4,x_0)}{p(s_1,x_0)} \\ \dfrac{p(s_1,u_3,x_1)}{p(s_1,x_1)} - \dfrac{\det(P_{ws_1u_3x_1})}{\det(P_{ws_1u_4x_1})}\dfrac{p(s_1,u_4,x_1)}{p(s_1,x_1)} \end{pmatrix}.$$

Then, from the discussion of Section A.2,

$$\begin{pmatrix} p(u_1\,|\,v_1) \\ p(u_2\,|\,v_1) \end{pmatrix} = A_{11w}^{-1}\boldsymbol{b}_{11w}, \; \begin{pmatrix} p(u_3\,|\,v_1) \\ p(u_4\,|\,v_1) \end{pmatrix} = A_{12w}^{-1}\boldsymbol{b}_{12w}, \; \begin{pmatrix} p(u_1\,|\,v_2) \\ p(u_2\,|\,v_2) \end{pmatrix} = A_{21w}^{-1}\boldsymbol{b}_{21w}, \; \begin{pmatrix} p(u_3\,|\,v_2) \\ p(u_4\,|\,v_2) \end{pmatrix} = A_{22w}^{-1}\boldsymbol{b}_{22w},$$
$$\begin{pmatrix} p(u_1\,|\,v_3) \\ p(u_2\,|\,v_3) \end{pmatrix} = A_{31w}^{-1}\boldsymbol{b}_{31w}, \; \begin{pmatrix} p(u_3\,|\,v_3) \\ p(u_4\,|\,v_3) \end{pmatrix} = A_{32w}^{-1}\boldsymbol{b}_{32w}, \; \begin{pmatrix} p(u_1\,|\,v_4) \\ p(u_2\,|\,v_4) \end{pmatrix} = A_{41w}^{-1}\boldsymbol{b}_{41w}, \; \begin{pmatrix} p(u_3\,|\,v_4) \\ p(u_4\,|\,v_4) \end{pmatrix} = A_{42w}^{-1}\boldsymbol{b}_{42w}. \tag{B.77}$$

To estimate $p(u \mid v)$, consider $\widehat{A}_{ijw}$ and $\widehat{\boldsymbol{b}}_{ijw}$, which are derived by replacing P_{wsux}, $p(w,s,x)$, and $p(w,s,u,x)$ of A_{ijw} and b_{ijw} with the sample probabilities $\widehat{P}_{wsux}$, $\widehat{p}(y,s \mid w)$, and $\widehat{p}(y,z,s \mid w)$, respectively, for $i = 1,2,3,4$ and $j = 1,2$.

From equations (A.67)-(A.70) in Supplementary Material A.2, using the solutions of the following optimization problems

$$\widehat{\theta}_{u_1v_1}, \widehat{\theta}_{u_2v_1} = \underset{\theta_{u_1v_1},\theta_{u_2v_1}}{\operatorname{argmin}} \sum_{w \in \{w_0,w_1\}} \left\| \widehat{A}_{11w} \begin{pmatrix} \frac{1}{1+\exp(-\theta_{u_1v_1})} \\ \frac{1}{1+\exp(-\theta_{u_2v_1})} \end{pmatrix} - \widehat{\boldsymbol{b}}_{11w} \right\|_2^2,$$

$$\widehat{\theta}_{u_3v_1}, \widehat{\theta}_{u_4v_1} = \underset{\theta_{u_3v_1},\theta_{u_4v_1}}{\operatorname{argmin}} \sum_{w \in \{w_0,w_1\}} \left\| \widehat{A}_{12w} \begin{pmatrix} \frac{1}{1+\exp(-\theta_{u_3v_1})} \\ \frac{1}{1+\exp(-\theta_{u_4v_1})} \end{pmatrix} - \widehat{\boldsymbol{b}}_{12w} \right\|_2^2,$$

$$\widehat{\theta}_{u_1v_2}, \widehat{\theta}_{u_2v_2} = \underset{\theta_{u_1v_2},\theta_{u_2v_2}}{\operatorname{argmin}} \sum_{w \in \{w_0,w_1\}} \left\| \widehat{A}_{21w} \begin{pmatrix} \frac{1}{1+\exp(-\theta_{u_1v_2})} \\ \frac{1}{1+\exp(-\theta_{u_2v_2})} \end{pmatrix} - \widehat{\boldsymbol{b}}_{21w} \right\|_2^2,$$

