ϵ-Identifiability of Causal Quantities

Ang Li
Florida State University

Scott Mueller
Toyota Research Institute

Xin Shu
Florida State University

Judea Pearl
UCLA

Abstract

Identifying the effects of causes and causes of effects is vital in virtually every scientific field. Often, however, the needed probabilities may not be fully identifiable from the available data sources. This paper shows how approximate identifiability is still possible for several probabilities of causation. We term this ϵ-identifiability and demonstrate its usefulness in cases where the behavior of certain subpopulations can be restricted within sufficiently narrow bounds. In particular, we show how unidentifiable causal effects and counterfactual probabilities can be ϵ-identified when such allowances are made. Often, these allowances are easily measured and reasonably assumed. Finally, ϵ-identifiability is applied to the unit selection problem.

1 INTRODUCTION

Both Effects of Causes (EoC) and Causes of Effects (CoE) play an important role in many fields, such as health science, social science, and business. This is due, in part, to several features of EoC and CoE analysis. First, causal effects identified by the adjustment formula (Pearl, 1993) help decision-makers avoid randomized controlled trials and use purely observational data. Second, probabilities of causation have been proven critical in policy-making and personalized decision-making (Mueller and Pearl, 2022). Third, a linear combination of probabilities of causation has been used to solve the unit selection problem defined by Li and Pearl (2019, 2022b, 2024b). Additionally, causal quantities can increase the accuracy of machine learning models by combining causal quantities with the model's label (Li et al., 2020).

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300.

The identification of causal quantities has been studied for decades. Pearl first defined causal quantities such as causal effects (Pearl, 1993), probability of necessity and sufficiency (PNS), probability of sufficiency (PS), and probability of necessity (PN) (Pearl, 1999) and their identifiability (Pearl, 2009) using the structural causal model (SCM) (Galles and Pearl, 1998; Halpern, 2000). Pearl also developed identification conditions of causal effects, such as back-door and front-door criteria (Pearl, 1993), along with the *do*-calculus mathematically complete axiomatic identification system (Pearl, 1994). Pearl, Shpitser, and Bareinboim have studied more conditions for identifying causal effects (Shpitser and Pearl, 2009a; Bareinboim and Pearl, 2012). If causal effects are not identifiable, informative bounds can be obtained. Balke and Pearl (1997a) bounded causal effects from a randomized controlled trial (RCT) with imperfect compliance. Li and Pearl (2022a) bounded causal effects when adjustment variables are partially observed. Tian and Pearl (2000) proposed monotonicity as an identification condition of the binary probabilities of causation. Tian and Pearl (2000) additionally derived tight bounds, using linear programming, when probabilities of causation are not identifiable (Balke and Pearl, 1997b). Mueller et al. (2022), as well as Dawid et al. (2017), narrowed those bounds using additional covariate information and the corresponding causal structure. Recently, Li and Pearl (2024a) proposed theoretical work for non-binary probabilities of causation. Zhang et al. (2022) proposed numerical bounds for non-binary probabilities of causation and named it "Partial identification" (the difference between partial identification and the proposed ϵ-Identifiability will be discussed in the conclusion section). Furthermore, Shpitser and Pearl (2009b) introduced another essential causal measure, known as the effect of treatment on the treated (ETT), within the SCM framework and extensively investigated its identification.

In real-world applications, decision-makers are more willing to work with identifiable cases (i.e., the causal quantities have point estimations) because the bounds under unidentifiable cases may be less informative (e.g., $0.1 \leq \text{PNS} \leq 0.9$). Besides, estimating the bounds often requires a combination of experimental and observational data. However, the identifiability is often hard to achieve, so we wonder if something is sitting between the identifiable and the bounds. Inspired by the idea of the confidence interval, in this paper, we proposed the definition of ϵ-identifiability, in which more conditions of ϵ-identifiability can be found while the estimations of the causal quantities are still near point estimations.

2 PRELIMINARIES

Here, we review the definition of ETT, PNS, PS, and PN defined by Pearl (Pearl, 1999; Shpitser and Pearl, 2009a), as well as the definition of identifiability and the conditions for identifying ETT, PNS, PS, and PN (Tian and Pearl, 2000; Shpitser and Pearl, 2009b). Besides, we review the tight bounds of PNS, PS, and PN when they are unidentifiable (Tian and Pearl, 2000). Readers who are familiar with the above knowledge may skip this section.

Similarly to any works mentioned above, we used the causal language of the SCM (Galles and Pearl, 1998; Halpern, 2000). The introductory counterfactual sentence "Variable Y would have the value y, had X been x" in this language is denoted by $Y_x = y$, and shorted as y_x. We have two types of data: experimental data, which is in the form of causal effects (denoted as $P(y_x)$), and observational data, which is in the form of a joint probability function (denoted as $P(x, y)$). Besides, in the absence of further specification, let X and Y be two binary variables in a causal model M, let x and y stand for the propositions $X = true$ and $Y = true$, respectively, and x' and y' for their complements. For notational simplicity, we limit the discussion to binary treatments and effects; extensions to multi-valued variables are discussed by Pearl (see (Pearl, 2009) p. 286, footnote 5).

First, the definition of identifiable for any causal quantities defined using SCM is as follows:

Definition 1 (Identifiability)**.** *Let $Q(M)$ be any computable quantity of a class of SCM* $\mathbf{M}$ *that is compatible with graph G. We say that Q is identifiable in* $\mathbf{M}$ *if, for any pairs of models M_1 and M_2 from* $\mathbf{M}$, *$Q(M_1) = Q(M_2)$ whenever $P_{M_1}(v) = P_{M_2}(v)$, where $P(v)$ is the statistical distribution over the set V of observed variables. If our observations are limited and permit only a partial set F_M of features (of $P_M(v)$) to be estimated, we define Q to be identifiable from F_M if $Q(M_1) = Q(M_2)$ whenever $F_{M_1} = F_{M_2}$. (Pearl, 2009)*

Second, the definitions of four binary probabilities of causation defined using SCM are as follows (Pearl, 1999; Shpitser and Pearl, 2009a):

Definition 2 (Effect of treatment on the treated (ETT))**.** *Let X and Y be two binary variables in a causal model M, let x and y stand for the propositions $X = true$ and $Y = true$, respectively, and x' and y' for their complements. The effect of treatment on the treated is defined as the expression*

$$ETT \triangleq P(y_x|x')$$

Definition 3 (Probability of necessity (PN))**.** *Let X and Y be two binary variables in a causal model M, let x and y stand for the propositions $X = true$ and $Y = true$, respectively, and x' and y' for their complements. The probability of necessity is defined as the expression*

$$PN \triangleq P(y'_{x'}|x, y)$$

Definition 4 (Probability of sufficiency (PS))**.** *Let X and Y be two binary variables in a causal model M, let x and y stand for the propositions $X = true$ and $Y = true$, respectively, and x' and y' for their complements. The probability of sufficiency is defined as the expression*

$$PS \triangleq P(y_x|x', y')$$

Definition 5 (Probability of necessity and sufficiency (PNS))**.** *Let X and Y be two binary variables in a causal model M, let x and y stand for the propositions $X = true$ and $Y = true$, respectively, and x' and y' for their complements. The probability of necessity and sufficiency is defined as the expression*

$$PNS \triangleq P(y_x, y'_{x'})$$

Third, we review the identification conditions for causal effects (Pearl, 1993, 1995).

Definition 6 (Back-door criterion)**.** *Given an ordered pair of variables (X, Y) in a directed acyclic graph G, a set of variables Z satisfies the back-door criterion relative to (X, Y), if no node in Z is a descendant of X, and Z blocks every path between X and Y that contains an arrow into X.*

If a set of variables Z satisfies the back-door criterion for X and Y, the causal effects of X on Y are identifiable and given by the adjustment formula:

$$P(y_x) = \sum_z P(y|x, z)P(z). \quad (1)$$

Definition 7 (Front-door criterion)**.** *A set of variables Z is said to satisfy the front-door criterion relative to an ordered pair of variables (X, Y) if 1) Z intercepts all directed paths from X to Y; 2) there is no back-door path from X to Z; and 3) all back-door paths from Z to Y are blocked by X.*

If a set of variables Z satisfies the front-door criterion for X and Y, and $P(x, Z) > 0$, then the causal effects of X on Y are identifiable and given by the adjustment formula:

$$P(y_x) = \sum_z P(z|x) \sum_{x'} P(y|x', z)P(x').$$

If causal effects are not identifiable, Tian and Pearl (2000) provided the following bounds for the causal effects.

$$P(x, y) \leq P(y_x) \leq 1 - P(x, y'). \tag{2}$$

Fourth, we review the identification conditions for ETT (Shpitser and Pearl, 2009b).

