

Statistica Sinica Preprint No: SS-2025-0374	
Title	Bidirectional Causal Inference for Binary Outcomes in the Presence of Unmeasured Confounding
Manuscript ID	SS-2025-0374
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0374
Complete List of Authors	Yafang Deng, Kang Shuai and Shanshan Luo
Corresponding Authors	Shanshan Luo
E-mails	shanshanluo@btbu.edu.cn
Notice: Accepted author version.	

Bidirectional causal inference for binary outcomes in the presence of unmeasured confounding

Yafang Deng¹, Kang Shuai², Shanshan Luo¹

¹ *School of Mathematics and Statistics,*

Beijing Technology and Business University

² *The Dartmouth Institute for Health Policy and Clinical Practice,*

Geisel School of Medicine at Dartmouth

Abstract: Bidirectional causal relationships arising from mutual interactions between variables are commonly observed within biomedical, econometrical, and social science contexts. When such relationships are further complicated by unobserved factors, identifying causal effects in both directions becomes especially challenging. For continuous variables, methods that utilize two instrumental variables from both directions have been proposed to explore bidirectional causal effects within linear models. However, the existing techniques are not immediately applicable when the key variables of interest are binary. To address these issues, we propose a structural equation modeling approach that links observed binary variables to latent continuous variables through a constrained mapping. We firstly establish the identification results for bidirectional causal effects using a pair of instrumental variables. Then, we develop an estimation method for

the corresponding causal parameters. We also conduct sensitivity analysis under scenarios where certain identification conditions are violated. Finally, we apply our approach to investigate the bidirectional causal relationship between heart disease and diabetes, demonstrating its practical utility in biomedical researches.

Key words and phrases: Bidirectional causal inference; Binary outcomes; Instrumental variables; Selection mechanism; Sensitivity analysis.

1. Introduction

The emergence of feedback mechanisms between entities often signifies mutual causal influences, a ubiquitous phenomenon across real-world systems (Smith, 1999; Engelgau et al., 2019; Riaz et al., 2017). In economics, a bidirectional causal relationship exists between health status and wealth accumulation: poor health erodes economic resources, while economic deprivation compromises health, forming a self-reinforcing cycle (Smith, 1999; Engelgau et al., 2019). In biomedicine, tumor cells and the immune system engage in mutual causation—immune surveillance exerts selective pressures that sculpt cancer clonal evolution, whereas cancer cells deploy immunosuppressive molecules to subvert immune responses, establishing a malignant “immunoediting-escape” loop (Riaz et al., 2017). Similarly, the gut microbiota and brain maintain bidirectional communication via neuro-immune-endocrine pathways: stress-induced neural activity alters microbial compo-

sition, and dysbiosis independently disrupts brain function (Foster et al., 2017; Cryan et al., 2019). Psychologically, sleep and depression exhibit reciprocal causality: chronic insomnia elevates risks for psychiatric disorders, while over 90% of depressed patients manifest sleep disruptions consistent with insomnia (Palagini et al., 2013). All these practical phenomenon highlights the significance of investigating reverse causality, and methods for bidirectional causal effects definitely require further exploration.

However, in the study of bidirectional causal relationships, the presence of unobserved confoundings complicates the real-world problems. It may come with confounding bias so that the true causal effect cannot be recovered, which leads to invalid causal conclusions. To address this issue, instrumental variable (IV) approach offers a potential solution for identifying causal effects by eliminating the confounding bias. The related IV researches on bidirectional causal inference are primarily conducted within Mendelian Randomization (MR) framework. Specifically, MR utilizes independent genetic IVs for two traits to effectively disentangle the causal relations between them, thereby separating the additional associations caused by reverse causation (Davey Smith and Hemani, 2014). For example, Gao et al. (2019) regarded reverse causality as a source of confounding and used two-sample MR to explore the causal relationships between insomnia and

five major psychiatric disorders. Xue and Pan (2020) refined the traditional bidirectional MR approach, enabling robust inference of causal direction between two traits even in the presence of horizontal pleiotropy. Li and Ye (2024) assessed commonly used identifiability assumptions in bidirectional MR models and proposed a new focusing framework to test bidirectional causal effects with potentially invalid instruments. Xie et al. (2024) addressed the identifiability within bidirectional MR models by correctly selecting valid instrument sets using observational data and developed a cluster fusion-like method to discover valid instruments.

Previous research on bidirectional causal inference based on MR primarily focuses on recovering causal effects and related applications with continuous outcomes. However, binary outcomes are also frequently encountered in applied settings, particularly in epidemiological studies involving the disease risk factors. For instance, Ten Have et al. (2003) conducted a placebo-controlled randomized trial on cholesterol levels, where participants in the treatment group listened to audio tapes delivering dietary and nursing interventions, and the binary outcome was defined as whether or not participants had high cholesterol levels. Similarly, Borraacci et al. (2013) evaluated the impact of preoperative oral morphine sulfate on postoperative pain relief, where two binary outcome variables were constructed based on resting and

movement pain scores using established thresholds (Lupparelli and Mattei, 2020). Compared with continuous variables, simple linear models are not applicable for binary outcomes. Clarke and Windmeijer (2012) provided a comprehensive review of IV estimation methods for binary outcomes, focusing on identification strategies that rely on linear selection models and the control function approach. However, these methods have thus far been applied exclusively within traditional unidirectional causal frameworks.

In this paper, we propose bidirectional causal frameworks under unmeasured confounding for observed binary variables, employing IV approaches.

This paper makes the following three main contributions: First, we formulate the causal problem based on a linear structural equation modeling (SEM) framework to capture bidirectional causal relationships between continuous latent variables. Second, we use a pair of IVs to establish the identifiability of bidirectional causal effects and derive their estimators, along with the corresponding asymptotic properties. Third, we perform sensitivity analysis to evaluate the proposed method when key assumptions are possibly violated, including the exclusion restriction assumption and the restrictive normality assumption.

The remainder of this paper is organized as follows. Section 2 presents the proposed causal framework, including the problem formulation, core

assumptions and the identification results. Section 3 conducts sensitivity analysis for two main assumptions, including the normality and exclusion restriction assumptions. To illustrate the numerical performance of our method, in Section 4, we conduct detailed simulation experiments under various settings when the key conditions are met or not met. We further apply the proposed method to study the relationship between heart disease and diabetes in Section 5. Finally, we briefly conclude this paper in Section 6.

2. Notation and Framework

2.1 Binary bidirectional model

Let $X \in \{0, 1\}$ be a binary treatment variable and $Y \in \{0, 1\}$ an observed binary outcome. In observational studies, both X and Y may be influenced by background covariates, including observed covariates C and unobserved confounders U and V . In our motivating application, X indicates whether one has coronary heart disease or not, and Y indicates whether one has diabetes or not. Previous studies (Henning, 2018; Katakami, 2018) have shown that individuals with diabetes ($Y = 1$) are at a significantly high risk of having heart disease ($X = 1$); conversely, complications or treatments associated with heart disease ($X = 1$) may impair

2.1 Binary bidirectional model

glucose regulation, thereby increasing the risk of diabetes ($Y = 1$).

In such scenarios, assuming a unidirectional causal relationship (e.g., $X \rightarrow Y$) may ignore potential reverse effects, leading to biased estimates of causal effects. To more accurately capture the mutual dependence between the variables, we introduce a bidirectional causal mechanism to account for the possible reciprocal influences. Specifically, we assume the following generating mechanism (Hausman, 1983):

$$\begin{aligned} X^\circ &= \mu_{x0} + \beta_{y \rightarrow x} Y^\circ + \mu_{xc} C + U, & X &= \mathbb{I}(X^\circ > 0), \\ Y^\circ &= \mu_{y0} + \beta_{x \rightarrow y} X^\circ + \mu_{yc} C + V, & Y &= \mathbb{I}(Y^\circ > 0), \end{aligned} \quad (2.1)$$

where $C \perp\!\!\!\perp (U, V)$, X° and Y° represent the latent continuous variables corresponding to the observed binary outcomes X and Y , and $\mathbb{I}(\cdot)$ is the indicator function that maps the latent variables to binary responses. We have provided a discussion on the presence of the latent continuous variables in Section E of the Supplementary Materials.