$$\widehat{\theta}_{u_3v_2}, \widehat{\theta}_{u_4v_2} = \underset{\theta_{u_3v_2},\theta_{u_4v_2}}{\operatorname{argmin}} \sum_{w \in \{w_0,w_1\}} \left\| \widehat{A}_{22w} \begin{pmatrix} \frac{1}{1+\exp(-\theta_{u_3v_2})} \\ \frac{1}{1+\exp(-\theta_{u_4v_2})} \end{pmatrix} - \widehat{\boldsymbol{b}}_{22w} \right\|_2^2,$$

$$\widehat{\theta}_{u_1v_3}, \widehat{\theta}_{u_2v_3} = \underset{\theta_{u_1v_3},\theta_{u_2v_3}}{\operatorname{argmin}} \sum_{w \in \{w_0,w_1\}} \left\| \widehat{A}_{31w} \begin{pmatrix} \frac{1}{1+\exp(-\theta_{u_1v_3})} \\ \frac{1}{1+\exp(-\theta_{u_2v_3})} \end{pmatrix} - \widehat{\boldsymbol{b}}_{31w} \right\|_2^2,$$

$$\widehat{\theta}_{u_3v_3}, \widehat{\theta}_{u_4v_3} = \underset{\theta_{u_3v_3},\theta_{u_4v_3}}{\operatorname{argmin}} \sum_{w \in \{w_0,w_1\}} \left\| \widehat{A}_{32w} \begin{pmatrix} \frac{1}{1+\exp(-\theta_{u_3v_3})} \\ \frac{1}{1+\exp(-\theta_{u_4v_3})} \end{pmatrix} - \widehat{\boldsymbol{b}}_{32w} \right\|_2^2,$$

$$\widehat{\theta}_{u_1v_4}, \widehat{\theta}_{u_2v_4} = \underset{\theta_{u_1v_4},\theta_{u_2v_4}}{\operatorname{argmin}} \sum_{w \in \{w_0,w_1\}} \left\| \widehat{A}_{41w} \begin{pmatrix} \frac{1}{1+\exp(-\theta_{u_1v_4})} \\ \frac{1}{1+\exp(-\theta_{u_2v_4})} \end{pmatrix} - \widehat{\boldsymbol{b}}_{41w} \right\|_2^2,$$

$$\widehat{\theta}_{u_3v_4}, \widehat{\theta}_{u_4v_4} = \underset{\theta_{u_3v_4},\theta_{u_4v_4}}{\operatorname{argmin}} \sum_{w \in \{w_0,w_1\}} \left\| \widehat{A}_{42w} \begin{pmatrix} \frac{1}{1+\exp(-\theta_{u_3v_4})} \\ \frac{1}{1+\exp(-\theta_{u_4v_4})} \end{pmatrix} - \widehat{\boldsymbol{b}}_{42w} \right\|_2^2,$$

we have consistent estimators $\widehat{p}(u \mid v) = 1/\left(1+\exp(-\widehat{\theta}_{uv})\right)$ of the probabilities $p(u \mid v) = 1/\left(1+\exp(-\theta_{uv})\right)$ for $u \in \{u_1,u_2,u_3,u_4\}$ and $v \in \{v_1,v_2,v_3,v_4\}$.

Let

$$\widehat{Q}_{s_0ux_0} = \begin{pmatrix} 1 & \widehat{p}(u \mid v_1) \\ 1 & \widehat{p}(u \mid v_2) \end{pmatrix}, \quad \widehat{Q}_{s_1ux_0} = \begin{pmatrix} 1 & \widehat{p}(u \mid v_3) \\ 1 & \widehat{p}(u \mid v_4) \end{pmatrix},$$

$$\widehat{Q}_{s_0ux_1} = \begin{pmatrix} 1 & \widehat{p}(u \mid v_1) \\ 1 & \widehat{p}(u \mid v_3) \end{pmatrix}, \quad \widehat{Q}_{s_1ux_1} = \begin{pmatrix} 1 & \widehat{p}(u \mid v_2) \\ 1 & \widehat{p}(u \mid v_4) \end{pmatrix},$$