Theorem 8. *If a set of variables Z satisfies the back-door criterion for X and Y, the ETT of X on Y are identifiable and given by the following formula:*

$$P(y_x|x') = \sum_z P(y|x, z)P(z|x'). \tag{3}$$

Theorem 9. *The ETT of X on Y are identifiable if $P(Y_X)$ are identifiable and given by the following formula:*

$$P(y_x|x') = \frac{P(y_x) - P(x, y)}{P(x')}. \tag{4}$$

Finally, we review the identification conditions for PNS, PS, and PN (Tian and Pearl, 2000).

Definition 10. *(Monotonicity) A Variable Y is said to be monotonic relative to variable X in a causal model M iff*

$$y'_x \wedge y_{x'} = false.$$

Theorem 11. *If Y is monotonic relative to X, then PNS, PN, and PS are all identifiable, and*

$$\begin{aligned} PNS &= P(y_x) - P(y_{x'}), \\ PN &= \frac{P(y) - P(y_{x'})}{P(x, y)}, \\ PS &= \frac{P(y_x) - P(y)}{P(x', y')}. \end{aligned}$$

If PNS, PN, and PS are not identifiable, informative bounds are given by Tian and Pearl (2000).

$$\max \left\{ \begin{array}{c} 0, \\ P(y_x) - P(y_{x'}), \\ P(y) - P(y_{x'}), \\ P(y_x) - P(y) \end{array} \right\} \leq \text{PNS} \tag{5}$$

$$\min \left\{ \begin{array}{c} P(y_x), \\ P(y'_{x'}), \\ P(x, y) + P(x', y'), \\ P(y_x) - P(y_{x'}) + \\ P(x, y') + P(x', y) \end{array} \right\} \geq \text{PNS} \tag{6}$$

$$\max \left\{ \begin{array}{c} 0, \\ \frac{P(y) - P(y_{x'})}{P(x,y)} \end{array} \right\} \leq \text{PN} \tag{7}$$

$$\min \left\{ \begin{array}{c} 1, \\ \frac{P(y'_{x'}) - P(x',y')}{P(x,y)} \end{array} \right\} \geq \text{PN} \tag{8}$$

$$\max \left\{ \begin{array}{c} 0, \\ \frac{P(y') - P(y'_x)}{P(x',y')} \end{array} \right\} \leq \text{PS} \tag{9}$$

$$\min \left\{ \begin{array}{c} 1, \\ \frac{P(y_x) - P(x,y)}{P(x',y')} \end{array} \right\} \geq \text{PS} \tag{10}$$

The identification conditions mentioned above (i.e., back-door and front-door criteria and monotonicity) are robust. However, it may still be hard to achieve in real-world applications. In this work, we extend the definition of identifiability, in which a sufficiently small interval is allowed. By the new definition, the estimates of causal quantities are still near point estimations, and more conditions for identifiability could be discovered. Again, if nothing is specified, the discussion in this paper will be restricted to binary treatment and effect (i.e., X and Y are binary).

3 MAIN RESULTS

First, we extend the definition of identifiability, which we call ϵ-identifiability.

Definition 12 (ϵ-Identifiability)**.** *Let $Q(M)$ be any computable quantity of a class of SCM $\mathbf{M}$ that is compatible with graph G. We say that Q is ϵ-identifiable in $\mathbf{M}$ (and ϵ-identified to q) if for every $P(v)$, there exists q s.t. for any model M from $\mathbf{M}$, $Q(M) \in [q - \epsilon, q + \epsilon]$ whenever the statistical data $P_M(v) = P(v)$, where $P(v)$ is the statistical distribution over the set V of observed variables. If our observations are limited and permit only a partial set F_M of features (of $P_M(v)$) to be estimated, we define Q to be ϵ-identifiable from F_M if $Q(m) \in [q - \epsilon, q + \epsilon]$ with statistical data F_M.*

With the above definition, the causal quantity is at a maximum distance of ϵ from its true value. We will use the infix operator symbol $\approx_\epsilon$ to represent its left-hand side being within ϵ of its right-hand side:

$$r \approx_\epsilon q \iff r \in [q - \epsilon, q + \epsilon]. \tag{11}$$

The following sections explicate conditions for ϵ-identifiability of causal effects, ETT, PNS, PS, and PN. The proofs of all the following theorems are provided in the appendix.

3.1 ϵ-Identifiability of Causal Effects

The causal effect $P(Y_X)$ can be ϵ-identified with information about the observational joint distribution $P(X, Y)$. This can be seen by rewriting Equation (2) as:

$$P(x, y) \leqslant P(y_x) \leqslant P(x, y) + P(x'). \tag{12}$$

Here, $P(y_x)$ is ϵ-identified to $P(x, y) + \epsilon$ when $P(x') \leqslant 2\epsilon$. This ϵ-identification indicates a lower bound of $P(x, y)$ and an upper bound of $P(x, y) + 2\epsilon$. Since $P(x') \leqslant 2\epsilon$, these bounds are equivalent to (12). Notably, only $P(x, y)$ and an upper bound on $P(x')$ are necessary to ϵ-identify $P(y_x)$. This is generalized in Theorem 13, without any assumptions of the causal structure.

Theorem 13. *The causal effect $P(Y_X)$ is ϵ-identified as follows:*

$$P(y_x) \approx_\epsilon P(x, y) + \epsilon \qquad \textit{if } P(x') \leqslant 2\epsilon, \tag{13}$$
$$P(y'_x) \approx_\epsilon P(x, y') + \epsilon \qquad \textit{if } P(x') \leqslant 2\epsilon, \tag{14}$$
$$P(y_{x'}) \approx_\epsilon P(x', y) + \epsilon \qquad \textit{if } P(x) \leqslant 2\epsilon, \tag{15}$$
$$P(y'_{x'}) \approx_\epsilon P(x', y') + \epsilon \qquad \textit{if } P(x) \leqslant 2\epsilon. \tag{16}$$

When the complete distribution $P(X, Y)$ is known, Theorem 13 provides no extra precision over Equation (12). Its power comes from when only part of the distribution is known and only an upper bound on $P(X)$ is available or able to be assumed.

Knowledge of a causal structure can aid in ϵ-identification, such as the structures shown in Figures 1a and 1b. (We will use Figure 1a, as the two figures are actually equivalent; see Definition 6 in Li and Pearl (2022a) for the detailed proof of their equivalence.) If the full joint distribution $P(X, Y, U)$ were available, the causal effect $P(Y_X)$ could be computed directly via the backdoor adjustment formula. In the absence of the full joint distribution, Theorem 14 allows for ϵ-identification of $P(y_x)$ using only $P(x)$, the conditional probability $P(y|x)$, and an upper bound on $P(u)$.

Theorem 14. *Given the causal graph in Figure 1a and $P(u) \leqslant P(x) - c$ for some constant c, where $0 < c \leqslant P(x)$,*

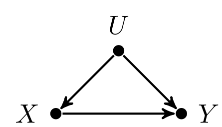
$$P(y_x) \approx_\epsilon P(y|x) + \frac{P(x) - c}{2cP(x) + P(x) + c} \cdot \epsilon,$$
$$\textit{if } P(u) \leqslant \frac{2cP(x)}{2cP(x) + P(x) + c} \cdot \epsilon. \tag{17}$$

Specifically, if $P(x) \geqslant 0.5$ and $P(u) \leqslant \min\{0.1, \frac{4}{13}\epsilon\}$, then the causal effect $P(y_x)$ is ϵ-identified to $P(y|x) + \frac{\epsilon}{13}$.

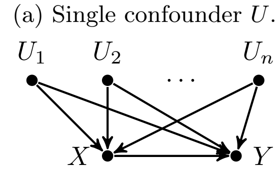
Note that $x \in \{x, x'\}$, $y \in \{y, y'\}$ (i.e., x may be replaced by x', and similarly for y), and $u \in \mathcal{U}$, where $\mathcal{U}$ denotes the (potentially multi-valued) domain of U, in Theorem 14 (e.g., you may substitute x with x' or y with y', and vice versa). The constant c should be maximized subject to the constraints $c \leqslant P(x) - P(u)$ and the condition specified in Equation (17) for a given ϵ. A larger c implies that $P(y_x)$ is more closely ϵ-identified with $P(y|x)$, but this must be balanced against minimizing ϵ. Importantly, when U takes on more values (i.e., when there are more potential confounders), it becomes more likely that one can find a value $u \in \mathcal{U}$ satisfying the conditions of the theorem.

