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

The Supplementary Materials contain proofs of the theoretical results and additional simulation studies in the main paper.

A. Details on identification and estimation under Assumption 1

A.1 Proof of Proposition 1

Proof. Given the Model (2):

$$X^\circ = \mu_{x0} + \beta_{y\to x}Y^\circ + \mu_{xz}Z + U, \quad X = \mathbb{I}(X^\circ > 0),$$

$$Y^\circ = \mu_{y0} + \beta_{x\to y}X^\circ + \mu_{yw}W + V, \quad Y = \mathbb{I}(Y^\circ > 0),$$

we can derive the reduced form:

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

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

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

A.1 Proof of Proposition 1

Under Assumption 1, the unobserved confounders U and V are independent normal random variables with zero means and variances σ^2, which implies that the composite error terms in the reduced form follow normal distributions:

$$\theta_{xu}U + \theta_{xv}V \sim N(0, (\theta_{xu}^2 + \theta_{xv}^2)\sigma^2),$$

$$\theta_{yu}U + \theta_{yv}V \sim N(0, (\theta_{yu}^2 + \theta_{yv}^2)\sigma^2).$$

Therefore, we can derive the Probit model specifications:

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

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

where $\lambda_1 = \text{Var}(\theta_{xu}U + \theta_{xv}V) = (\theta_{xu}^2 + \theta_{xv}^2)\sigma^2$, $\lambda_2 = \text{Var}(\theta_{yu}U + \theta_{yv}V) = (\theta_{yu}^2 + \theta_{yv}^2)\sigma^2$, $\xi_{x0} = \theta_{x0}/\sqrt{\lambda_1}$, $\xi_{xz} = \theta_{xz}/\sqrt{\lambda_1}$, $\xi_{xw} = \theta_{xw}/\sqrt{\lambda_1}$; $\xi_{y0} = \theta_{y0}/\sqrt{\lambda_2}$, $\xi_{yz} = \theta_{yz}/\sqrt{\lambda_2}$, $\xi_{yw} = \theta_{yw}/\sqrt{\lambda_2}$, and we can identify the ratios ξ_{xz}, ξ_{xw}, ξ_{yz} and ξ_{yw} through the observed data. Here we focus on four key parameters:

$$\xi_{xz} = \frac{c\mu_{xz}}{|c|\sigma\sqrt{1+\beta_{y\to x}^2}}, \quad \xi_{xw} = \frac{c\beta_{y\to x}\mu_{yw}}{|c|\sigma\sqrt{1+\beta_{y\to x}^2}},$$
$$\xi_{yz} = \frac{c\beta_{x\to y}\mu_{xz}}{|c|\sigma\sqrt{1+\beta_{x\to y}^2}}, \quad \xi_{yw} = \frac{c\mu_{yw}}{|c|\sigma\sqrt{1+\beta_{x\to y}^2}}.$$

We then define the coefficient ratios: $k_1 = \xi_{yz}/\xi_{xz}$, and $k_2 = \xi_{xw}/\xi_{yw}$. These ratios can be directly estimated from the probit regression coefficients. Substituting the expressions for the standardized coefficients, we obtain:

$$k_1 \equiv \frac{\xi_{yz}}{\xi_{xz}} = \beta_{x\to y}\frac{\sqrt{1+\beta_{y\to x}^2}}{\sqrt{1+\beta_{x\to y}^2}}, \quad k_2 \equiv \frac{\xi_{xw}}{\xi_{yw}} = \beta_{y\to x}\frac{\sqrt{1+\beta_{x\to y}^2}}{\sqrt{1+\beta_{y\to x}^2}}. \tag{S1}$$

Thus, we obtain the relational equations between the causal effect parameters and k_1, k_2:

$$\begin{cases} k_1 k_2 = \beta_{x\to y}\beta_{y\to x}, \\ \\ k_1^2 = \beta_{x\to y}^2 \dfrac{1+\beta_{y\to x}^2}{1+\beta_{x\to y}^2}, \end{cases}$$