$$\widehat{P}_{ws_0ux_0} = \begin{pmatrix} \widehat{p}(s_0,x_0) & \widehat{p}(w,s_0,x_0) \\ \widehat{p}(s_0,u,x_0) & \widehat{p}(w,s_0,u,x_0) \end{pmatrix}, \quad \widehat{P}_{ws_1ux_0} = \begin{pmatrix} \widehat{p}(s_1,x_0) & \widehat{p}(w,s_1,x_0) \\ \widehat{p}(s_1,u,x_0) & \widehat{p}(w,s_1,u,x_0) \end{pmatrix},$$

$$\widehat{P}_{ws_0ux_1} = \begin{pmatrix} \widehat{p}(s_0,x_1) & \widehat{p}(w,s_0,x_1) \\ \widehat{p}(s_0,u,x_1) & \widehat{p}(w,s_0,u,x_1) \end{pmatrix}, \quad \widehat{P}_{ws_1ux_1} = \begin{pmatrix} \widehat{p}(s_1,x_1) & \widehat{p}(w,s_1,x_1) \\ \widehat{p}(s_1,u,x_1) & \widehat{p}(w,s_1,u,x_1) \end{pmatrix}$$

for $u \in \{u_1, u_2, u_3, u_4\}$. Using the solutions of the following optimization problems

$$\widehat{\theta}_{x_0v_1}, \widehat{\theta}_{x_0v_2} = \underset{\theta_{x_0v_1}, \theta_{x_0v_2}}{\text{argmin}} \sum_{w\in\{w_0,w_1\}} \sum_{u\in\{u_1,u_2,u_3,u_4\}} \left\| \widehat{Q}^{\top}_{s_0ux_0} \begin{pmatrix} \dfrac{1}{1+\exp(-\theta_{x_0v_1})} \\ \dfrac{1}{1+\exp(-\theta_{x_0v_2})} \end{pmatrix} - \widehat{P}_{ws_0ux_0}\boldsymbol{e}_1 \right\|_2^2,$$

$$\widehat{\theta}_{x_0v_3}, \widehat{\theta}_{x_0v_4} = \underset{\theta_{x_0v_3}, \theta_{x_0v_4}}{\text{argmin}} \sum_{w\in\{w_0,w_1\}} \sum_{u\in\{u_1,u_2,u_3,u_4\}} \left\| \widehat{Q}^{\top}_{s_1ux_0} \begin{pmatrix} \dfrac{1}{1+\exp(-\theta_{x_0v_3})} \\ \dfrac{1}{1+\exp(-\theta_{x_0v_4})} \end{pmatrix} - \widehat{P}_{ws_1ux_0}\boldsymbol{e}_1 \right\|_2^2,$$

$$\widehat{\theta}_{x_1v_1}, \widehat{\theta}_{x_1v_3} = \underset{\theta_{x_1v_1}, \theta_{x_1v_3}}{\text{argmin}} \sum_{w\in\{w_0,w_1\}} \sum_{u\in\{u_1,u_2,u_3,u_4\}} \left\| \widehat{Q}^{\top}_{s_0ux_1} \begin{pmatrix} \dfrac{1}{1+\exp(-\theta_{x_1v_1})} \\ \dfrac{1}{1+\exp(-\theta_{x_1v_3})} \end{pmatrix} - \widehat{P}_{ws_0ux_1}\boldsymbol{e}_1 \right\|_2^2,$$

$$\widehat{\theta}_{x_1v_2}, \widehat{\theta}_{x_1v_4} = \underset{\theta_{x_1v_2}, \theta_{x_1v_4}}{\text{argmin}} \sum_{w\in\{w_0,w_1\}} \sum_{u\in\{u_1,u_2,u_3,u_4\}} \left\| \widehat{Q}^{\top}_{s_1ux_1} \begin{pmatrix} \dfrac{1}{1+\exp(-\theta_{x_1v_2})} \\ \dfrac{1}{1+\exp(-\theta_{x_1v_4})} \end{pmatrix} - \widehat{P}_{ws_1ux_1}\boldsymbol{e}_1 \right\|_2^2,$$

we have consistent estimators $\widehat{p}(x,v) = 1/\left(1+\exp(-\widehat{\theta}_{xv})\right)$ of the probabilities $p(x,v) = 1/(1+\exp(-\theta_{xv}))$ for $x \in \{x_0, x_1\}$ and $v \in \{v_1, v_2, v_3, v_4\}$, and consistent estimators $\widehat{p}(u,v) = \widehat{p}(u \,|\, v)\sum_{x\in\{x_0,x_1\}} \widehat{p}(x,v)$ of the probabilities $p(u \,|\, v)$ for $u \in \{u_1, u_2, u_3, u_4\}$ and $v \in \{v_1, v_2, v_3, v_4\}$.