As an example, if $P(x) \geqslant 0.5$ and $P(u) \leqslant 0.1$, then the causal effect $P(y_x)$ is ϵ-identified to $P(y|x) + \frac{\epsilon}{13}$ if $P(u) \leqslant \frac{4}{13}\epsilon$.

Essentially, $P(y_x)$ is ϵ-identified as $P(y|x)$ plus some fraction of ϵ when there exists a $P(u)$ that is sufficiently small. Therefore, the causal effect $P(y_x)$ is close to $P(y|x)$ if $P(U)$ is specific (i.e., there exists a $P(u)$ that is sufficiently small). In this case, Theorem 14 can be advantageous over the backdoor adjustment formula to compute $P(y_x)$, even when data on X, Y, and U are available, because $P(Y|X, U)$, required for the adjustment formula, is impractical to estimate with any $P(u)$ close to 0.



(a) Single confounder U.



(b) Multiple confounders $U_1, \ldots, U_n$.

Figure 1: Causal graphs with confounders.

3.2 ϵ-Identifiability of ETT

The effect of treatment on the treated (ETT) is a crucial quantity in the field of causal inference, as defined in Definition 2. ETT has been shown to be valuable for detecting latent heterogeneity (Pearl, 2015). Shpitser and Pearl proposed the identification of ETT through backdoor criteria, as outlined in Theorem 8. Furthermore, Shpitser and Pearl demonstrated the relationship between the ETT and both experimental and observational data as shown in Theorem 9. Therefore, Theorem 14 can be extended to incorporate ETT accordingly.

Theorem 15. *Given the causal graph in Figure 1a and $P(u) \leqslant P(x) - c$ for some constant c, where $0 <$*

$c \leqslant P(x) < 1$,

$$P(y_x|x') \approx_\epsilon \frac{P(y|x) - P(x,y)}{P(x')} + \\ + \frac{P(x) - c}{2cP(x) + P(x) + c} \cdot \epsilon, \\ \textit{if } P(u) \leqslant \frac{2cP(x)P(x')}{2cP(x) + P(x) + c} \cdot \epsilon. \tag{18}$$

Specifically, if $P(x) = 0.5$ *and* $P(u) \leqslant \min\{0.1, \frac{2}{13}\epsilon\}$, *then the* $P(y_x|x')$ *is* ϵ*-identified to* $P(y|x) + \frac{\epsilon}{13}$.

Note that $P(x)$ cannot equal 1, as this would make the conditional space x' in $P(y_x \mid x')$ undefined. Additionally, it should be noted that $x \in x, x'$, $y \in y, y'$, and $u \in \mathcal{U}$, where $\mathcal{U}$ denotes the (potentially multi-valued) domain of U. It is also worth mentioning that both Theorems 14 and 15 apply to the multi-confounder case shown in Figure 1b, by using the equivalent transformation method described in Theorem 7 of (Li and Pearl, 2022a), making these two theorems quite general.

3.3 ϵ-Identifiability of PNS

Even though Tian and Pearl derived tight bounds on PNS (Tian and Pearl, 2000), the PNS can potentially be sufficiently narrow when taking into account particular upper-bound assumptions on causal effects or observational probabilities. This can be seen by analyzing the bounds of PNS in Equations (5) and (6). Picking any of the arguments of the max function of the lower bound and any of the arguments to the min function of the upper bound, we can make a condition that the range of those two values is less than 2ϵ. For example, let us pick the second argument of the max function, $P(y_x) - P(y_{x'})$, and the first argument of the min function, $P(y_x)$:

$$P(y_{x'}) = P(y_x) - [P(y_x) - P(y_{x'})] \leqslant 2\epsilon. \tag{19}$$

Equation (19) is the assumption and the PNS is the ϵ-identified to ϵ above the lower bound or ϵ below the upper bound:

$$\text{PNS} \approx_\epsilon P(y_x) - P(y_{x'}) + \epsilon, \text{ or} \tag{20}$$

$$\text{PNS} \approx_\epsilon P(y_x) - \epsilon. \tag{21}$$

Since it is assumed that $P(y_{x'}) \leqslant 2\epsilon$, Equation (20) is equivalent to Equation (21). The complete set of ϵ-identifications and associated conditions are stated in Theorem 16.

Theorem 16. *The PNS is* ϵ*-identified as follows:*

$$PNS \approx_\epsilon \epsilon \quad \textit{if } P(y_x) \leqslant 2\epsilon, \tag{22}$$

$$PNS \approx_\epsilon \epsilon \quad \textit{if } P(y'_{x'}) \leqslant 2\epsilon, \tag{23}$$

$$PNS \approx_\epsilon \epsilon \quad \textit{if } P(x,y) + P(x',y') \leqslant 2\epsilon, \tag{24}$$

$$PNS \approx_\epsilon \epsilon \quad \textit{if } P(y_x) - P(y_{x'}) + P(x,y') + P(x',y) \leqslant 2\epsilon, \tag{25}$$

$$PNS \approx_\epsilon P(y_x) - \epsilon \quad \textit{if } P(y_{x'}) \leqslant 2\epsilon, \tag{26}$$

$$PNS \approx_\epsilon P(y'_{x'}) - \epsilon \quad \textit{if } P(y'_x) \leqslant 2\epsilon, \tag{27}$$

$$PNS \approx_\epsilon P(y_x) - P(y_{x'}) + \epsilon \quad \textit{if } P(x,y') + P(x',y) \leqslant 2\epsilon, \tag{28}$$

$$PNS \approx_\epsilon P(y_x) - P(y_{x'}) + \epsilon \quad \textit{if } P(y_{x'}) - P(y_x) + P(x,y) + P(x',y') \leqslant 2\epsilon, \tag{29}$$

$$PNS \approx_\epsilon P(x,y) + P(x',y') - \epsilon \quad \textit{if } P(y_{x'}) - P(y_x) + P(x,y) + P(x',y') \leqslant 2\epsilon, \tag{30}$$

$$PNS \approx_\epsilon P(y'_{x'}) - \epsilon \quad \textit{if } P(y') \leqslant 2\epsilon, \tag{31}$$

$$PNS \approx_\epsilon P(y_x) - \epsilon \quad \textit{if } P(y_x) + P(y_{x'}) - P(y) \leqslant 2\epsilon, \tag{32}$$

$$PNS \approx_\epsilon P(y) - P(y_{x'}) + \epsilon \quad \textit{if } P(y_x) + P(y_{x'}) - P(y) \leqslant 2\epsilon, \tag{33}$$

$$PNS \approx_\epsilon P(x,y) + P(x',y') - \epsilon \quad \textit{if } P(x',y') + P(y_{x'}) - P(x',y) \leqslant 2\epsilon, \tag{34}$$

$$PNS \approx_\epsilon P(y) - P(y_{x'}) + \epsilon \quad \textit{if } P(x',y') + P(y_{x'}) - P(x',y) \leqslant 2\epsilon, \tag{35}$$

$$PNS \approx_\epsilon P(y) - P(y_{x'}) + \epsilon \quad \textit{if } P(x,y') + P(y_x) - P(x,y) \leqslant 2\epsilon, \tag{36}$$

$$PNS \approx_\epsilon P(y_x) - \epsilon \quad \textit{if } P(y) \leqslant 2\epsilon, \tag{37}$$

$$PNS \approx_\epsilon P(y'_{x'}) - \epsilon \quad \textit{if } P(y'_{x'}) - P(y_x) + P(y) \leqslant 2\epsilon, \tag{38}$$

$$PNS \approx_\epsilon P(y_x) - P(y) + \epsilon \quad \textit{if } P(y'_{x'}) - P(y_x) + P(y) \leqslant 2\epsilon, \tag{39}$$

$$PNS \approx_\epsilon P(x,y) + P(x',y') - \epsilon \quad \textit{if } P(x,y) + P(y'_x) - P(x,y') \leqslant 2\epsilon, \tag{40}$$

$$PNS \approx_\epsilon P(y_x) - P(y) + \epsilon \quad \textit{if } P(x,y) + P(y'_x) - P(x,y') \leqslant 2\epsilon, \tag{41}$$

$$PNS \approx_\epsilon P(y_x) - P(y) + \epsilon \quad \textit{if } P(x',y) + P(y'_{x'}) - P(x',y') \leqslant 2\epsilon. \tag{42}$$

Note that in the above theorem, eight conditions consist solely of experimental probabilities or solely of observational probabilities. This potentially eliminates the need for some types of studies, at least partially, even when estimating a counterfactual quantity such as PNS. For example, if a decision-maker knows that $P(y)$ is large ($P(y) \geqslant 0.95$), they can immediately conclude PNS $\approx_{0.025} P(y'_{x'}) - 0.025$ (Equation (31)), without knowing the specific value of $P(y)$. Thus, only a control group study would be sufficient to estimate PNS.