Substituting the first equation into the second, we get:

$$(k_1^2 - 1)\beta_{x\to y}^2 = k_1^2(k_2^2 - 1).$$

According to Equation (S1), we can see that $\beta_{x\to y}$ and k_1 should have the same sign. Therefore, we can solve for the bidirectional causal effects as:

$$\begin{cases} \beta_{x\to y} = \mathrm{sgn}(k_1)\sqrt{\dfrac{k_1^2(k_2^2-1)}{(k_1^2-1)}} = k_1\sqrt{\dfrac{k_2^2-1}{k_1^2-1}} = \dfrac{\xi_{yz}}{\xi_{xz}}\sqrt{\dfrac{\xi_{xw}^2\xi_{xz}^2 - \xi_{yw}^2\xi_{xz}^2}{\xi_{yz}^2\xi_{yw}^2 - \xi_{yw}^2\xi_{xz}^2}}, \\ \\ \beta_{y\to x} = k_1k_2/\beta_{x\to y} = \dfrac{k_1k_2}{\sqrt{\dfrac{k_1^2(k_2^2-1)}{(k_1^2-1)}}} = k_2\sqrt{\dfrac{k_1^2-1}{k_2^2-1}} = \dfrac{\xi_{xw}}{\xi_{yw}}\sqrt{\dfrac{\xi_{yz}^2\xi_{yw}^2 - \xi_{xz}^2\xi_{yw}^2}{\xi_{xw}^2\xi_{xz}^2 - \xi_{xz}^2\xi_{yw}^2}}. \end{cases}$$

We can identify $\beta_{y\to x}$ and $\beta_{x\to y}$ in observed data using Equation (5).

□

A.2 Estimation and inference

Proposition 1 establishes the identification formula (2.5), which further allows us to estimate the bidirectional causal effects. Accordingly, we propose a two-step estimation procedure, as described in Algorithm S1.

Algorithm S1 Estimation procedure for bidirectional causal effects

1: Apply maximum likelihood estimation (MLE) to the binary response models specified in Model (2.4), and obtain the estimated model parameters $\hat{\xi}_{xz}$, $\hat{\xi}_{yz}$, $\hat{\xi}_{xw}$, and $\hat{\xi}_{yw}$.

2: Substitute these parameter estimates into the identification formula (2.5) in Proposition 1, and compute the estimators of the bidirectional causal effects:

$$\hat{\beta}_{x\to y} = \frac{\hat{\xi}_{yz}}{\hat{\xi}_{xz}}\sqrt{\frac{\hat{\xi}^2_{xw}\hat{\xi}^2_{xz} - \hat{\xi}^2_{yw}\hat{\xi}^2_{xz}}{\hat{\xi}^2_{yz}\hat{\xi}^2_{yw} - \hat{\xi}^2_{yw}\hat{\xi}^2_{xz}}}, \qquad \hat{\beta}_{y\to x} = \frac{\hat{\xi}_{xw}}{\hat{\xi}_{yw}}\sqrt{\frac{\hat{\xi}^2_{yz}\hat{\xi}^2_{yw} - \hat{\xi}^2_{xz}\hat{\xi}^2_{yw}}{\hat{\xi}^2_{xw}\hat{\xi}^2_{xz} - \hat{\xi}^2_{xz}\hat{\xi}^2_{yw}}}.$$

In the following, we provide the asymptotic properties of the estimators $\hat{\beta}_{x\to y}$ and $\hat{\beta}_{y\to x}$. Let $\mathbf{\Lambda} = (\xi_{x0}, \xi_{xz}, \xi_{xw}, \xi_{y0}, \xi_{yz}, \xi_{yw})^{\mathrm{T}}$ denote the vector of true parameters in Model (2.4), and let $\hat{\mathbf{\Lambda}} = (\hat{\xi}_{x0}, \hat{\xi}_{xz}, \hat{\xi}_{xw}, \hat{\xi}_{y0}, \hat{\xi}_{yz}, \hat{\xi}_{yw})^{\mathrm{T}}$ denote the corresponding estimators. Based on Algorithm S1, we apply the Delta method to obtain the asymptotic distributions of $\hat{\beta}_{x\to y}$ and $\hat{\beta}_{y\to x}$. The main results are summarized in Proposition S1.