The structural equation (2.1) characterizes the bidirectional causal relationship at equilibrium. We provide a detailed discussion of the bidirectional causal mechanism and the causal interpretation of the parameters $\beta_{x \rightarrow y}$ and $\beta_{y \rightarrow x}$ in Sections F and G of the Supplementary Materials. When $\beta_{y \rightarrow x} = 0$, Model (2.1) reduces to a unidirectional causal model.

Model (2.1) presents substantial challenges due to the coexistence of

2.2 Assumptions

the bidirectional causation, latent variable structures, and unobserved confounding. First, the reciprocal dependence between the latent variables X° and Y° forms a simultaneous equation system, complicating identification and estimation, as standard regression techniques fail to recover true causal effects. Second, the observed binary variables X and Y are thresholded versions of continuous latent variables, which necessitates additional distributional assumptions to characterize the dependence structure between the observed variables and latent variables. Third, the unobserved confounders U and V that simultaneously affect both equations introduce endogeneity, leading to biased estimates unless properly addressed.

To establish identification results, the next subsection introduces the structural and IV assumptions. These assumptions form the basis for the estimation procedures developed in this paper. Section 3 further conducts sensitivity analyses to examine the robustness of our conclusions to possible violations of key assumptions.

2.2 Assumptions

The covariate vector C includes two IVs, Z and W , each influencing a single causal direction. Specifically, Z only directly affects X° and W only directly affects Y° . The remaining components of C , denoted by Q ,

2.2 Assumptions

represent other observed covariates that may impact both X and Y . For notational simplicity, we omit Q in the following discussion. Figure 1a illustrates the structure of the bidirectional causal framework for binary variables with IVs Z and W . By incorporating Z and W , the bidirectional causal Model (2.1) can be reformulated as the following form:

$$\begin{aligned} X^\circ &= \mu_{x0} + \beta_{y \rightarrow x} Y^\circ + \mu_{xz} Z + U, \quad X = \mathbb{I}(X^\circ > 0), \\ Y^\circ &= \mu_{y0} + \beta_{x \rightarrow y} X^\circ + \mu_{yw} W + V, \quad Y = \mathbb{I}(Y^\circ > 0). \end{aligned} \quad (2.2)$$

The parameters μ_{xz} and μ_{yw} in Model (2.2) characterize the effects of Z and W on the latent variables X° and Y° , respectively. From Model (2.2), it is clear that Z and W serve as valid IVs for each causal direction. Throughout this paper, we maintain this terminology for IV in the bidirectional framework because they satisfy the exclusion restriction (ER) by having no direct effects on the outcome variables, analogous to their role in traditional unidirectional models (Angrist et al., 1996). This terminology is consistent with other bidirectional causal inference literature (Xue and Pan, 2020; Li and Ye, 2024; Xie et al., 2024).

In our data illustration, the binary variables X and Y in Figure 1a represent two observed binary indicators: X indicates whether an individual has coronary heart disease (or myocardial infarction), and Y indicates whether an individual has diabetes. The variables X° and Y° correspond

2.2 Assumptions

to their underlying continuous latent variables. In this context, “whether an individual has had a stroke” can be used as the IV Z for X° , and “body mass index (BMI)” can be used as the IV W for Y° . This choice is based on the following knowledge: stroke will possibly have direct effect on heart disease but unlikely has a direct influence on diabetes, while BMI can directly affects diabetes and is less likely to directly impact the heart disease. Also, the covariates Q can represent other background characteristics such as age, gender, and race, while U and V denotes unobserved confounding factors not captured in the dataset, such as genetic predisposition and lifestyle factors. Based on our setting, we need to further impose the following normality assumption to achieve identifiability.

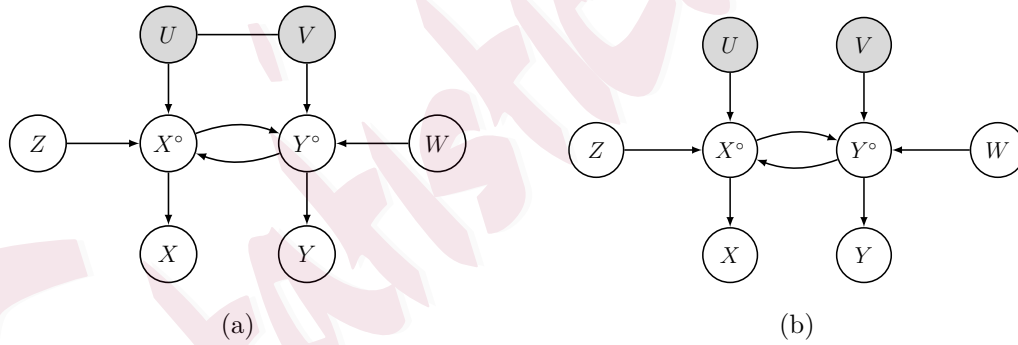


Figure 1: Graphic illustration of Model (2.2) with IVs Z and W .

Assumption 1 (Restrictive normality). (i) $V, U \sim N(0, \sigma^2)$, where σ is an unknown constant, (ii) $\text{Cov}(V, U) = 0$.

2.2 Assumptions

Assumption 1 imposes the distributional restrictions on U and V to facilitate identification and simplify the modeling framework. Assumption 1(i) assumes that U and V both follow normal distributions with zero means and equal variances. Assumption 1(ii) further assumes that U and V are uncorrelated, i.e., mutually independent in the normal setting, and the endogeneity in the bidirectional causal model (2.2) for each direction arises only from reverse causality. Figure 1b shows the causal diagram that satisfies Assumption 1. This assumption is primarily introduced to simplify computation. Under this framework, Section 2.3 discusses the identification of causal effects within the bidirectional causal framework. Additionally, Section 3.1 presents the sensitivity analysis results under more general settings, evaluating the possible violations of Assumption 1.

Before formally introducing the identification results, we firstly rewrite the structural equation Model (2.2) in the following reduced form when

$\beta_{x \rightarrow y} \beta_{y \rightarrow x} \neq 1$:

$$X^\circ = \theta_{x0} + \theta_{xz}Z + \theta_{xw}W + \theta_{xu}U + \theta_{xv}V, \quad X = \mathbb{I}(X^\circ > 0), \quad (2.3)$$

$$Y^\circ = \theta_{y0} + \theta_{yz}Z + \theta_{yw}W + \theta_{yu}U + \theta_{yv}V, \quad Y = \mathbb{I}(Y^\circ > 0),$$

where $c = 1/(1 - \beta_{x \rightarrow y} \beta_{y \rightarrow x})$, $\theta_{x0} = c(\mu_{x0} + \beta_{y \rightarrow x} \mu_{y0})$, $\theta_{xz} = c\mu_{xz}$, $\theta_{xw} = c\beta_{y \rightarrow x} \mu_{yw}$, $\theta_{xu} = c$, $\theta_{xv} = c\beta_{y \rightarrow x}$; $\theta_{y0} = c(\beta_{x \rightarrow y} \mu_{x0} + \mu_{y0})$, $\theta_{yz} = c\beta_{x \rightarrow y} \mu_{xz}$, $\theta_{yw} = c\mu_{yw}$, $\theta_{yu} = c\beta_{x \rightarrow y}$, and $\theta_{yv} = c$.

2.3 Identification

Under the structural assumptions and Assumption 1, the conditional probabilities of the two binary variables X and Y given the two observed IVs Z and W can be derived through their latent continuous representations X° and Y° . Under Model (2.3), we can derive the two observed conditional probabilities as follows, with $\Phi(\cdot)$ being the cumulative distribution function (CDF) of standard normal distribution:

$$\begin{aligned}\text{pr}(X = 1 \mid Z, W) &= \text{pr}(X^\circ > 0 \mid Z, W) = \Phi(\xi_{x0} + \xi_{xz}Z + \xi_{xw}W), \\ \text{pr}(Y = 1 \mid Z, W) &= \text{pr}(Y^\circ > 0 \mid Z, W) = \Phi(\xi_{y0} + \xi_{yz}Z + \xi_{yw}W),\end{aligned}\quad (2.4)$$

where $\lambda_1 = \text{Var}(\theta_{xu}U + \theta_{xv}V)$, $\xi_{x0} = \theta_{x0}/\sqrt{\lambda_1}$, $\xi_{xz} = \theta_{xz}/\sqrt{\lambda_1}$, $\xi_{xw} = \theta_{xw}/\sqrt{\lambda_1}$, and λ_2 , ξ_{y0} , ξ_{yz} , and ξ_{yw} are similarly denoted. Expressions in Model (2.4) depend only on the observed variables Z and W , allowing us to estimate the parameters directly using observational data. We summarize the identification results in Proposition 1.