There are real-world cases that exhibit near monotonicity, though not strictly so. For example, consider a scenario where a car is offered at a large discount (e.g., 50%). The proportion of customers who would have purchased the car at the original price but choose not to buy it when the discount is applied is likely small—indicating approximate monotonicity. In such cases, the PNS is not identifiable under the traditional definition. However, it becomes ϵ-identifiable under the following theorem.

Theorem 17. *The PNS is ϵ-identified to $P(y_x) - P(y_{x'}) + \epsilon$ if Y is nearly monotonic relative to X, such that, $P(y'_x, y_{x'}) \leq 2\epsilon$.*

Note that monotonicity is defined in terms of counterfactuals; however, as discussed above, in some real-world cases, this probability can be estimated from experience or through other tests, as illustrated in (Mueller and Pearl, 2023).

3.4 ϵ-Identifiability of PN and PS

Tian and Pearl derived tight bounds on PN and PS, in addition to PNS. Similar to the derivation in Theorem 16, we can potentially obtain sufficiently narrow bounds by assuming upper bounds on causal effects or observational probabilities. The sets of ϵ-identifications and their associated conditions are presented in Theorems 18 and 19. Moreover, similar to Theorem 17, in cases of near monotonicity, the ϵ-identifications of PN and PS are provided in Theorems 20 and 21.

Theorem 18. *The PN is ϵ-identified as follows:*

$$PN \approx_\epsilon \epsilon \quad \text{if } P(y'_{x'}) - P(x', y') \leqslant 2\epsilon P(x, y), \tag{43}$$

$$PN \approx_\epsilon 1 - \epsilon \quad \text{if } P(y_{x'}) - P(x', y) \leqslant 2\epsilon P(x, y), \tag{44}$$

$$PN \approx_\epsilon \frac{P(y) - P(y_{x'})}{P(x, y)} + \epsilon \quad \text{if } P(y_{x'}) - P(x', y) \leqslant 2\epsilon P(x, y), \tag{45}$$

$$PN \approx_\epsilon \frac{P({y'}_{x'}) - P(x', y')}{P(x, y)} - \epsilon \quad \text{if } P(x, y') \leqslant 2\epsilon P(x, y), \tag{46}$$

$$PN \approx_\epsilon \frac{P(y) - P(y_{x'})}{P(x, y)} + \epsilon \quad \text{if } P(x, y') \leqslant 2\epsilon P(x, y). \tag{47}$$

Theorem 19. *The PS is ϵ-identified as follows:*

$$PS \approx_\epsilon \epsilon \quad \text{if } P(y_x) - P(x, y) \leqslant 2\epsilon P(x', y'), \tag{48}$$

$$PS \approx_\epsilon 1 - \epsilon \quad \text{if } P(y'_x) - P(x, y') \leqslant 2\epsilon P(x', y'), \tag{49}$$

$$PS \approx_\epsilon \frac{P(y') - P(y'_x)}{P(x', y')} + \epsilon \quad \text{if } P(y'_x) - P(x, y') \leqslant 2\epsilon P(x', y'), \tag{50}$$

$$PS \approx_\epsilon \frac{P(y_x) - P(x, y)}{P(x', y')} - \epsilon \quad \text{if } P(x', y) \leqslant 2\epsilon P(x', y'), \tag{51}$$

$$PS \approx_\epsilon \frac{P(y') - P(y'_x)}{P(x', y')} + \epsilon \quad \text{if } P(x', y) \leqslant 2\epsilon P(x', y'). \tag{52}$$

Theorem 20. *The PN is ϵ-identified to $\frac{P(y)-P(y_{x'})}{P(x,y)} + \epsilon$ if Y is nearly monotonic relative to X, such that, $P(y'_x, y_{x'}) \leq 2\epsilon P(x, y)$.*

Theorem 21. *The PS is ϵ-identified to $\frac{P(y')-P(y'_x)}{P(x',y')} + \epsilon$ if Y is nearly monotonic relative to X, such that, $P(y'_x, y_{x'}) \leq 2\epsilon P(x', y')$.*

4 EXAMPLES

Here, we illustrate how to apply ϵ-Identifiability in real applications by two simulated examples.

4.1 Causal Effects of Medicine

Consider a medicine manufacturer who wants to know the causal effect of a new medicine on a disease. They conducted an observational study where 1500 patients were given access to the medicine; the results of the study are summarized in Table 1. In addition, the expert from the medicine manufacturer acknowledged that family history is the only confounder of taking medicine and recovery, and the family history of the disease is extremely rare; only 1% of the people have the family history.

Let $X = x$ denote that a patient chose to take the medicine, and $X = x'$ denote that a patient chose not to take the medicine. Let $Y = y$ denote that a patient recovered, and $Y = y'$ denote that a patient did not recover. Let $U = u$ denote that a patient has the family history, and $U = u'$ denote that a patient has no family history.

The simulated data in Table 1 provides the estimates

$$\begin{aligned}
P(x) &= \frac{1260}{1500} = 0.84, \\
P(x') &= \frac{240}{1500} = 0.16, \\
P(y|x) &= \frac{780}{1260} \approx 0.62, \\
P(y|x') &= \frac{210}{240} = 0.875.
\end{aligned}$$

To obtain the causal effect of the medicine (i.e., using adjustment formula (1)), we have to know the observational data associated with family history, which is difficult to obtain.

Table 1: Results of an observational study with 1500 individuals who have access to the medicine, where 1260 individuals chose to receive the medicine and 240 individuals chose not to.

	Medicine	No medicine
Recovered	780	210
Not recovered	480	30

Fortunately, we have the prior that $P(u) = 0.01$. Since $0.01 = P(u) \leq P(x) - 0.8$ (let $c = 0.8$) and $0.01 = P(u) < \frac{2c*0.025P(x)}{2cP(x)+P(x)+c} \approx 0.0113$, we can apply Theorem 14 to obtain that $P(y_x)$ is 0.025-identified to $P(y|x) + \frac{P(x)-c}{2cP(x)+P(x)+c}0.025 \approx 0.62$. This means the causal effect of the medicine is very close to 0.62 (i.e., 0.025 close), which can not be 0.025 far from 0.62.

Or even simpler, note that $P(x) = 0.84 > 0.5$ and $P(u) = 0.01 < 0.1$, $P(u) = 0.01 < \frac{4}{13} * 0.035 \approx 0.0108$. We obtain that $P(y_x)$ is 0.035-identified to $P(y|x) + \frac{0.035}{13} \approx 0.62$.

Similarly, since $0.01 = P(u) \leq P(x') - 0.15$ (let $c = 0.15$) and $0.01 = P(u) < \frac{2c*0.1P(x')}{2cP(x')+P(x')+c} \approx 0.0134$, we can apply Theorem 14 again to obtain that $P(y_{x'})$ is 0.1-identified to $P(y|x') + \frac{P(x')-c}{2cP(x')+P(x')+c}0.1 \approx 0.878$.

In contrast, without the concept of "ϵ-Identifiability," we would have to apply the general bounds in equation (2), because the joint distribution $P(X,Y,U)$ is not available; therefore, the adjustment formula (1) cannot be applied. We have, by equation (2), $0.52 \leq P(y_x) \leq 0.68$ and $0.14 \leq P(y_{x'}) \leq 0.98$, making it impossible to reach a decision. However, by Theorem 14, we find $P(y_x) \approx_{0.025} 0.62$ and $P(y_{x'}) \approx_{0.1} 0.878$, leading to the conclusion that the medicine is ineffective, without knowing the observational data associated with the family history.

4.2 PNS of Flu Shot

Consider a newly invented flu shot. After a vaccination company introduced a new flu shot, the number of people infected by flu reached the lowest point in 20 years (i.e., less than 5% of people infected by flu). The government concluded that the new flu shot is the key to success. However, some anti-vaccination associations believe it is because people's physical quality increases yearly. Therefore, they all want to know how many percentages of people are uninfected because of the flu shot. The PNS of the flu shot (i.e., the percentage of individuals who would not infect by the flu if they had taken the flu shot and would infect otherwise) is indeed what they want.