Since

$$\begin{aligned}
&p(Y_{x_0} = y_0, Y_{x_1} = y_0) = \sum_{s,s'\in\{s_0,s_1\}} p(Y_{x_0,S_{x_0}} = y_0, Y_{x_1,S_{x_1}} = y_0, S_{x_0} = s, S_{x_1} = s') \\
&= \sum_{s,s'\in\{s_0,s_1\}} p(Y_s = y_0, Y_{s'} = y_0, S_{x_0} = s, S_{x_1} = s') \\
&= p(Y_{s_0} = y_0, Y_{s_0} = y_0, S_{x_0} = s_0, S_{x_1} = s_0) + p(Y_{s_0} = y_0, Y_{s_1} = y_0, S_{x_0} = s_0, S_{x_1} = s_1) \\
&\quad + p(Y_{s_1} = y_0, Y_{s_0} = y_0, S_{x_0} = s_1, S_{x_1} = s_0) + p(Y_{s_1} = y_0, Y_{s_1} = y_0, S_{x_0} = s_1, S_{x_1} = s_1) \\
&= p(U \in \{u_1, u_2\}, V = v_1) + p(U = u_1, V = v_2) + p(U = u_1, V = v_3) + p(U \in \{u_1, u_3\}, V = v_4) \\
&= p(u_1, v_1) + p(u_2, v_1) + p(u_1, v_2) + p(u_1, v_3) + p(u_1, v_4) + p(u_3, v_4), \\
&p(Y_{x_0} = y_0, Y_{x_1} = y_1) = \sum_{s,s'\in\{s_0,s_1\}} p(Y_{x_0,S_{x_0}} = y_0, Y_{x_1,S_{x_1}} = y_1, S_{x_0} = s, S_{x_1} = s') \\
&= \sum_{s,s'\in\{s_0,s_1\}} p(Y_s = y_0, Y_{s'} = y_1, S_{x_0} = s, S_{x_1} = s') \\
&= p(Y_{s_0} = y_0, Y_{s_0} = y_1, S_{x_0} = s_0, S_{x_1} = s_0) + p(Y_{s_0} = y_0, Y_{s_1} = y_1, S_{x_0} = s_0, S_{x_1} = s_1) \\
&\quad + p(Y_{s_1} = y_0, Y_{s_0} = y_1, S_{x_0} = s_1, S_{x_1} = s_0) + p(Y_{s_1} = y_0, Y_{s_1} = y_1, S_{x_0} = s_1, S_{x_1} = s_1) \\
&= p(U = u_2, V = v_2) + p(U = u_3, V = v_3) \\
&= p(u_2, v_2) + p(u_3, v_3), \\
&p(Y_{x_0} = y_1, Y_{x_1} = y_0) = \sum_{s,s'\in\{s_0,s_1\}} p(Y_{x_0,S_{x_0}} = y_1, Y_{x_1,S_{x_1}} = y_0, S_{x_0} = s, S_{x_1} = s') \\
&= \sum_{s,s'\in\{s_0,s_1\}} p(Y_s = y_1, Y_{s'} = y_0, S_{x_0} = s, S_{x_1} = s') \\
&= p(Y_{s_0} = y_1, Y_{s_0} = y_0, S_{x_0} = s_0, S_{x_1} = s_0) + p(Y_{s_0} = y_1, Y_{s_1} = y_0, S_{x_0} = s_0, S_{x_1} = s_1) \\
&\quad + p(Y_{s_1} = y_1, Y_{s_0} = y_0, S_{x_0} = s_1, S_{x_1} = s_0) + p(Y_{s_1} = y_1, Y_{s_1} = y_0, S_{x_0} = s_1, S_{x_1} = s_1) \\
&= p(U = u_3, V = v_2) + p(U = u_2, V = v_3) \\
&= p(u_3, v_2) + p(u_2, v_3),
\end{aligned}$$