Proposition 1. *Under Model (2.2) and Assumption 1, the bidirectional causal effects are identifiable:*

$$\beta_{x \rightarrow y} = \frac{\xi_{yz}}{\xi_{xz}} \sqrt{\frac{\xi_{xw}^2 \xi_{xz}^2 - \xi_{yw}^2 \xi_{xz}^2}{\xi_{yz}^2 \xi_{yw}^2 - \xi_{yw}^2 \xi_{xz}^2}}, \quad \beta_{y \rightarrow x} = \frac{\xi_{xw}}{\xi_{yw}} \sqrt{\frac{\xi_{yz}^2 \xi_{yw}^2 - \xi_{xz}^2 \xi_{yw}^2}{\xi_{xw}^2 \xi_{xz}^2 - \xi_{xz}^2 \xi_{yw}^2}}. \quad (2.5)$$

where the terms ξ_{xz} , ξ_{yz} , ξ_{xw} , and ξ_{yw} are the coefficients in Model (2.4).

2.3 Identification

The detailed proof of Proposition 1 is provided in Section A.1 of the Supplementary Materials. Proposition 1 gives the identification formulas for the causal parameters $\beta_{x \rightarrow y}$ and $\beta_{y \rightarrow x}$ based on the identifiable coefficients from Model (2.4). As shown in Equation (2.5), the introduction of two IVs is essential, as they differentially affect the binary variables X and Y . This difference is captured through the coefficients ξ_{xz} , ξ_{yz} , ξ_{xw} , and ξ_{yw} in Probit models. In addition, assuming zero covariance between U and V in Assumption 1 simplifies the parameters in Model (2.4), yielding a symmetric and simple identification formula (2.5). Proposition 1 can be interpreted as the identification result when the two unobserved confounders are uncorrelated. Furthermore, it can be seen from Equation (2.5) that the omission of the baseline covariate Q does not affect identifiability. In the numerical simulations in Section 4, we will conduct experiments based on the full model that includes the baseline covariate. In Section 3.1, we will extend this framework to the scenario depicted in Figure 1a, which incorporates the correlation between confounders, thereby deriving more general identification results. Furthermore, we provide Algorithm S1 for estimation in Section A.2 of the Supplementary Materials and establish asymptotic properties in Proposition S1, with the corresponding proofs given in Section A.3.

We further consider extending Assumption 1 to non-Gaussian cases and conducting a robustness analysis of Assumption 1. For details, see Section H of the Supplementary Materials.

3. Sensitivity Analysis

3.1 Sensitivity analysis for Assumption 1

In Section 2.3, we have introduced a strong normality Assumption 1 for identification. Specifically, this assumption requires that (i) the unobserved confounders U and V follow normal distributions with mean zero and equal variance σ^2 ; and (ii) they are independent, i.e., $\text{Cov}(U, V) = 0$. These two requirements are rather restrictive, since in practice it is often unrealistic to assume that the unobserved confounders in the two models have identical variances and are completely independent. In this subsection, while still utilizing the IVs, we conduct a sensitivity analysis for Assumption 1. We propose the following new assumption that is weaker than Assumption 1, allowing U and V to have different variances and be potentially correlated.

Assumption 2 (Normality). (i) $V \sim N(0, \sigma^2)$ and $U \sim N(0, \gamma_1 \sigma^2)$, (ii) $\text{Cov}(V, U) = \gamma_2 \sigma^2$. The parameter σ is an unknown constant, while γ_1 and γ_2 are sensitivity parameters.

3.1 Sensitivity analysis for Assumption 1

To quantify the influence of the relationship between unobserved confounders U and V , we introduce two sensitivity parameters in Assumption 2: $\gamma_1 = \text{Var}(U)/\text{Var}(V)$ and $\gamma_2 = \text{Cov}(U, V)/\text{Var}(V)$. The parameter γ_1 measures the relative strength of the variances of U and V , γ_2 characterizes the ratio of the covariance of (U, V) to the variance of V , and the presence of γ_2 implies that the endogeneity in model (2.2) arises from both reverse causality and unobserved confounders. We will examine the robustness of the proposed method by varying these two sensitivity parameters. It is clear that the two sensitivity parameters satisfy $\gamma_1 \geq \gamma_2^2$. When $\gamma_1 = 1$ and $\gamma_2 = 0$, Assumption 2 is equivalent to Assumption 1. The two parameters enable us to investigate how the relative variances of U and V , together with their covariance $\text{Cov}(U, V)$, influence the identification and estimation of the bidirectional causal effects.

In this context, we introduce two notations: $k_1 \equiv \xi_{yz}/\xi_{xz}$ and $k_2 \equiv \xi_{xw}/\xi_{yw}$, to simplify subsequent discussion. They represent the ratios of Probit regression coefficients of X and Y on (Z, W) , respectively. The following proposition provides the identification results:

Proposition 2. *Under Model (2.2) and Assumption 2, given the sensitivity parameters γ_1 and γ_2 , the bidirectional causal effects can be obtained through*

3.1 Sensitivity analysis for Assumption 1

the following two candidate solution pairs:

$$\begin{aligned}\beta_{x \rightarrow y,1} &= \frac{\gamma_2 k_1 (k_2 - k_1) + |k_1| \sqrt{\Delta_1}}{\gamma_1 (k_1^2 - 1)}, & \beta_{y \rightarrow x,1} &= \frac{\gamma_1 k_1 k_2 (k_1^2 - 1)}{\gamma_2 k_1 (k_2 - k_1) + |k_1| \sqrt{\Delta_1}}, \\ \beta_{x \rightarrow y,2} &= \frac{\gamma_2 k_1 (k_2 - k_1) - |k_1| \sqrt{\Delta_1}}{\gamma_1 (k_1^2 - 1)}, & \beta_{y \rightarrow x,2} &= \frac{\gamma_1 k_1 k_2 (k_1^2 - 1)}{\gamma_2 k_1 (k_2 - k_1) - |k_1| \sqrt{\Delta_1}},\end{aligned}\tag{3.6}$$

where $\Delta_1 = \gamma_1 k_1^2 k_2^2 - \gamma_1 k_1^2 - \gamma_1 k_2^2 + \gamma_1 + \gamma_2^2 k_1^2 - 2\gamma_2^2 k_1 k_2 + \gamma_2^2 k_2^2$. For $j \in \{1, 2\}$, $(\beta_{x \rightarrow y,j}, \beta_{y \rightarrow x,j})$ denotes the j th candidate solution pair. This pair is admissible only if $\text{sgn}(\beta_{x \rightarrow y,j}) = \text{sgn}(k_1)$ and $\text{sgn}(\beta_{y \rightarrow x,j}) = \text{sgn}(k_2)$.

When the unobserved confounders have unequal variances and are correlated, the causal structure becomes more complex, making the bidirectional causal effects more difficult to distinguish. Proposition 2 provides at most two candidate solution pairs obtained from a quadratic equation. The existence and number of real candidate pairs depend on the discriminant Δ_1 . Given specified values of the sensitivity parameters, Δ_1 can be evaluated using the observed data, and when $\Delta_1 > 0$, the quadratic equation yields two distinct real candidate pairs. For each $j \in \{1, 2\}$, the candidate pair $(\beta_{x \rightarrow y,j}, \beta_{y \rightarrow x,j})$ is admissible only if $\text{sgn}(\beta_{x \rightarrow y,j}) = \text{sgn}(k_1)$ and $\text{sgn}(\beta_{y \rightarrow x,j}) = \text{sgn}(k_2)$. Thus, the sign restrictions are applied separately to the two candidate pairs to determine which of them are admissible. If exactly one candidate pair satisfies these restrictions, the bidirectional causal effects are uniquely identified. Otherwise, all admissible candidate pairs

3.2 Further sensitivity analysis for ER assumption

constitute the identification results.

In practice, by varying the sensitivity parameters γ_1 and γ_2 , we can investigate how changes in the confounding structure influence the identified bidirectional causal effects. The proof and further discussion of Proposition 2 are provided in Section B of the Supplementary Materials.