Let $X = x$ denote that an individual has taken the flu shot and $X = x'$ denote that an individual has not taken the flu shot. Let $Y = y$ denote an individual infected by the flu and $Y = y'$ denote an individual not infected by the flu.

If they want to apply the bounds of PNS in Equations (5) and (6), they must conduct both experimental and observational studies. However, note that $P(y) < 0.05$, one could apply Equation (37) in Theorem 16, which PNS is 0.025-identified to $P(y_x) - 0.025$ (i.e., PNS is very close to $P(y_x)$). Thus, according to the analysis of sample size in (Li et al., 2022), an experimental study for the treated group with a sample size of only 385 is adequate for estimating PNS.

5 ϵ-IDENTIFIABILITY IN UNIT SELECTION PROBLEM

One utility of the causal quantities is the unit selection problem (Li and Pearl, 2019, 2024b), in which Li and Pearl defined an objective causal function to select a set of individuals that have the desired mode of behavior.

Let X denote the binary treatment and Y denote the binary effect. According to Li and Pearl, individuals were divided into four response types: Complier (i.e., $P(y_x, {y'}_{x'})$), always-taker (i.e., $P(y_x, y_{x'})$), never-taker (i.e., $P({y'}_x, {y'}_{x'})$), and defier (i.e., $P({y'}_x, y_{x'})$). Suppose the payoff of selecting a complier, always-taker, never-taker, and defier is $\beta, \gamma, \theta, \delta$, respectively (i.e., benefit vector). The objective function (i.e., benefit function) that optimizes the composition of the four types over the selected set of individuals c is as follows:

$$\begin{aligned} f(c) = \beta P(y_x, y'_{x'}|c) + \gamma P(y_x, y_{x'}|c) + \\ + \theta P(y'_x, y'_{x'}|c) + \delta P(y'_x, y_{x'}|c). \end{aligned}$$

Li and Pearl provided two types of identifiability conditions for the benefit function. One is about the response type such that there is no defier in the population (i.e., monotonicity). Another is about the benefits vector's relations, such that $\beta + \delta = \gamma + \theta$ (i.e., gain equality). These two conditions are helpful but still too specific and challenging to satisfy in real-world applications. If the benefit function is not identifiable, it can be bounded using experimental and observational data. Here in this paper, we extend the gain equality to the ϵ-identifiability as stated in the following theorem.

Theorem 22. *Given a causal diagram G and distribution compatible with G, let C be a set of variables that does not contain any descendant of X in G, then the benefit function* $f(c) = \beta P(y_x, y'_{x'}|c) + \gamma P(y_x, y_{x'}|c) + \theta P(y'_x, y'_{x'}|c) + \delta P(y_{x'}, y'_x|c)$ *is* $\frac{|\beta-\gamma-\theta+\delta|}{2}$*-identified to* $(\gamma - \delta)P(y_x|c) + \delta P(y_{x'}|c) + \theta P(y'_{x'}|c) + \frac{\beta-\gamma-\theta+\delta}{2}$.

One critical use case of the above theorem is that decision-makers usually only care about the sign (gain

Table 2: Results of an experimental study with 1500 randomly selected customers were forced to apply the discount, and 1500 randomly selected customers were forced not to.

	Discount	No discount
Bought the purchase	900	750
No purchase	600	750

or lose) of the benefit function. Decision-makers can apply the above theorem before conducting any observational study to see if the sign of the benefit function can be determined, as we will illustrate in the next section.

5.1 Example: Non-immediate Profit

Consider the most common example in (Li and Pearl, 2019). A sale company proposed a discount on a purchase in order to increase the total non-immediate profit. The company assessed that the profit of offering the discount to complier, always-taker, never-taker, and defier is $\$100, -\$60, \$0, -\$140$, respectively. Let $X = x$ denote that a customer applied the discount, and $X = x$ denote that a customer did not apply the discount. Let $Y = y$ denote that a customer bought the purchase and $Y = y'$ denote that a customer did not. The benefit function is then (here c denote all customers)

$$\begin{aligned} f(c) = 100P(y_x, y'_{x'}|c) - 60P(y_x, y_{x'}|c) + \\ +0P(y'_x, y'_{x'}|c) - 140P(y'_x, y_{x'}|c). \end{aligned}$$

The company conducted an experimental study where 1500 randomly selected customers were forced to apply the discount, and 1500 randomly selected customers were forced not to. The results are summarized in Table 2. The experimental data reads $P(y_x|c) = 0.6$ and $P(y_{x'}|c) = 0.5$.

Before conducting any observational study, one can conclude that the benefit function is 10-identified to -12 using Theorem 22. This result indicates that the benefit function is at most 10 away from -12; thus, the benefit function is negative regardless of the observational data. The decision-maker then can easily conclude that the discount should not offer to the customers.

6 CONCLUSION

We introduced the concept of ϵ-identifiability for causal quantities and established corresponding conditions for causal effects, ETT, PNS, PN, PS, and the unit selection problem. The idea is meant to capture situations where full identifiability is too demanding, yet "near-identifiability" within a controlled error margin is sufficient for making sound decisions. This relaxation is particularly valuable in applied settings, where strict conditions for identification are rarely met but practitioners still need principled ways to draw inferences.

Most of the conditions we presented, apart from Theorems 14 and 15, can be derived directly from observational or experimental distributions. They therefore remain useful even when the underlying causal graph is only partially known. This makes the results immediately relevant to practice. At the same time, the two graph-dependent theorems provide a glimpse of how additional structural assumptions may help unlock ϵ-identifiability in cases where standard approaches would fail. A natural next step is to search for purely graphical conditions that complement the distribution-based ones and together give a more complete picture.

Another promising direction is the incorporation of covariates and their causal relationships. Prior studies (Dawid et al., 2017; Li and Pearl, 2022b; Mueller et al., 2022) already suggest that appropriate use of covariates can shrink bounds and sometimes restore identifiability, and it is plausible that similar benefits extend to the ϵ-identifiable setting. Exploring this systematically could significantly broaden the scope of the framework.

A key insight is that ϵ-identifiability is closely related to partial identification. While partial identification provides bounds when exact values are inaccessible, ϵ-identifiability introduces explicit conditions under which the bounds are guaranteed to be within a tolerable ϵ margin. In this way, it formalizes the idea of "near-identifiability" and allows practitioners to reason about causal quantities with quantified imprecision deemed acceptable for practical purposes.

For each of the causal quantities considered, we provided several ϵ-identifiable conditions based on different types of data. Our theorems not only show that bounded causal estimates exist but also give explicit ranges, thereby turning intractable questions into informative inference problems. This dual emphasis on existence and explicitness is one of the strengths of the approach.

Our main contribution is to lay the foundation for this new line of research. While tightness was not the focus here, we aimed to find minimal conditions to achieve a specified ϵ. Future work could explore how to define and prove tightness for multiple overlapping conditions.

We hope this framework inspires future theoretical developments and practical applications of ϵ-identifiability in causal inference.

Acknowledgements

This research was supported in parts by grants from the National Science Foundation [#IIS-2106908 and #CNS-2321786], and Toyota Research Institute of North America [#PO-000897].

References

Alexander Balke and Judea Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92 (439):1171–1176, 1997a.

Alexander A Balke and Judea Pearl. Probabilistic counterfactuals: Semantics, computation, and applications. Technical report, UCLA Dept. of Computer Science, 1997b.

Elias Bareinboim and Judea Pearl. Causal inference by surrogate experiments: z-identifiability. In Nando de Freitas and Kevin Murphy, editors, *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, pages 113–120, Corvallis, OR, 2012. AUAI Press.

Philip Dawid, Monica Musio, and Rossella Murtas. The probability of causation. *Law, Probability and Risk*, 16:163–179, 2017.

David Galles and Judea Pearl. An axiomatic characterization of causal counterfactuals. *Foundations of Science*, 3(1):151–182, 1998.

Joseph Y Halpern. Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research*, 12:317–337, 2000.

Ang Li and Judea Pearl. Unit selection based on counterfactual logic. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 1793–1799. International Joint Conferences on Artificial Intelligence Organization, 7 2019. doi: 10.24963/ijcai.2019/248.

Ang Li and Judea Pearl. Bounds on causal effects and application to high dimensional data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36(5), pages 5773–5780, 2022a.

Ang Li and Judea Pearl. Unit selection with causal diagram. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36(5), pages 5765–5772, 2022b.

Ang Li and Judea Pearl. Probabilities of causation with nonbinary treatment and effect. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 20465–20472, 2024a.