$$
\begin{aligned}
p(Y_{x_0}=y_1,Y_{x_1}=y_1) &= \sum_{s,s'\in\{s_0,s_1\}} p(Y_{x_0,S_{x_0}}=y_1,Y_{x_1,S_{x_1}}=y_1,S_{x_0}=s,S_{x_1}=s') \\
&= \sum_{s,s'\in\{s_0,s_1\}} p(Y_s=y_1,Y_{s'}=y_1,S_{x_0}=s,S_{x_1}=s') \\
&= p(Y_{s_0}=y_1,Y_{s_0}=y_1,S_{x_0}=s_0,S_{x_1}=s_0)+p(Y_{s_0}=y_1,Y_{s_1}=y_1,S_{x_0}=s_0,S_{x_1}=s_1) \\
&\quad + p(Y_{s_1}=y_1,Y_{s_0}=y_1,S_{x_0}=s_1,S_{x_1}=s_0)+p(Y_{s_1}=y_1,Y_{s_1}=y_1,S_{x_0}=s_1,S_{x_1}=s_1) \\
&= p(U\in\{u_3,u_4\},V=v_1)+p(U=u_4,V=v_2)+p(U=u_4,V=v_3)+p(U\in\{u_2,u_4\},V=v_4) \\
&= p(u_3,v_1)+p(u_4,v_1)+p(u_4,v_2)+p(u_4,v_3)+p(u_2,v_4)+p(u_4,v_4),
\end{aligned}
$$

we can obtain the consistent estimators

$$
\begin{aligned}
&\widehat{p}(Y_{x_0}=y_0,Y_{x_1}=y_0)=\widehat{p}(u_1,v_1)+\widehat{p}(u_2,v_1)+\widehat{p}(u_1,v_2)+\widehat{p}(u_1,v_3)+\widehat{p}(u_1,v_4)+\widehat{p}(u_3,v_4), \\
&\widehat{p}(Y_{x_0}=y_0,Y_{x_1}=y_1)=\widehat{p}(u_2,v_2)+\widehat{p}(u_3,v_3), \\
&\widehat{p}(Y_{x_0}=y_1,Y_{x_1}=y_0)=\widehat{p}(u_3,v_2)+\widehat{p}(u_2,v_3), \\
&\widehat{p}(Y_{x_0}=y_1,Y_{x_1}=y_1)=\widehat{p}(u_3,v_1)+\widehat{p}(u_4,v_1)+\widehat{p}(u_4,v_2)+\widehat{p}(u_4,v_3)+\widehat{p}(u_2,v_4)+\widehat{p}(u_4,v_4).
\end{aligned}
$$

C CASE STUDY

We illustrate our results through the data set reported by LaLonde (1986) and re-analyzed by Dehejia and Wahba (1999). The aim of this study was to evaluate the effect on trainee earnings of the National Supported Work (NSW) demonstration, a job training program, in the field experiment. According to LaLonde (1986), in this study, individuals were randomly assigned to treatment (attendance) and control (non-attendance) groups with the estimates that would have been produced by an econometrician. However it seems that the random assignment was not successful. The data set used in this section is available from Dehejia's homepage (`https://users.nber.org/~rdehejia/nswdata2.html`). The sample size given in the homepage is 445, and the variables that we are interested in are:

X**:** an indicator of whether the individual attended the job training program (x_1: "attended"; x_0: "did not attend"),

Y**:** an indicator of whether the individual's earning increment increased between 1975 and 1978 (y_1: "increased"; y_0: "did not increase"),

S**:** a mediator for whether the individual's earnings was zero in 1975 (y_1: "non-zero"; y_0: "zero"),

Z**:** a joint indicator of marriage status and high school degree (z_3: "no degree" and "marriage"; z_2: "no degree" and "no marriage"; z_1: otherwise),

W**:** an indicator of age in years (w_0: age < 27; w_1: age ≥ 27).