Similarly, we can construct plugging-in based estimators and establish the corresponding asymptotic theory by following the approach in Section A.2 of the Supplementary Materials. For simplicity, we omit the details.

3.2 Further sensitivity analysis for ER assumption

Selecting valid IVs that satisfy Model (2.2) poses significant challenges in practice. A key concern is that the IV Z may have direct effects on Y° , and W may have direct effects on X° . To address this issue, we extend the sensitivity analysis framework introduced in Section 3.1 by incorporating sensitivity parameters for the IVs and evaluating the robustness of our results to possible violations of the exclusion restriction assumption.

In addition to the sensitivity parameters γ_1 and γ_2 in Assumption 2, we introduce two additional sensitivity parameters, η and δ , to account for potential violations of the IVs. These parameters represent the extent to which the instrumental variables Z and W exert direct effects on Y° and

3.2 Further sensitivity analysis for ER assumption

X° , respectively. The modified structural model is given by:

$$\begin{aligned} X^\circ &= \mu_{x0} + \beta_{y \rightarrow x} Y^\circ + \mu_{xz} Z + \delta W + U, \quad X = \mathbb{I}(X^\circ > 0), \\ Y^\circ &= \mu_{y0} + \beta_{x \rightarrow y} X^\circ + \eta Z + \mu_{yw} W + V, \quad Y = \mathbb{I}(Y^\circ > 0). \end{aligned} \quad (3.7)$$

When the sensitivity parameters η and δ are set to zero, Model (3.7) reduces to Model (2.2). Given Assumption 2, we can derive a Probit model similar to Model (2.4). While the coefficients are the same as in the previous model, as they correspond to the regression coefficients in the Probit model, they now involve more complex expressions due to the additional direct effects η and δ introduced in Model (3.7). To illustrate these differences in parameter complexity, we present the key parameter expressions below:

$$\text{pr}(X = 1 \mid Z, W) = \text{pr}(X^\circ > 0 \mid Z, W) = \Phi(\xi_{x0} + \xi_{xz} Z + \xi_{xw} W),$$

$$\text{pr}(Y = 1 \mid Z, W) = \text{pr}(Y^\circ > 0 \mid Z, W) = \Phi(\xi_{y0} + \xi_{yz} Z + \xi_{yw} W),$$

where $\xi_{x0} = \theta_{x0}/\sqrt{\lambda_1}$, $\xi_{xz} = \theta_{xz}^*/\sqrt{\lambda_1}$, $\xi_{xw} = \theta_{xw}^*/\sqrt{\lambda_1}$, $\theta_{xz}^* = c(\mu_{xz} + \beta_{y \rightarrow x} \eta)$, $\theta_{xw}^* = c(\delta + \beta_{y \rightarrow x} \mu_{yw})$, $\theta_{yz}^* = c(\beta_{x \rightarrow y} \mu_{xz} + \eta)$, and $\theta_{yw}^* = c(\beta_{x \rightarrow y} \delta + \mu_{yw})$ and similarly for ξ_{y0} , ξ_{yz} , and ξ_{yw} . The coefficients c , λ_1 , λ_2 , θ_{x0} , θ_{y0} , θ_{xu} , θ_{xv} , θ_{yu} , and θ_{yv} are already defined in (2.3) and (2.4).

Given the sensitivity parameters η and δ , to establish identifiable results, we introduce reparameterizations which will be used in the following theoretical results. We define $\eta_0 = \eta/\mu_{xz}$, which represents the relative strength of the effect of Z on (X°, Y°) , and $\delta_0 = \delta/\mu_{yw}$, which represents

3.2 Further sensitivity analysis for ER assumption

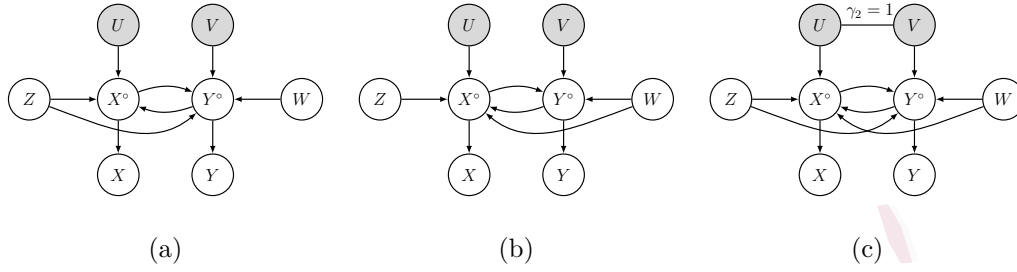


Figure 2: Three graphical illustrations of Model (3.7).

the relative strength of the effect of W on (X°, Y°) .

Building on Proposition 2, we further extend our identification strategy to enable more comprehensive sensitivity analysis by incorporating direct effect parameters (η_0, δ_0) under Model (3.7) (see Proposition S2 in Section C.1 of Supplementary Materials).

However, solving the quartic equation in Proposition S2 poses significant computational challenges in practical applications. Intuitively, this difficulty arises because the direct effects of Z and W on X° and Y° are inherently intertwined, making it challenging to disentangle their individual contributions to the causal effects. To make the sensitivity analysis more tractable and practically interpretable, we examine three special cases that yield simplified analytical expressions, thereby facilitating implementation in empirical applications. Figure 2 illustrates the causal frameworks under the sensitivity analysis settings specified in the following three corollaries.

3.2 Further sensitivity analysis for ER assumption

Corollary 1. *Under Assumption 2, with the settings $\delta_0 = 0$, $\gamma_1 = 1$, $\gamma_2 = 0$, and an arbitrary η_0 (illustrated in Figure 2a), the bidirectional causal effects can be obtained through the following two candidate solution pairs:*

$$\begin{aligned}\beta_{y \rightarrow x,1} &= \frac{-s_2 + \sqrt{s_2^2 - 4s_1s_3}}{2s_1}, & \beta_{x \rightarrow y,1} &= \eta_0(t_1t_2 - 1) + \frac{2s_1t_1t_2}{-s_2 + \sqrt{s_2^2 - 4s_1s_3}}, \\ \beta_{y \rightarrow x,2} &= \frac{-s_2 - \sqrt{s_2^2 - 4s_1s_3}}{2s_1}, & \beta_{x \rightarrow y,2} &= \eta_0(t_1t_2 - 1) + \frac{2s_1t_1t_2}{-s_2 - \sqrt{s_2^2 - 4s_1s_3}},\end{aligned}$$

where $t_1 = \xi_{xw}/\xi_{xz}$, $t_2 = \xi_{yz}/\xi_{yw}$, $t_3 = \xi_{xz}/\xi_{yz}$, $t_4 = \xi_{xw}/\xi_{yw}$, $s_1 = \eta_0^2(t_1t_2 - 1)^2 - t_4^2 + 1$, $s_2 = 2t_1t_2\eta_0(t_1t_2 - 1)$, and $s_3 = t_1^2t_2^2 - t_4^2$. For each $j \in \{1, 2\}$, $(\beta_{x \rightarrow y,j}, \beta_{y \rightarrow x,j})$ denotes the j th candidate solution pair. The j th candidate pair is admissible only if $\text{sgn}(\beta_{y \rightarrow x,j}) = \text{sgn}(t_4)$.

Corollary 2. *Under Assumption 2, with the settings $\eta_0 = 0$, $\gamma_1 = 1$, $\gamma_2 = 0$, and an arbitrary δ_0 (illustrated in Figure 2b), the bidirectional causal effects can be obtained through the following two candidate solution pairs:*

$$\begin{aligned}\beta_{x \rightarrow y,1} &= \frac{-s_5 + \sqrt{s_5^2 - 4s_4s_6}}{2s_4}, & \beta_{y \rightarrow x,1} &= \delta_0(t_1t_2 - 1) + \frac{2s_4t_1t_2}{-s_5 + \sqrt{s_5^2 - 4s_4s_6}}, \\ \beta_{x \rightarrow y,2} &= \frac{-s_5 - \sqrt{s_5^2 - 4s_4s_6}}{2s_4}, & \beta_{y \rightarrow x,2} &= \delta_0(t_1t_2 - 1) + \frac{2s_4t_1t_2}{-s_5 - \sqrt{s_5^2 - 4s_4s_6}},\end{aligned}$$

where t_1 , t_2 , t_3 and t_4 are already defined in Corollary 1, $s_4 = t_3t_4 + t_3t_4\delta_0^2(t_1t_2 - 1)^2 - t_1t_2$, $s_5 = 2t_4\delta_0(t_1t_2 - 1)$, and $s_6 = t_1t_2(t_4^2 - 1)$. For each $j \in \{1, 2\}$, $(\beta_{x \rightarrow y,j}, \beta_{y \rightarrow x,j})$ denotes the j th candidate solution pair. The j th candidate pair is admissible only if $\text{sgn}(\beta_{x \rightarrow y,j}) = \text{sgn}(t_3)$.