Ang Li and Judea Pearl. Unit selection with nonbinary treatment and effect. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 20473–20480, 2024b.

Ang Li, Suming J. Chen, Jingzheng Qin, and Zhen Qin. Training machine learning models with causal logic. In *Companion Proceedings of the Web Conference 2020*, pages 557–561, 2020.

Ang Li, Ruirui Mao, and Judea Pearl. Probabilities of causation: Adequate size of experimental and observational samples. Technical report, Department of Computer Science, University of California, Los Angeles, CA, 2022.

Scott Mueller and Judea Pearl. Personalized decision making – a conceptual introduction. Technical Report R-513, Department of Computer Science, University of California, Los Angeles, CA, 2022.

Scott Mueller and Judea Pearl. Monotonicity: Detection, refutation, and ramification. Technical report, UCLA Technical report, 2023.

Scott Mueller, Ang Li, and Judea Pearl. Causes of effects: Learning individual responses from population data. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)*, pages 2712–2718, 2022.

Judea Pearl. Aspects of graphical models connected with causality. *Proceedings of the 49th Session of the international Statistical Institute, Italy*, pages 399–401, 1993.

Judea Pearl. From imaging and stochastic control to a calculus of actions. *Papers from the 1994 AAAI Spring Symposium*, 1994.

Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

Judea Pearl. Probabilities of causation: Three counterfactual interpretations and their identification. *Synthese*, pages 93–149, 1999.

Judea Pearl. *Causality*. Cambridge university press, 2nd edition, 2009.

Judea Pearl. Detecting latent heterogeneity. *Sociological Methods and Research*, pages 1–20 (online), DOI: 10.1177/0049124115600597, 2015.

Ilya Shpitser and Judea Pearl. Effects of treatment on the treated: Identification and generalization. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pages 514–521. AUAI Press, Montreal, Quebec, 2009a.

Ilya Shpitser and Judea Pearl. Effects of treatment on the treated: Identification and generalization. Technical report, Department of Computer Science, University of California, Los Angeles, CA, 2009b. Extended version of paper that appeared in UAI-09.

Jin Tian and Judea Pearl. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28(1-4):287–313, 2000.

Junzhe Zhang, Jin Tian, and Elias Bareinboim. Partial counterfactual identification from observational and experimental data. In K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162, pages 26548–26558. PMLR, Jul 2022.

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. [Not Applicable]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Not Applicable]

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). [Not Applicable]
   (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). [Not Applicable]
   (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. [Not Applicable]
   (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]

Supplementary Materials

A Technical Appendices and Supplementary Material

A.1 Proof of Theorem 13

Proof. From Equation (2) we have,

$$P(x,y) \leq P(y_x) \leq 1 - P(x,y').$$

Let $1 - P(x,y') - P(x,y) \leq 2\epsilon$, we obtain $P(x') \leq 2\epsilon$.
Therefore, $P(y_x)$ is ϵ-identified to $P(x,y)+\epsilon$ if $P(x') \leq 2\epsilon$, Equation (13) holds. Similarily, we can substitute x, y with x', y', respectively. Equations (14) to (16) hold. □

A.2 Proof of Theorem 14

Proof. Let $u \in \mathcal{U}$, and $u' = \cup_{u'' \in \mathcal{U}, u'' \neq u} u''$. Then, we have,

$$\begin{aligned} P(y_x) &= P(y_x|u)P(u) + P(y_x|u')P(u') \\ &= P(y|x,u)P(u) + P(y|x,u')P(u') \text{ since u satisfied the back-door} \end{aligned}$$

Thus,

$$\begin{aligned} & P(y_x) \\ \geq\ & P(y|x,u')P(u') \\ =\ & P(y|x,u')(1-P(u)) \\ =\ & \frac{P(x,y,u')}{P(x,u')}(1-P(u)) \\ \geq\ & \frac{P(x,y)-P(u)}{P(x)}(1-P(u)) \\ =\ & P(y|x) - P(y|x)P(u) - \frac{P(u)}{P(x)} + \frac{P^2(u)}{P(x)} \\ \geq\ & P(y|x) - P(u) - \frac{P(u)}{P(x)} \\ =\ & P(y|x) - (1+\frac{1}{P(x)})P(u). \end{aligned}$$

Also if $P(x) \geq P(u) + c$ for some constant $c > 0$, we have,

$$\begin{aligned}
& P(y_x) \\
\leq\ & P(u) + P(y|x,u')(1-P(u)) \\
\leq\ & P(u) + \frac{P(x,y,u')}{P(x,u')}(1-P(u)) \\
\leq\ & P(u) + \frac{P(x,y)}{P(x)-P(u)}(1-P(u)) \\
\leq\ & P(u) + \frac{P(x,y)}{P(x)-P(u)} \\
=\ & P(u) + \frac{P(x,y)}{P(x)(1-\frac{P(u)}{P(x)})} \\
=\ & P(u) + \frac{P(x,y)(1-\frac{P(u)}{P(x)}) + P(y|x)P(u)}{P(x)(1-\frac{P(u)}{P(x)})} \\
=\ & P(u) + P(y|x) + \frac{P(y|x)P(u)}{P(x)-P(u)} \\
\leq\ & P(y|x) + P(u) + \frac{P(u)}{P(x)-P(u)} \\
\leq\ & P(y|x) + P(u) + \frac{P(u)}{c} \\
=\ & P(y|x) + P(u)(1+\frac{1}{c})
\end{aligned}$$

Therefore, we have,

$$P(y|x) - (1+\frac{1}{P(x)})P(u) \leq P(y_x),$$
$$P(y|x) + (1+\frac{1}{c})P(u) \geq P(y_x).$$

Let

$$(1+\frac{1}{c})P(u) + (1+\frac{1}{P(x)})P(u) \leq 2\epsilon.$$

We have,

$$\begin{aligned}
& P(u) \\
\leq\ & \frac{2}{2+\frac{1}{c}+\frac{1}{P(x)}}\epsilon \\
=\ & \frac{2cP(x)}{2cP(x)+P(x)+c}\epsilon.
\end{aligned}$$

Then we know that if $P(u) \leq \frac{2cP(x)}{2cP(x)+P(x)+c}\epsilon$,

$$\begin{aligned}
P(y|x) - (1+\frac{1}{P(x)})\frac{2cP(x)}{2cP(x)+P(x)+c}\epsilon \leq\ & P(y_x), \\
P(y|x) + (1+\frac{1}{c})\frac{2cP(x)}{2cP(x)+P(x)+c}\epsilon \geq\ & P(y_x), \\
P(y|x) - \frac{2cP(x)+2c}{2cP(x)+P(x)+c}\epsilon \leq\ & P(y_x), \\
P(y|x) + \frac{2cP(x)+2P(x)}{2cP(x)+P(x)+c}\epsilon \geq\ & P(y_x).
\end{aligned}$$

Therefore, $P(y_x)$ is ϵ-identified to $P(y|x) - \frac{2cP(x)+2c}{2cP(x)+P(x)+c}\epsilon + \epsilon = P(y|x) + \frac{P(x)-c}{2cP(x)+P(x)+c}\epsilon$.
Besides, if $P(x) \geq 0.5$ and $P(u) \leq 0.1$, let $c = 0.4$, we have

$$P(y|x) - (1 + \frac{1}{P(x)})P(u) \leq P(y_x),$$
$$P(y|x) + (1 + \frac{1}{c})P(u) \geq P(y_x).$$
$$P(y|x) - (1 + \frac{1}{0.5})P(u) \leq P(y_x),$$
$$P(y|x) + (1 + \frac{1}{0.4})P(u) \geq P(y_x).$$
$$P(y|x) - 3P(u) \leq P(y_x) \leq P(y|x) + 3.5P(u).$$

Let $3.5P(u) + 3P(u) \leq 2\epsilon$, we have $P(u) \leq \frac{4}{13}\epsilon$, and

$$P(y|x) - \frac{12}{13}\epsilon \leq \quad P(y_x) \quad \leq P(y|x) + \frac{14}{13}\epsilon.$$

Therefore, $P(y_x)$ is ϵ-identified to $P(y|x) - \frac{12}{13}\epsilon + \epsilon = P(y|x) + \frac{\epsilon}{13}$. □