Proposition S1. *Under Model (2.2), suppose Assumption 1 and the regularity conditions established in Theorem 5.41 of Van der Vaart (2000) hold. Then, as $n \to \infty$, we have:*

$$\sqrt{n}(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}) \xrightarrow{\text{d}} N(0, \Sigma_{\mathbf{\Lambda}}),$$

$$\sqrt{n}(\hat{\beta}_{y\to x} - \beta_{y\to x}) \xrightarrow{\text{d}} N(0, \Sigma_X);$$

$$\sqrt{n}(\hat{\beta}_{x\to y} - \beta_{x\to y}) \xrightarrow{\text{d}} N(0, \Sigma_Y),$$

where the explicit expressions of the variance terms Σ_Λ, Σ_X, and Σ_Y are provided in the Section A.3.

Although Proposition S1 establishes the asymptotic theory, the explicit variance expressions may be overly complicated in practice. As an alternative, we recommend using bootstrap methods to obtain the variance estimator in empirical applications.

A.3 Proof of Proposition S1

Proof. We assume N independent and identically distributed samples $\{Z_i, W_i, X_i, Y_i\}_{i=1}^N$ drawn from an infinite population and estimate using the proposed method. Consider Model (2) under Assumption 1 and the regularity conditions established in Theorem 5.41 of Van der Vaart (2000).

First, we define the notation $R = (1, Z, W)^{\mathrm{T}}$ and $\mathbf{\Lambda} = (\xi_{x0}, \xi_{xz}, \xi_{xw}, \xi_{y0}, \xi_{yz}, \xi_{yw})^{\mathrm{T}}$, where $\Lambda_2 = \xi_{xz}$, $\Lambda_3 = \xi_{xw}$, $\Lambda_5 = \xi_{yz}$, and $\Lambda_6 = \xi_{yw}$. The corresponding estimator is given by $\hat{\mathbf{\Lambda}} = (\hat{\xi}_{x0}, \hat{\xi}_{xz}, \hat{\xi}_{xw}, \hat{\xi}_{y0}, \hat{\xi}_{yz}, \hat{\xi}_{yw})^{\mathrm{T}}$.

The conditional probabilities are given by:

$$p_X(R; \mathbf{\Lambda}) = \mathrm{pr}\left(X = 1 \mid Z, W\right) = \Phi\left(\xi_{x0} + \xi_{xz} Z + \xi_{xw} W\right),$$

$$p_Y(R; \mathbf{\Lambda}) = \mathrm{pr}\left(Y = 1 \mid Z, W\right) = \Phi\left(\xi_{y0} + \xi_{yz} Z + \xi_{yw} W\right).$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

The estimator $\hat{\mathbf{\Lambda}}$ maximize the following log-likelihood functions:

$$\ell_X(X, R; \mathbf{\Lambda}) = \sum_{i=1}^{N} \left[X_i \mathrm{log}\{p_X(R; \mathbf{\Lambda})\} + (1 - X_i)\mathrm{log}\{1 - p_X(R; \mathbf{\Lambda})\}\right],$$
$$\ell_Y(Y, R; \mathbf{\Lambda}) = \sum_{i=1}^{N} \left[Y_i \mathrm{log}\{p_Y(R; \mathbf{\Lambda})\} + (1 - Y_i)\mathrm{log}\{1 - p_Y(R; \mathbf{\Lambda})\}\right].$$