3.2 Further sensitivity analysis for ER assumption

Corollary 1 examines the identification of the bidirectional causal effects when Z has a direct effect on Y° , under the assumptions that W is a valid instrumental variable and that the unobserved confounders U and V are independent. Corollary 2 considers the corresponding setting in which W has a direct effect on X° . Each corollary yields at most two candidate solution pairs. In Corollary 1, a candidate pair indexed by j is admissible only if $\text{sgn}(\beta_{y \rightarrow x, j}) = \text{sgn}(t_4)$, whereas in Corollary 2, it is admissible only if $\text{sgn}(\beta_{x \rightarrow y, j}) = \text{sgn}(t_3)$. If exactly one candidate pair satisfies the sign restriction, the bidirectional causal effects are uniquely identified. This approach is motivated by sensitivity analyses for violations of the exclusion restriction in traditional IV frameworks (Angrist et al., 1996).

Beyond the scenarios in Corollaries 1 and 2, we explore cases where the unobserved confounders U and V are correlated, while simultaneously allowing both Z and W to have direct effects on X° and Y° .

Corollary 3. *We introduce the signal-to-noise ratios $\eta_0 = \eta/\sigma$ and $\delta_0 = \delta/\sigma$ to reparameterize η and δ in Model (3.7). We then consider the setting with $\gamma_1 = 1$, $\gamma_2 = 1$, and fixed values of δ_0 and η_0 (as illustrated in Figure 2c). Suppose that both $\beta_{x \rightarrow y} > -1$ and $\beta_{y \rightarrow x} > -1$, that $\beta_{x \rightarrow y}\beta_{y \rightarrow x} \neq 1$. If $\beta_{x \rightarrow y}\beta_{y \rightarrow x} < 1$, the identification expressions for the bidirectional causal*

3.2 Further sensitivity analysis for ER assumption

effects are given by:

$$\beta_{y \rightarrow x} = \frac{(\xi_{yz} - \xi_{xz})(\xi_{xw} - \delta_0)}{(\xi_{xw} - \xi_{yw})(\xi_{xz} - \eta_0)}, \quad \beta_{x \rightarrow y} = \frac{(\xi_{xw} - \xi_{yw})(\xi_{yz} - \eta_0)}{(\xi_{yz} - \xi_{xz})(\xi_{yw} - \delta_0)}. \quad (3.8)$$

If $\beta_{x \rightarrow y}\beta_{y \rightarrow x} > 1$, the identification expressions for the bidirectional causal effects are given by:

$$\beta_{y \rightarrow x} = \frac{(\xi_{yz} - \xi_{xz})(\xi_{xw} + \delta_0)}{(\xi_{xw} - \xi_{yw})(\xi_{xz} + \eta_0)}, \quad \beta_{x \rightarrow y} = \frac{(\xi_{xw} - \xi_{yw})(\xi_{yz} + \eta_0)}{(\xi_{yz} - \xi_{xz})(\xi_{yw} + \delta_0)}. \quad (3.9)$$

Equations (3.8) and (3.9) each yield one candidate pair. They provide identification expressions under different ranges of the causal effect parameters. In fact, except for values that make the denominators undefined, namely $\beta_{y \rightarrow x} = -1$, $\beta_{x \rightarrow y} = -1$, and $\beta_{y \rightarrow x}\beta_{x \rightarrow y} = 1$, identification expressions exist for all other parameter regions. More detailed results are presented in Section C.4 of the Supplementary Materials. Although multiple pairs may arise during the identification process due to the sign configurations of the parameters, under each sign configuration the identification expression is unique and has a neat closed-form expression; these solutions together constitute the complete set of candidate pairs. Therefore, with further prior information, for different signal-to-noise ratios, we can compute the causal effects based on the candidate pairs and subsequently perform selection to achieve more accurate identification. Figure S1 of the Supplementary Materials illustrates the regions of these configurations.

Sections C.2–C.4 of the Supplementary Materials provide the detailed proofs of Corollaries 1–3, while Section C.5 presents the discussion of Corollary 3 under the alternative parameterization based on η/μ_{xz} and δ/μ_{yw} .

4. Numerical Experiments

4.1 Simulation studies

In this section, we conduct numerical simulations to evaluate the finite sample performance of the proposed method in Section 2. We generate observed covariate $Q \sim N(0, 1)$ and unobserved confounders $U, V \sim N(0, 0.75^2)$. Instrumental variables Z and W follow either $N(0, 1)$ (Scenario 1) or $\text{Unif}(-1, 1)$ (Scenario 2). Latent variables X° and Y° are generated from Model (2.2) with parameters $\mu_{x0} = \mu_{y0} = 0$, $\beta_{y \rightarrow x} = 0.45$, $\beta_{x \rightarrow y} = -0.25$, $\mu_{xz} = \mu_{yw} = 0.65$, $\mu_{xq} = \mu_{yq} = 0.15$, and then thresholded to obtain binary X and Y .

For the two data-generating scenarios, we consider two estimation methods with sample sizes $N = 2000, 5000$, and 10000 . The first method (GLS) is the direct regression approach, where we fit two Probit models: one by regressing Y on (X, Z, W, Q) , and another by regressing X on (Y, Z, W, Q) , the resulting coefficients on X and Y are then interpreted as estimates of $\hat{\beta}_{x \rightarrow y}$ and $\hat{\beta}_{y \rightarrow x}$, respectively. The second method (IV) is Algorithm S1 in

4.1 Simulation studies

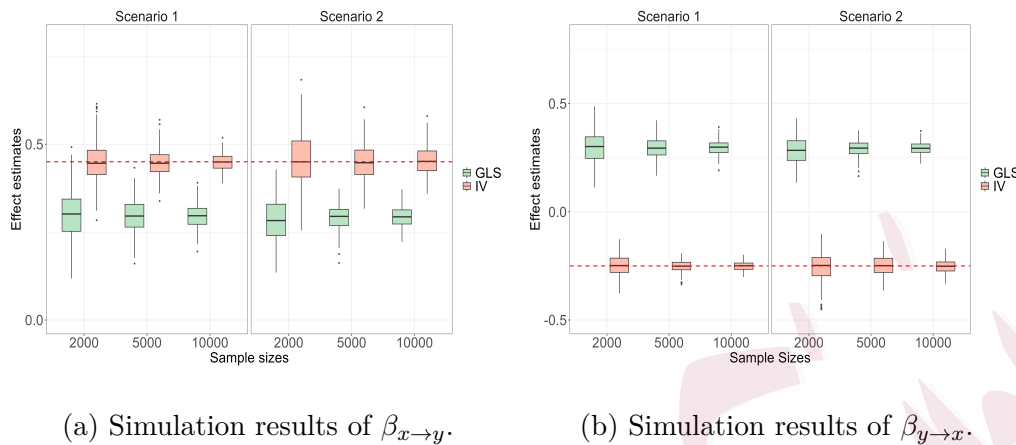


Figure 3: Boxplot of simulation results under bidirectional framework.

Section A.2 of the Supplementary Materials. We perform Probit regression on X , Y , Z , W and Q to estimate the coefficients in Equation (2.4), namely, ξ_{xz} , ξ_{xw} , ξ_{yz} and ξ_{yw} . These estimated coefficients are then substituted into the identification formula (2.5) to obtain the estimates $\hat{\beta}_{x \rightarrow y}$ and $\hat{\beta}_{y \rightarrow x}$.