We assume that the data generating process of this study is encoded in Figure 4b. Then, under the assumption that Condition 5 of Theorem 2 holds, together with Conditions 4 and 6, PNS $p(Y_{x_0}=y_0,Y_{x_1}=y_1)$ and the causal risk difference $p(Y_{x_0}=y_0,Y_{x_1}=y_1)-p(Y_{x_0}=y_1,Y_{x_1}=y_0)$ are evaluated by $\widehat{p}(Y_{x_0}=y_0,Y_{x_1}=y_1)=0.092$ and $\widehat{p}(Y_{x_0}=y_0,Y_{x_1}=y_1)-\widehat{p}(Y_{x_0}=y_1,Y_{x_1}=y_0)=0.092$, respectively, as shown in Table C.1. Here, the 2.5th and 97.5th percentiles of 1000 bootstrap replications of the estimates were used to derive the 95% confidential intervals[1]. According to Table C.1, it may be reasonable to assume monotonicity, i.e., $p(Y_{x_0}=y_1,Y_{x_1}=y_0)=0$, to evaluate the causal effects since there is a little difference between the estimates of PNS and causal risk. Additionally, Table C.2 and Figure C.1 show summary statistics and boxplots for 1000 bootstrap replications of the estimates ($\widehat{p}(Y_{x_0}=y_0,Y_{x_1}=y_0)$, $\widehat{p}(Y_{x_0}=y_0,Y_{x_1}=y_1)$, $\widehat{p}(Y_{x_0}=y_1,Y_{x_1}=y_0)$, $\widehat{p}(Y_{x_0}=y_1,Y_{x_1}=y_1)$) based on the proposed method, respectively. It seems that $\widehat{p}(Y_{x_0}=y_0,Y_{x_1}=y_1)$ and $\widehat{p}(Y_{x_0}=y_1,Y_{x_1}=y_0)$ are relatively stable from Table C.2 and Figure C.1.

[1]Strictly speaking, only the 761 replications that yielded invertible matrices $\widehat{P}_{xyzs\cdot w}$ were used.

Table C.1: Estimates of PNS and the causal risk difference in the NSW dataset.

	Estimate (95%CI)
PNS	0.092 (0.000, 0.372)
Causal risk difference	0.092 (−0.141, 0.29)

Table C.2: Basic statistics for 1000 bootstrap replications of estimates ($\widehat{p}(Y_{x_0} = y_0, Y_{x_1} = y_0)$, $\widehat{p}(Y_{x_0} = y_0, Y_{x_1} = y_1)$, $\widehat{p}(Y_{x_0} = y_1, Y_{x_1} = y_0)$, $\widehat{p}(Y_{x_0} = y_1, Y_{x_1} = y_1)$) based on our proposed method in the NSW dataset.

	$\widehat{p}(Y_{x_0} = y_0, Y_{x_1} = y_0)$	$\widehat{p}(Y_{x_0} = y_0, Y_{x_1} = y_1)$	$\widehat{p}(Y_{x_0} = y_1, Y_{x_1} = y_0)$	$\widehat{p}(Y_{x_0} = y_1, Y_{x_1} = y_1)$
Minimum	0.011	0.000	0.000	0.066
1st Quantile	0.244	0.044	0.023	0.420
Median	0.316	0.102	0.068	0.501
Mean	0.334	0.121	0.084	0.500
3rd Quantile	0.408	0.176	0.123	0.575
Maximum	0.984	0.800	0.640	1.052
s.e.	0.138	0.102	0.079	0.133

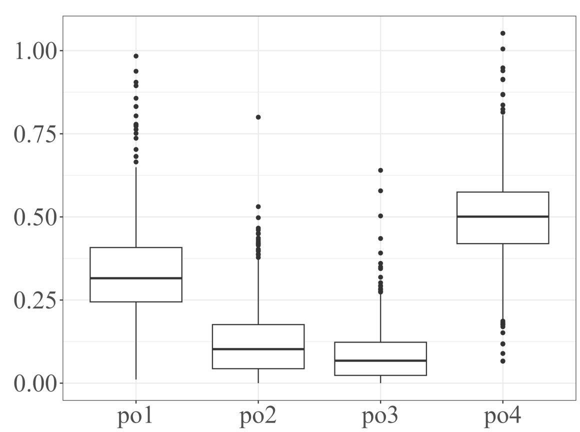


Figure C.1: Boxplots of estimates ($\widehat{p}(Y_{x_0} = y_0, Y_{x_1} = y_0)$, $\widehat{p}(Y_{x_0} = y_0, Y_{x_1} = y_1)$, $\widehat{p}(Y_{x_0} = y_1, Y_{x_1} = y_0)$, $\widehat{p}(Y_{x_0} = y_1, Y_{x_1} = y_1)$) based on our proposed method in the NSW dataset.