The combined score function is:

$$\mathrm{G}(R; \Lambda) = \begin{pmatrix} \dfrac{\partial \ell_X(X, R; \mathbf{\Lambda})}{\partial \mathbf{\Lambda}} \\ \\ \dfrac{\partial \ell_Y(Y, R; \mathbf{\Lambda})}{\partial \mathbf{\Lambda}} \end{pmatrix}.$$

By the M-estimator theory and under the stated regularity conditions, we have:

$$\sqrt{n}(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}) \xrightarrow{\mathrm{d}} N(0, \Sigma_{\mathbf{\Lambda}}),$$

where the asymptotic covariance matrix is:

$$\Sigma_{\mathbf{\Lambda}} = \{I(\mathbf{\Lambda})\}^{-1}\mathrm{E}\begin{pmatrix} \dfrac{\partial \ell_X(X,R;\mathbf{\Lambda})}{\partial \mathbf{\Lambda}} \\ \\ \dfrac{\partial \ell_Y(Y,R;\mathbf{\Lambda})}{\partial \mathbf{\Lambda}} \end{pmatrix} \left\{ \mathrm{E}\begin{pmatrix} \dfrac{\partial \ell_X(X,R;\mathbf{\Lambda})}{\partial \mathbf{\Lambda}} \\ \\ \dfrac{\partial \ell_Y(Y,R;\mathbf{\Lambda})}{\partial \mathbf{\Lambda}} \end{pmatrix} \right\}^{\mathrm{T}} \{I(\mathbf{\Lambda})^{\mathrm{T}}\}^{-1},$$

and the Fisher information matrix is:

$$I(\mathbf{\Lambda}) = \mathrm{E}\begin{pmatrix} \dfrac{\partial^2 \ell_X}{\partial \mathbf{\Lambda} \partial \mathbf{\Lambda}^{\mathrm{T}}} \\ \\ \dfrac{\partial^2 \ell_Y}{\partial \mathbf{\Lambda} \partial \mathbf{\Lambda}^{\mathrm{T}}} \end{pmatrix}.$$

Define the transformation functions that map parameters to causal effects:

$$\mathrm{g}_X(\mathbf{\Lambda}) = \frac{\Lambda_3(\Lambda_5 - \Lambda_2)}{\Lambda_2(\Lambda_3 - \Lambda_6)}, \quad \mathrm{g}_Y(\mathbf{\Lambda}) = \frac{\Lambda_5(\Lambda_3 - \Lambda_6)}{\Lambda_6(\Lambda_5 - \Lambda_2)},$$

The causal effect estimators can be expressed as:

$$\hat{\beta}_{y\to x} - \beta_{y\to x} = \mathrm{g}_X(\hat{\mathbf{\Lambda}}) - \mathrm{g}_X(\mathbf{\Lambda}), \quad \hat{\beta}_{x\to y} - \beta_{x\to y} = \mathrm{g}_Y(\hat{\mathbf{\Lambda}}) - \mathrm{g}_Y(\mathbf{\Lambda}).$$

By the Delta method, the asymptotic distributions of the causal effect estimators are:

$$\sqrt{n}(\hat{\beta}_{y\to x} - \beta_{y\to x}) \xrightarrow{\mathrm{d}} N(0, \Sigma_X),$$

$$\sqrt{n}(\hat{\beta}_{x\to y} - \beta_{x\to y}) \xrightarrow{\mathrm{d}} N(0, \Sigma_Y),$$

where

$$\Sigma_X = \Omega_X^{\mathrm{T}} \Sigma_\beta \Omega_X, \quad \Sigma_Y = \Omega_Y^{\mathrm{T}} \Sigma_\beta \Omega_Y.$$