A.3 Proof of Theorem 15

Proof. Similar to the proof of Theorem 14, let $u \in \mathcal{U}$ and $u' = \cup_{u'' \in \mathcal{U}, u'' \neq u} u''$, we have,

$$P(y|x) - (1 + \frac{1}{P(x)})P(u) \leq P(y_x),$$
$$P(y|x) + (1 + \frac{1}{c})P(u) \geq P(y_x).$$

then we obtain,

$$\frac{P(y|x)}{P(x')} - (1 + \frac{1}{P(x)})\frac{P(u)}{P(x')} \leq \frac{P(y_x)}{P(x')},$$
$$\frac{P(y|x)}{P(x')} + (1 + \frac{1}{c})\frac{P(u)}{P(x')} \geq \frac{P(y_x)}{P(x')}.$$

Let

$$(1 + \frac{1}{c})\frac{P(u)}{P(x')} + (1 + \frac{1}{P(x)})\frac{P(u)}{P(x')} \leq 2\epsilon.$$

We have,

$$\begin{aligned} & P(u) \\ \leq \quad & \frac{2P(x')}{2 + \frac{1}{c} + \frac{1}{P(x)}}\epsilon \\ = \quad & \frac{2cP(x)P(x')}{2cP(x) + P(x) + c}\epsilon. \end{aligned}$$

Then we know that if $P(u) \leq \frac{2cP(x)P(x')}{2cP(x)+P(x)+c}\epsilon$,

$$\frac{P(y|x)}{P(x')} - (1 + \frac{1}{P(x)})\frac{2cP(x)}{2cP(x) + P(x) + c}\epsilon \leq \quad \frac{P(y_x)}{P(x')},$$
$$\frac{P(y|x)}{P(x')} + (1 + \frac{1}{c})\frac{2cP(x)}{2cP(x) + P(x) + c}\epsilon \geq \quad \frac{P(y_x)}{P(x')},$$
$$\frac{P(y|x)}{P(x')} - \frac{2cP(x) + 2c}{2cP(x) + P(x) + c}\epsilon \leq \quad \frac{P(y_x)}{P(x')},$$
$$\frac{P(y|x)}{P(x')} + \frac{2cP(x) + 2P(x)}{2cP(x) + P(x) + c}\epsilon \geq \quad \frac{P(y_x)}{P(x')}.$$

By Theorem 9, we have,

$$P(y_x|x') = \frac{P(y_x)}{P(x')} - \frac{P(x,y)}{P(x')}.$$

Therefore, $P(y_x|x')$ is ϵ-identified to $\frac{P(y|x)}{P(x')} - \frac{P(x,y)}{P(x')} - \frac{2cP(x)+2c}{2cP(x)+P(x)+c}\epsilon + \epsilon = \frac{P(y|x)}{P(x')} - \frac{P(x,y)}{P(x')} + \frac{P(x)-c}{2cP(x)+P(x)+c}\epsilon$. Besides, if $P(x) = 0.5$ and $P(u) \leq 0.1$, let $c = 0.4$, similarly to the proof of Theorem 14, we have,

$$\begin{aligned}
P(y|x) - (1 + \frac{1}{0.5})P(u) &\leq P(y_x), \\
P(y|x) + (1 + \frac{1}{0.4})P(u) &\geq P(y_x). \\
2P(y|x) - 2(1 + \frac{1}{0.5})P(u) &\leq 2P(y_x), \\
2P(y|x) + 2(1 + \frac{1}{0.4})P(u) &\geq 2P(y_x). \\
2P(y|x) - 6P(u) \leq 2P(y_x) &\leq 2P(y|x) + 7P(u).
\end{aligned}$$

By Theorem 9 and $P(x) = P(x') = 0.5$, we have,

$$\begin{aligned}
P(y_x|x') &= \frac{P(y_x)}{P(x')} - \frac{P(x,y)}{P(x')} \\
&= \frac{P(y_x)}{0.5} - \frac{P(x,y)}{P(x)} \\
&= 2P(y_x) - P(y|x).
\end{aligned}$$

Let $7P(u) + 6P(u) \leq 2\epsilon$, we have $P(u) \leq \frac{2}{13}\epsilon$, and

$$P(y|x) - \frac{12}{13}\epsilon \leq \quad P(y_x|x') \quad \leq P(y|x) + \frac{14}{13}\epsilon.$$

Therefore, $P(y_x|x')$ is ϵ-identified to $P(y|x) - \frac{12}{13}\epsilon + \epsilon = P(y|x) + \frac{\epsilon}{13}$. ∎

A.4 Proof of Theorem 16

Proof. From the bounds of PNS in Equations (5) and (6) is as follows:

$$\max\left\{\begin{array}{c} 0, \\ P(y_x) - P(y_{x'}), \\ P(y) - P(y_{x'}), \\ P(y_x) - P(y) \end{array}\right\} \leq \text{PNS}$$

$$\min\left\{\begin{array}{c} P(y_x), \\ P(y'_{x'}), \\ P(x,y) + P(x',y'), \\ P(y_x) - P(y_{x'}) + \\ +P(x,y') + P(x',y) \end{array}\right\} \geq \text{PNS}.$$

Let $P(y_x) - 0 \leq 2\epsilon$, we obtain that PNS is ϵ-identified to ϵ if $P(y_x) \leq 2\epsilon$, Equation (22) holds. Similarly, the rest of 20 equations can be obtained by letting

$$\begin{aligned}
P({y'}_{x'}) - 0 &\leq 2\epsilon,\\
P(x,y) + P(x',y') - 0 &\leq 2\epsilon,\\
P(y_x) - P(y_{x'}) + P(x,y') + P(x',y) - 0 &\leq 2\epsilon,\\
P(y_x) - (P(y_x) - P(y_{x'})) &\leq 2\epsilon,\\
P(y'_{x'}) - (P(y_x) - P(y_{x'})) &\leq 2\epsilon,\\
P(x,y) + P(x',y') - (P(y_x) - P(y_{x'})) &\leq 2\epsilon,\\
P(y_x) - P(y_{x'}) + P(x,y') + P(x',y) - &\\
(P(y_x) - P(y_{x'})) &\leq 2\epsilon,\\
P(y_x) - (P(y) - P(y_{x'})) &\leq 2\epsilon,\\
P(y'_{x'}) - (P(y) - P(y_{x'})) &\leq 2\epsilon,\\
P(x,y) + P(x',y') - (P(y) - P(y_{x'})) &\leq 2\epsilon,\\
P(y_x) - P(y_{x'}) + P(x,y') + P(x',y) - &\\
(P(y) - P(y_{x'})) &\leq 2\epsilon,\\
P(y_x) - (P(y_x) - P(y)) &\leq 2\epsilon,\\
P(y'_{x'}) - (P(y_x) - P(y)) &\leq 2\epsilon,\\
P(x,y) + P(x',y') - (P(y_x) - P(y)) &\leq 2\epsilon,\\
P(y_x) - P(y_{x'}) + P(x,y') + P(x',y) - &\\
(P(y_x) - P(y)) &\leq 2\epsilon.
\end{aligned}$$

□

A.5 Proof of Thereom 17

Proof. First, we have,

$$\begin{aligned}
& P(y_x, y'_{x'}) - P(y_{x'}, y'_x)\\
=\ & P(y_x, y'_{x'}, x) + P(y_x, y'_{x'}, x') - P(y_{x'}, y'_x, x) - P(y_{x'}, y'_x, x')\\
=\ & P(y, y'_{x'}, x) + P(y_x, y', x') - P(y_{x'}, y', x) - P(y, y'_x, x')\\
=\ & P(y, y'_{x'}, x) - P(y_{x'}, y', x) + P(y_x, y', x') - P(y, y'_x, x')\\
=\ & P(x,y) - P(y, y_{x'}, x) - P(y_{x'}, y', x) + P(y_x, y', x') + P(y, y_x, x') - P(x', y)\\
=\ & P(x,y) - P(y_{x'}, x) + P(y_x, x') - P(x', y)\\
=\ & P(x,y) - P(y_{x'}) + P(y_{x'}, x') + P(y_x) - P(y_x, x) - P(x', y)\\
=\ & P(x,y) - P(y_{x'}) + P(y, x') + P(y_x) - P(y, x) - P(x', y)\\
=\ & P(y_x) - P(y_{x'})
\end{aligned}$$

Then, we have,

$$\begin{aligned}
& PNS\\
=\ & P(y_x, y'_{x'})\\
=\ & P(y_x) - P(y_{x'}) + P(y'_x, y_{x'})
\end{aligned}$$