The estimation results show that the proposed method performs well across both scenarios, yielding consistent estimates. In contrast, the direct regression approach produces estimates that deviate significantly from the true values in both scenarios. Furthermore, we compute the standard deviation and bias of the proposed estimators over 200 repeated simulations. Figure 3 presents the simulation results for the bidirectional causal effect, where the red represents our proposed method and the green represents the naive regression method. The corresponding numerical results are summa-

4.2 Sensitivity analysis for Assumption 1

rized in Table 1. It can be observed that the standard deviation and bias of the estimators decrease as the sample size increases across both scenarios.

Table 1: Standard deviation and bias of estimators across scenarios and methods. All values are multiplied by 10^4 . Detailed results are provided in Section D.1 of the Supplementary Materials.

		Scenario 1 (GLS)		Scenario 2 (GLS)		Scenario 1 (IV)		Scenario 2 (IV)	
	N	2000	10000	2000	10000	2000	10000	2000	10000
$\hat{\beta}_{x \rightarrow y}$	bias	-1510	-1540	-1640	-1570	-2	10	80	-8
	sd	724	358	635	288	535	242	941	407
$\hat{\beta}_{y \rightarrow x}$	bias	5490	5460	5370	5440	-45	-7	10	-31
	sd	718	355	637	289	500	208	755	354

4.2 Sensitivity analysis for Assumption 1

In this section, we conduct a sensitivity analysis of Assumption 1, focusing on the influence of both the variance relationship and covariance of the unobserved confounders U and V on the estimation results. Accordingly, we modify the second step of the data generation process in Section 4.1 as follows: Under Assumption 2, we generate the unobserved confounders U and V , where $U \sim N(0, \gamma_1 \sigma^2)$, $V \sim N(0, \sigma^2)$, and $\text{cov}(U, V) = \gamma_2 \sigma^2$, with $\sigma = 0.75$. The parameter γ_1 varies within the range $(0.1, 1)$ at intervals of

4.2 Sensitivity analysis for Assumption 1

0.02, while γ_2 varies within the range $(-0.6, 0.6)$ at intervals of 0.02. All other settings remain consistent with Section 4.1.

For each parameter combination (γ_1, γ_2) and each of the 200 simulated datasets, we calculate the two candidate solution pairs given in Proposition 2 and retain those satisfying the sign restrictions $\text{sgn}(\beta_{x \rightarrow y, j}) = \text{sgn}(k_1)$ and $\text{sgn}(\beta_{y \rightarrow x, j}) = \text{sgn}(k_2)$. The estimates reported below refer to the resulting admissible candidate pairs. Figure 4 presents the bias and standard deviation of causal effect estimates across different sensitivity parameter values for a sample size of $N = 10000$. The black curve corresponds to the truncation of the parabola defined by $\gamma_1 = \gamma_2^2$. As stated in Proposition 2, within the parabola, specific point estimates of the causal effects can be obtained once the two sensitivity parameter values are determined.

The first row of Figure 4 presents the bias of the bidirectional causal effect estimates $\beta_{x \rightarrow y}$ and $\beta_{y \rightarrow x}$ across different combinations of sensitivity parameters (γ_1, γ_2) . The gradient from light to dark corresponds to values from large to small. Across a broad sensitivity range, the estimation bias remains small, around 0.002 (mainly shown in dark colors), with only minor fluctuations. The second row displays the standard deviations of the estimates, with the color gradient again corresponding to their values. Although the standard deviations vary across parameter values, they remain

4.3 Sensitivity analysis simulation for ER assumption

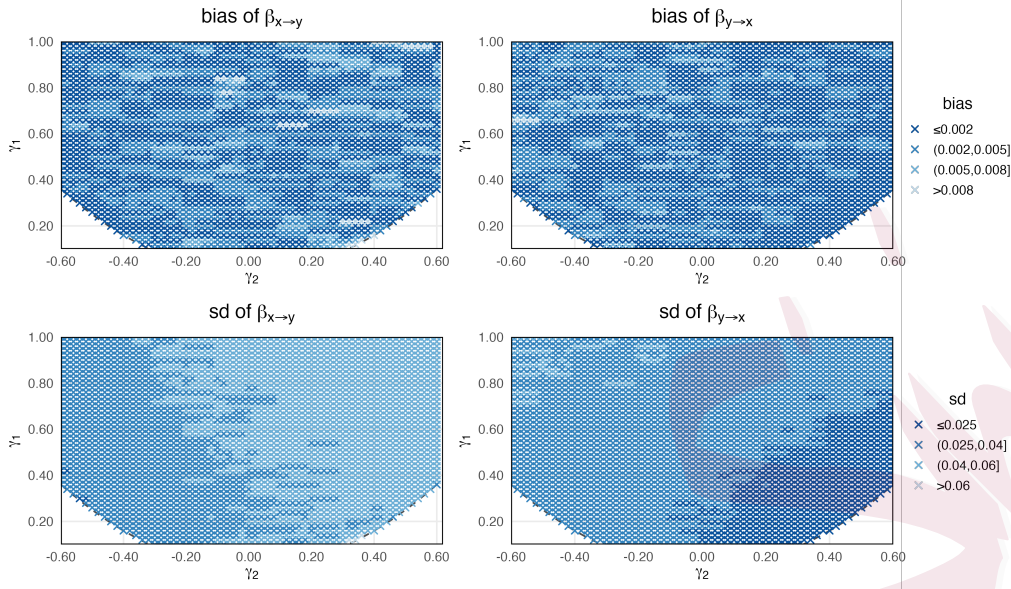


Figure 4: Bias and standard deviation estimation results.

consistently below 0.06, indicating the robustness of the method.

4.3 Sensitivity analysis simulation for ER assumption

This section conducts numerical simulations for the three instrumental variable sensitivity analysis settings discussed in Section 3.2. We assess how violating Model (2.2) by introducing direct effects from instrumental variables to target variables impacts causal effect estimation under finite samples. The simulation setup follows Section 4.1, with modifications corresponding to Corollaries 1–3. We consider three cases: Case 1 (Corollary 1) varies $\eta_0 \in [-0.16, 0.16]$ with $\delta_0 = 0$, $\gamma_1 = 1$, $\gamma_2 = 0$; Case 2 (Corollary

4.3 Sensitivity analysis simulation for ER assumption

2) varies $\delta_0 \in [-0.16, 0.16]$ with $\eta_0 = 0$, $\gamma_1 = 1$, $\gamma_2 = 0$; Case 3 (Corollary 3) varies $\eta_0 = \delta_0 \in [-0.16, 0.16]$ with $\gamma_1 = \gamma_2 = 1$. All sensitivity parameters vary by increments of 0.02.

For Cases 1 and 2, the two candidate solution pairs are calculated for each value of the sensitivity parameter. In Case 1, we retain the candidate pairs satisfying $\text{sgn}(\beta_{y \rightarrow x, j}) = \text{sgn}(t_4)$, whereas in Case 2, we retain the candidate pairs satisfying $\text{sgn}(\beta_{x \rightarrow y, j}) = \text{sgn}(t_3)$. The estimates reported below correspond to the resulting admissible candidate pairs.

Based on the simulation results from 200 repeated simulations under different combinations of parameters and data-generating scenarios, we plot error bar graphs in Figures 5. The red dashed lines represent the given true value of the causal effect parameter. It can be observed that across different parameter values, the point estimates under both data-generating scenarios remain consistently stable around the true value. Furthermore, when the arbitrary sensitivity parameters in the first two settings are changed to zero, the estimation results are equivalent to those under the specification of Model (2.2). Simultaneously, as the sample size increases, the estimation error decreases accordingly. These findings demonstrate that the proposed method retains its robustness even when the instrumental variables Z and W violate the assumptions of Model (2.2) to some extent. Here, we present

the results for $\beta_{x \rightarrow y}$ under three settings with a sample size of $N = 10000$. Simulation results for other sample sizes and for $\beta_{y \rightarrow x}$ are provided in detail in Section D.2 of the Supplementary Materials.