Here, Σ_β is the 4×4 submatrix of Σ_Λ corresponding to parameters $(\Lambda_2, \Lambda_3, \Lambda_5, \Lambda_6)$:

$$\Sigma_\beta = \begin{pmatrix} \Sigma_{\mathbf{\Lambda}(2,2)} & \Sigma_{\mathbf{\Lambda}(2,3)} & \Sigma_{\mathbf{\Lambda}(2,5)} & \Sigma_{\mathbf{\Lambda}(2,6)} \\ \Sigma_{\mathbf{\Lambda}(3,2)} & \Sigma_{\mathbf{\Lambda}(3,3)} & \Sigma_{\mathbf{\Lambda}(3,5)} & \Sigma_{\mathbf{\Lambda}(3,6)} \\ \Sigma_{\mathbf{\Lambda}(5,2)} & \Sigma_{\mathbf{\Lambda}(5,3)} & \Sigma_{\mathbf{\Lambda}(5,5)} & \Sigma_{\mathbf{\Lambda}(5,6)} \\ \Sigma_{\mathbf{\Lambda}(6,2)} & \Sigma_{\mathbf{\Lambda}(6,3)} & \Sigma_{\mathbf{\Lambda}(6,5)} & \Sigma_{\mathbf{\Lambda}(6,6)} \end{pmatrix},$$

where $\Sigma_{\mathbf{\Lambda}(i,j)}$ denotes the element in the i-th row and j-th column of $\Sigma_{\mathbf{\Lambda}}$ and the gradient vectors are computed as:

$$\begin{aligned} \Omega_X &= \left(\frac{\partial \mathrm{g}_X(\mathbf{\Lambda})}{\partial \Lambda_2}, \frac{\partial \mathrm{g}_X(\mathbf{\Lambda})}{\partial \Lambda_3}, \frac{\partial \mathrm{g}_X(\mathbf{\Lambda})}{\partial \Lambda_5}, \frac{\partial \mathrm{g}_X(\mathbf{\Lambda})}{\partial \Lambda_6}\right)^{\mathrm{T}} \\ &= \left(\frac{\Lambda_3\Lambda_6}{\Lambda_2^2(\Lambda_6 - \Lambda_3)}, \frac{\Lambda_6(\Lambda_2 - \Lambda_5)}{\Lambda_2(\Lambda_3 - \Lambda_6)^2}, \frac{\Lambda_3}{\Lambda_2(\Lambda_3 - \Lambda_6)}, \frac{\Lambda_3(\Lambda_5 - \Lambda_2)}{\Lambda_2(\Lambda_3 - \Lambda_6)^2}\right)^{\mathrm{T}}, \\ \Omega_Y &= \left(\frac{\partial \mathrm{g}_Y(\mathbf{\Lambda})}{\partial \Lambda_2}, \frac{\partial \mathrm{g}_Y(\mathbf{\Lambda})}{\partial \Lambda_3}, \frac{\partial \mathrm{g}_Y(\mathbf{\Lambda})}{\partial \Lambda_5}, \frac{\partial \mathrm{g}_Y(\mathbf{\Lambda})}{\partial \Lambda_6}\right)^{\mathrm{T}} \\ &= \left(\frac{\Lambda_5(\Lambda_3 - \Lambda_6)}{\Lambda_6(\Lambda_5 - \Lambda_2)^2}, \frac{\Lambda_5}{\Lambda_6(\Lambda_5 - \Lambda_2)}, \frac{\Lambda_2(\Lambda_6 - \Lambda_3)}{\Lambda_6(\Lambda_5 - \Lambda_2)^2}, \frac{\Lambda_5\Lambda_3}{\Lambda_6^2(\Lambda_2 - \Lambda_5)}\right)^{\mathrm{T}}. \end{aligned}$$

□

B. Discussion on Assumption 2 and Proposition 2

In Assumption 2, γ_1 and γ_2 are introduced as sensitivity parameters and are treated as known parameters in the identification process of causal effects under the sensitivity analysis (as described in Proposition 2). Specifically, in this paper we set γ_1 and γ_2 to shift their values to examine the robustness of the proposed method. In practice, prior information can be used to obtain the values of γ_1 and γ_2 for more accurate estimation of the bidirectional causal effects. Therefore, in the subsequent simulation experiments and the analysis of the heart disease dataset, we investigate the fluctuation of the causal effect estimates by specifying a range of values for γ_1 and γ_2. In this paper, γ_1 and γ_2 are treated as sensitivity parameters. From this perspective, the fact that they are taken as known is fundamentally different from the treatment of σ^2 as an unknown parameter.