Plus if $P(y'_x, y_{x'}) \leq 2\epsilon$, we have,

$$P(y_x) - P(y_{x'}) \leq PNS \leq P(y_x) - P(y_{x'}) + 2\epsilon$$

and we obtain that,

$$PNS \approx_\epsilon P(y_x) - P(y_{x'}) + \epsilon.$$

□

A.6 Proof of Theorem 18

Proof. From the bounds of PN in Equations (7) and (8) is as follows:

$$\max\left\{\begin{array}{c}0,\\ \frac{P(y)-P(y_{x'})}{P(x,y)}\end{array}\right\} \leq \text{PN} \leq \min\left\{\begin{array}{c}1,\\ \frac{P(y'_{x'})-P(x',y')}{P(x,y)}\end{array}\right\}$$

Let $\frac{P(y'_{x'})-P(x',y')}{P(x,y)} - 0 \leq 2\epsilon$, we obtain that PN is ϵ-identified to ϵ if $P(y'_{x'}) - P(x',y') \leq 2P(x,y)\epsilon$, Equation (43) holds.
Similarly, the rest of 4 equations can be obtained by letting

$$\begin{aligned} 1 - \frac{P(y)-P(y_{x'})}{P(x,y)} &\leq 2\epsilon,\\ \frac{P(y'_{x'})-P(x',y')}{P(x,y)} - \frac{P(y)-P(y_{x'})}{P(x,y)} &\leq 2\epsilon. \end{aligned}$$

□

A.7 Proof of Theorem 19

Proof. From the bounds of PS in Equations (9) and (10) is as follows:

$$\max\left\{\begin{array}{c}0,\\ \frac{P(y')-P(y'_{x})}{P(x',y')}\end{array}\right\} \leq \text{PS} \leq \min\left\{\begin{array}{c}1,\\ \frac{P(y_x)-P(x,y)}{P(x',y')}\end{array}\right\}$$

Let $\frac{P(y_x)-P(x,y)}{P(x',y')} - 0 \leq 2\epsilon$, we obtain that PS is ϵ-identified to ϵ if $P(y_x) - P(x,y) \leq 2P(x',y')\epsilon$, Equation (48).
Similarly, the rest of 4 conditions can be obtained by letting

$$\begin{aligned} 1 - \frac{P(y')-P({y'}_{x})}{P(x',y')} &\leq 2\epsilon,\\ \frac{P(y_x)-P(x,y)}{P(x',y')} - \frac{P(y')-P({y'}_{x})}{P(x',y')} &\leq 2\epsilon. \end{aligned}$$

□

A.8 Proof of Thereom 20

Proof.

$$\begin{aligned} & PN\\ =\ & P(y'_{x'}|x,y)\\ =\ & \frac{P(y'_{x'},x,y)}{P(x,y)}\\ =\ & \frac{P(y'_{x'})-P(x',y')-P(y'_{x'},x,y')}{P(x,y)}\\ =\ & \frac{P(y'_{x'})-P(x',y')-P(x,y')+P(y_{x'},x,y')}{P(x,y)}\\ =\ & \frac{P(y'_{x'})-P(y')+P(y_{x'},x,y'_x)}{P(x,y)}\\ =\ & \frac{P(y)-P(y_{x'})+P(y_{x'},y'_x,x)}{P(x,y)} \end{aligned}$$

Plus if $P(y'_x, y_{x'}) \leq 2\epsilon P(x,y)$, we have,

$$\frac{P(y'_x,y_{x'},x)}{P(x,y)} \leq \frac{P(y'_x,y_{x'})}{P(x,y)} \leq 2\epsilon$$

and then,

$$\frac{P(y) - P(y_{x'})}{P(x,y)} \leq PN \leq \frac{P(y) - P(y_{x'})}{P(x,y)} + 2\epsilon$$

and we obtain that,

$$PN \approx_{\epsilon} \frac{P(y) - P(y_{x'})}{P(x,y)} + \epsilon.$$

□

A.9 Proof of Thereom 21

Proof. The proof follows exactly the same steps as in the proof of Theorem 20 in the previous section, except that x is replaced by x', x' by x, y by y', and y' by y. □

A.10 Proof of Theorem 22

Proof.

$$\begin{aligned}
& f(c) \\
= \ & \beta P(y_x, y'_{x'}|c) + \gamma P(y_x, y_{x'}|c) + \\
& \theta P(y'_x, y'_{x'}|c) + \delta P(y'_x, y_{x'}|c) \\
= \ & \beta P(y_x, y'_{x'}|c) + \gamma [P(y_x|c) - P(y_x, y'_{x'}|c)] + \\
& \theta [P(y'_{x'}) - P(y_x, y'_{x'}|c)] + \delta P(y'_x, y_{x'}|c) \\
= \ & \gamma P(y_x|c) + \theta P(y'_{x'}|c) + (\beta - \gamma - \theta) P(y_x, y'_{x'}|c) + \\
& \delta P(y'_x, y_{x'}|c).
\end{aligned} \tag{53}$$

Note that, we have,

$$P(y'_x, y_{x'}|c) = P(y_x, y'_{x'}|c) - P(y_x|c) + P(y_{x'}|c). \tag{54}$$

Substituting Equation (54) into Equation (53), we have,

$$\begin{aligned}
& f(c) \\
= \ & \gamma P(y_x|c) + \theta P(y'_{x'}|c) + (\beta - \gamma - \theta) P(y_x, y'_{x'}|c) + \\
& \delta P(y'_x, y_{x'}|c) \\
= \ & \gamma P(y_x|c) + \theta P(y'_{x'}|c) + (\beta - \gamma - \theta) P(y_x, y'_{x'}|c) + \\
& \delta [P(y_x, y'_{x'}|c) - P(y_x|c) + P(y_{x'}|c)] \\
= \ & (\gamma - \delta) P(y_x|c) + \delta P(y_{x'}|c) + \theta P(y'_{x'}|c) + \\
& (\beta - \gamma - \theta + \delta) P(y_x, y'_{x'}|c).
\end{aligned}$$

Case 1: If $\beta - \gamma - \theta + \delta \geq 0$,

$$\begin{aligned}
& f(c) \\
\leq \ & (\gamma - \delta) P(y_x|c) + \delta P(y_{x'}|c) + \theta P(y'_{x'}|c) + \\
& \frac{\beta - \gamma - \theta + \delta}{2} + \frac{|\beta - \gamma - \theta + \delta|}{2} \\
= \ & (\gamma - \delta) P(y_x|c) + \delta P(y_{x'}|c) + \theta P(y'_{x'}|c) + \\
& \beta - \gamma - \theta + \delta.
\end{aligned}$$

and,

$$\begin{aligned}
& f(c) \\
\geq \ & (\gamma - \delta) P(y_x|c) + \delta P(y_{x'}|c) + \theta P(y'_{x'}|c) + \\
& \frac{\beta - \gamma - \theta + \delta}{2} - \frac{|\beta - \gamma - \theta + \delta|}{2} \\
= \ & (\gamma - \delta) P(y_x|c) + \delta P(y_{x'}|c) + \theta P(y'_{x'}|c).
\end{aligned}$$

Therefore, $f(c)$ is $\frac{|\beta-\gamma-\theta+\delta|}{2}$-identified to $(\gamma-\delta)P(y_x|c)+\delta P(y_{x'}|c)+\theta P(y'_{x'}|c)+\frac{\beta-\gamma-\theta+\delta}{2}$.
Case 2: If $\beta-\gamma-\theta+\delta<0$,

$$
\begin{aligned}
& f(c) \\
\leq\ & (\gamma-\delta)P(y_x|c)+\delta P(y_{x'}|c)+\theta P(y'_{x'}|c)+ \\
& \frac{\beta-\gamma-\theta+\delta}{2}+\frac{|\beta-\gamma-\theta+\delta|}{2} \\
=\ & (\gamma-\delta)P(y_x|c)+\delta P(y_{x'}|c)+\theta P(y'_{x'}|c).
\end{aligned}
$$

and,

$$
\begin{aligned}
& f(c) \\
\geq\ & (\gamma-\delta)P(y_x|c)+\delta P(y_{x'}|c)+\theta P(y'_{x'}|c)+ \\
& \frac{\beta-\gamma-\theta+\delta}{2}-\frac{|\beta-\gamma-\theta+\delta|}{2} \\
=\ & (\gamma-\delta)P(y_x|c)+\delta P(y_{x'}|c)+\theta P(y'_{x'}|c)+ \\
& \beta-\gamma-\theta+\delta.
\end{aligned}
$$

Therefore, $f(c)$ is $\frac{|\beta-\gamma-\theta+\delta|}{2}$-identified to $(\gamma-\delta)P(y_x|c)+\delta P(y_{x'}|c)+\theta P(y'_{x'}|c)+\frac{\beta-\gamma-\theta+\delta}{2}$. $\square$