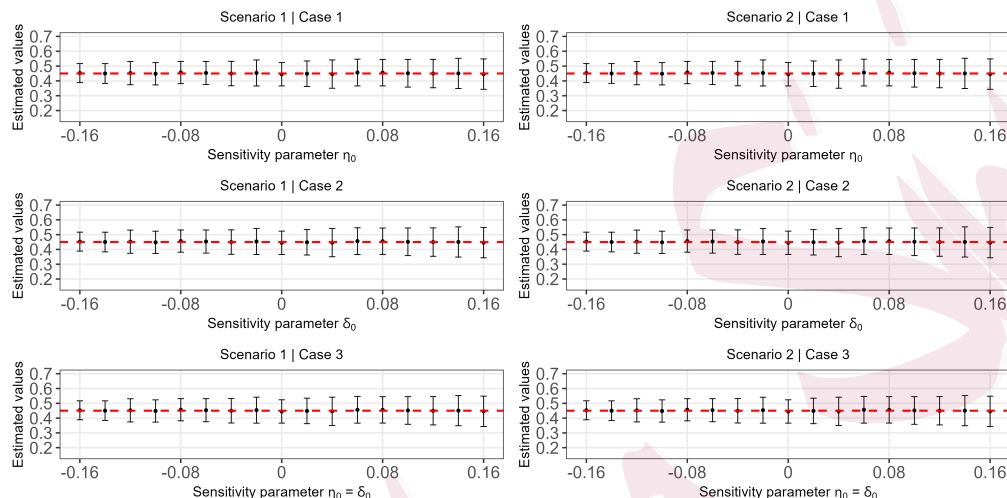


Figure 5: The point estimate and 95% confidence interval results of $\beta_{x \rightarrow y}$.

5. Application

5.1 Description

According to the Centers for Disease Control and Prevention (CDC), heart disease is the leading cause of death among most racial groups in the United States, including African Americans, American Indians/Alaska Natives, and Whites. In addition, nearly half (47%) of US adults have at least one of the three major risk factors: high blood pressure, high cholesterol, or

5.2 Analysis under the assumptions in Section 2.3

smoking. Other important risk factors include diabetes, obesity (high body mass index, BMI), physical inactivity, and excessive alcohol consumption.

In this study, we apply our method to investigate the bidirectional causal relationship between heart disease and diabetes. The data come from the 2023 Behavioral Risk Factor Surveillance System (BRFSS) by the CDC, containing 319795 records and 18 variables including coronary heart disease or myocardial infarction, diabetes status, smoking, BMI, and other health factors. We focus on the bidirectional relationship between X (coronary heart disease or myocardial infarction) and Y (diabetes). The remaining 16 variables (e.g., age, gender, race) are in Table S2 of Section D.3 in the Supplementary Materials. We standardize BMI prior to analysis given its larger scale compared to binary variables.

Since stroke is clinically associated with heart disease but has limited impact on diabetes, and higher BMI increases diabetes risk while having limited effect on heart disease, we treat “stroke” as the instrumental variable Z for X° and “BMI” as the instrumental variable W for Y° .

5.2 Analysis under the assumptions in Section 2.3

For the heart disease dataset, we apply our proposed method under Assumption 1 to estimate the bidirectional causal effects between heart

5.2 Analysis under the assumptions in Section 2.3

disease and diabetes, while also providing results from the direct regression approach for comparison. Based on 200 bootstrap replications, we compute the standard deviations (SD) and 95% confidence intervals (CIs) for $\beta_{x \rightarrow y}$ and $\beta_{y \rightarrow x}$. All results are presented in Table 2.

Table 2: Estimation results of the heart disease application.

	Estimate	SD	95% CI	Estimate	SD	95% CI
	GLS estimation			IV estimation		
$\hat{\beta}_{x \rightarrow y}$	0.2510	0.0097	[0.2321, 0.2698]	0.3108	0.0262	[0.2594, 0.3621]
$\hat{\beta}_{y \rightarrow x}$	0.2554	0.0090	[0.2379, 0.2729]	0.1786	0.0154	[0.1484, 0.2088]

We find that both point estimates are positive, indicating a bidirectional positive association between X and Y , that is, coronary heart disease or myocardial infarction may have a significant positive effect on diabetes, and vice versa. The 95% confidence intervals obtained via bootstrap do not include zero, further supporting the presence of statistically significant positive effects in both directions. These findings imply that the onset of heart disease may elevate the risk of developing diabetes, potentially due to physiological stress, inflammation, or lifestyle changes following a cardiovascular event. Conversely, individuals with diabetes may face an increased risk of heart disease, consistent with existing clinical evidence

5.3 Sensitivity analysis for Assumption 1

that links poor glycemic control to vascular damage and atherosclerosis (Brownlee, 2001). In addition, the direct regression approach yields point estimates with the same direction of effect, suggesting that the estimation results obtained using our proposed method are robust and reliable.

5.3 Sensitivity analysis for Assumption 1

In practice, it is challenging to determine the actual relative strength of unobserved confounding affecting the two target variables – heart disease” and diabetes”. Therefore, parallel to Section 4.2, we extend the study based on Assumption 1 in Section 5.2 by introducing parameters γ_1 and γ_2 to conduct a sensitivity analysis on this dataset. This aims to evaluate the accuracy of the proposed method in assessing possible real-world confounding scenarios. We vary γ_1 over $[0.5, 3]$ at intervals of 0.05, and γ_2 over $[-1.5, 1.5]$ at intervals of 0.1. For each parameter combination (γ_1, γ_2) , we calculate the two candidate solution pairs given in Proposition 2 and retain those satisfying the corresponding sign restrictions. The resulting admissible estimates are displayed in Figure 6. Figure 6 uses color shading to represent the magnitudes and variations of these estimates.

As can be seen from Figure 6, under different values of the parameter combinations (γ_1, γ_2) , the point estimates obtained from the empirical data

5.3 Sensitivity analysis for Assumption 1

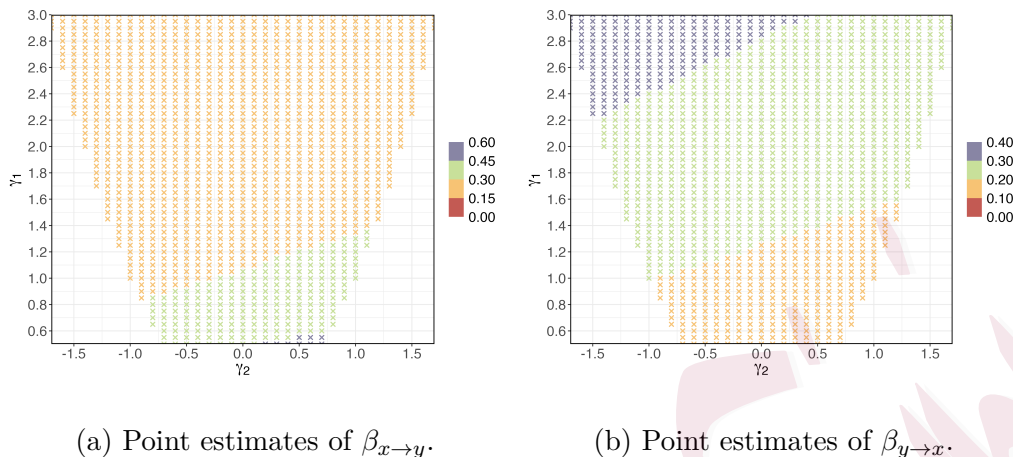


Figure 6: Sensitivity analysis results for various parameter settings with the heart disease and diabetes dataset.

are consistently positive. Moreover, according to the numerical range displayed in the figure, the range of the estimates when the absolute values of both parameters are close to zero is relatively consistent with the confidence interval obtained in Section 5.2. Overall, given the limited information on unobserved confounders, the method proposed in this paper can provide a reliable range for causal effect estimates, offering a practical reference value for real-world applications. Moreover, if more comprehensive information on confounders becomes available, such as knowledge from domain experts regarding the relative strength of the influence of unobserved confounders on heart disease and diabetes, the range of the causal effect point estimates can be further narrowed.

5.4 Sensitivity analysis for ER assumption

5.4 Sensitivity analysis for ER assumption

In Sections 5.2 and 5.3, we estimate the bidirectional causal effects between heart disease and diabetes and conduct corresponding sensitivity analyses under Assumption 1 and its weaker version, respectively. These investigations are carried out within the framework of Model (2.2), wherein the selected variables—“stroke” and “BMI”—are assumed to be valid instrumental variables for heart disease and diabetes, respectively. Therefore, building upon Model (3.7), we further incorporate a sensitivity analysis addressing potential violations of the ER assumption for IVs in this dataset. This analysis follows the three corollaries presented in Section 3.2. Similar to the numerical simulation approach in Section 4.3, the varying parameters across the three sensitivity parameter settings are all set to range from $[-0.16, 0.16]$ with an interval of 0.02.

Figure 7 presents the estimation results as error bar plots for three specific cases of sensitivity parameter values, with confidence intervals derived from 200 bootstrap samples. For Case 1, the sensitivity parameter η_0 has a significant impact on the causal relationship from X to Y . As η_0 increases, the effect gradually weakens, but the causal effect remains positive throughout. Similarly, η_0 has a smaller impact on the reverse causal effect from Y to X , and this effect also remains positive. For Case 2, the sensitiv-

5.4 Sensitivity analysis for ER assumption

ity parameter δ_0 has a larger impact on the causal relationship from Y to X . As δ_0 increases, the effect gradually weakens, but remains positive. In contrast, δ_0 has a smaller impact on the causal relationship from X to Y , and the effect remains positive. For Case 3, the inclusion of two sensitivity parameters affects the causal relationships in both directions. The causal effect of X on Y gradually weakens as the sensitivity parameters increase, but remains positive throughout. In contrast, the causal effect of Y on X is more substantially influenced and turns negative when the sensitivity parameter exceeds 0.04. Introducing a sensitivity parameter η_0 (or δ_0) can be interpreted as adding a causal path, which consequently influences the corresponding causal effect parameter $\beta_{x \rightarrow y}$ (or $\beta_{y \rightarrow x}$) to some extent.

Overall, whether under the change of η_0 or δ_0 , all bidirectional causal effects exhibit a positive relationship, suggesting a positive causal relationship between “heart disease” and “diabetes” based on real-world data, and the sensitivity analysis further validates the robustness of this relationship. Specifically, as the sensitivity parameters change, both the effect of heart disease on diabetes and its reverse both remain positive, indicating a significant bidirectional causal relationship. This relationship remains stable across a range of parameter values, providing reliable evidence supporting the causal relationship between heart disease and diabetes.

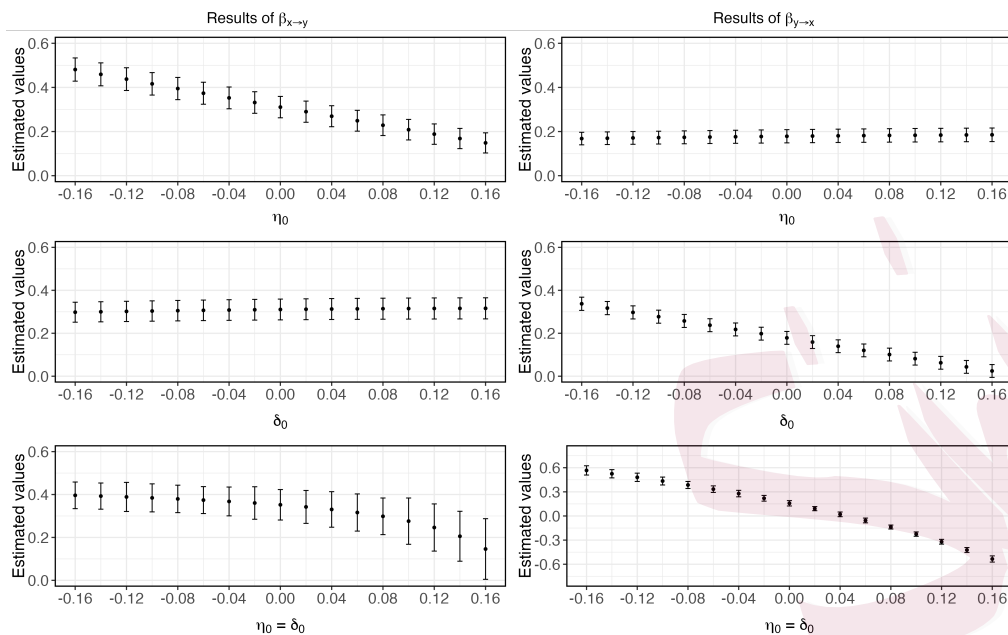


Figure 7: Sensitivity analysis results across three cases for the real dataset.

6. Summary

This paper develops a framework for evaluating bidirectional causal effects between binary variables under unobserved confounding by mapping continuous latent variables to observed binary outcomes via threshold constraints. We introduce instrumental variables to construct consistent estimators and propose a new identification strategy under the Probit model. We also relax the identification assumptions to varying degrees and conduct sensitivity analysis for the proposed method.

Several directions remain for future research. First, Clarke and Wind-

REFERENCES

meijer (2012) introduced a control function approach for binary outcomes, achieving identification of unidirectional causal effects under unobserved confounding. Extending this approach to bidirectional causal effects would be a promising direction. Second, since our method is designed for binary variables, a natural extension is to generalize the framework to multi-level settings using the ordered Probit model for categorical variables. These extensions are left for future work.

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.
- Borracci, T., Cappellini, I., Campiglia, L., Picciafuochi, F., Berti, J., Consales, G., and De Gaudio, A. (2013). Preoperative medication with oral morphine sulphate and postoperative pain. *Minerva Anestesiologica*, 79:525–533.
- Brownlee, M. (2001). Biochemistry and molecular cell biology of diabetic complications. *Nature*, 414:813–820.
- Clarke, P. S. and Windmeijer, F. (2012). Instrumental variable estimators

REFERENCES

- for binary outcomes. *Journal of the American Statistical Association*, 107(500):1638–1652.
- Cryan, J. F., O’Riordan, K. J., Cowan, C. S. M., et al. (2019). The microbiota-gut-brain axis. *Physiological Reviews*, 99(4):1877–2013.
- Davey Smith, G. and Hemani, G. (2014). Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics*, 23(R1):R89–R98.
- Engelgau, M. M., Zhang, P., Jan, S., and Mahal, A. (2019). Economic dimensions of health inequities: The role of implementation research. *Ethnicity & Disease*, 29(Suppl 1):103–112.
- Foster, J. A., Rinaman, L., and Cryan, J. F. (2017). Stress & the gut-brain axis: Regulation by the microbiome. *Neurobiology of Stress*, 7:124–136.
- Gao, X., Meng, L.-X., Ma, K.-L., et al. (2019). The bidirectional causal relationships of insomnia with five major psychiatric disorders: A mendelian randomization study. *European Psychiatry*, 60:79–85.
- Hausman, J. A. (1983). Specification and estimation of simultaneous equation models. *Handbook of econometrics*, 1:391–448.

REFERENCES

- Henning, R. J. (2018). Type-2 diabetes mellitus and cardiovascular disease. *Future Cardiology*, 14(6):491–509.
- Katakami, N. (2018). Mechanism of development of atherosclerosis and cardiovascular disease in diabetes mellitus. *Journal of Atherosclerosis and Thrombosis*, 25(1):27–39.
- Li, S. and Ye, T. (2024). A focusing framework for testing bi-directional causal effects in mendelian randomization. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 00:1–20.
- Lupparelli, M. and Mattei, A. (2020). Joint and marginal causal effects for binary non-independent outcomes. *Journal of Multivariate Analysis*, 178:104609.
- Palagini, L., Baglioni, C., Ciapparelli, A., Gemignani, A., and Riemann, D. (2013). REM sleep dysregulation in depression: State of the art. *Sleep Medicine Reviews*, 17(5):377–390.
- Riaz, N., Havel, J. J., Makarov, V., et al. (2017). Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell*, 171(4):934–949.e16.
- Smith, J. P. (1999). Healthy bodies and thick wallets: The dual relation

REFERENCES

- between health and economic status. *Journal of Economic Perspectives*, 13(2):145–166.
- Ten Have, T. R., Joffe, M., and Cary, M. (2003). Causal logistic models for non-compliance under randomized trials with non-compliance and a dichotomous outcome. *Statistics in Medicine*, 24:191–210.
- Xie, F., Yao, Z., Xie, L., Zeng, Y., and Geng, Z. (2024). Identification and estimation of the bi-directional MR with some invalid instruments. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pages 48256–48291.
- Xue, H. and Pan, W. (2020). Inferring causal direction between two traits in the presence of horizontal pleiotropy with GWAS summary data. *PLoS Genetics*, 16(11):e1009105.

Shanshan Luo

School of Mathematics and Statistics

Beijing Technology and Business University

E-mail: shanshanluo@btbu.edu.cn