Below we present the proof of Proposition 2.

Proof. Given Assumption 2, the unobserved confounders follow a bivariate normal distribution:

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim N\begin{pmatrix} \gamma_1\sigma^2 & \gamma_2\sigma^2 \\ \gamma_2\sigma^2 & \sigma^2 \end{pmatrix},$$

where $\gamma_1 \geq \gamma_2^2$ ensures that the covariance matrix is positive semidefinite.

The structural equations are given by:

$$X^\circ = \mu_{x0} + \beta_{y\to x}Y^\circ + \mu_{xz}Z + U, \quad X = \mathbb{I}(X^\circ > 0),$$

$$Y^\circ = \mu_{y0} + \beta_{x\to y}X^\circ + \mu_{yw}W + V, \quad Y = \mathbb{I}(Y^\circ > 0),$$

and solving the structural system yields the reduced form:

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

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

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

The error terms in the reduced form have the following variances:

$$\lambda_1 = \mathrm{Var}(\theta_{xu}U + \theta_{xv}V) = (\theta_{xu}^2\gamma_1 + \theta_{xv}^2 + 2\gamma_2\theta_{xu}\theta_{xv})\sigma^2,$$

$$\lambda_2 = \mathrm{Var}(\theta_{yu}U + \theta_{yv}V) = (\theta_{yu}^2\gamma_1 + \theta_{yv}^2 + 2\gamma_2\theta_{yu}\theta_{yv})\sigma^2.$$

And the observable probabilities follow probit models:

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

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

where $\xi_{x0} = \theta_{x0}/\sqrt{\lambda_1}$, $\xi_{xz} = \theta_{xz}/\sqrt{\lambda_1}$, $\xi_{xw} = \theta_{xw}/\sqrt{\lambda_1}$, $\xi_{y0} = \theta_{y0}/\sqrt{\lambda_2}$, $\xi_{yz} = \theta_{yz}/\sqrt{\lambda_2}$, $\xi_{yw} = \theta_{yw}/\sqrt{\lambda_2}$.

Similarly for the Proof in Section A.1, we define the observable ratios: $k_1 \equiv \xi_{yz}/\xi_{xz}$, and $k_2 \equiv \xi_{xw}/\xi_{yw}$. Computing these ratios using the reduced form coefficients:

$$k_1 \equiv \frac{\xi_{yz}}{\xi_{xz}} = \beta_{x\to y}\frac{\sqrt{(\theta_{xu}^2\gamma_1 + \theta_{xv}^2 + 2\gamma_2\theta_{xu}\theta_{xv})\sigma^2}}{\sqrt{(\theta_{yu}^2\gamma_1 + \theta_{yv}^2 + 2\gamma_2\theta_{yu}\theta_{yv})\sigma^2}} = \beta_{x\to y}\frac{\sqrt{\gamma_1 + \beta_{y\to x}^2 + 2\gamma_2\beta_{y\to x}}}{\sqrt{\gamma_1\beta_{x\to y}^2 + 1 + 2\gamma_2\beta_{x\to y}}},$$

$$k_2 \equiv \frac{\xi_{xw}}{\xi_{yw}} = \beta_{y\to x}\frac{\sqrt{(\theta_{yu}^2\gamma_1 + \theta_{yv}^2 + 2\gamma_2\theta_{yu}\theta_{yv})\sigma^2}}{\sqrt{(\theta_{xu}^2\gamma_1 + \theta_{xv}^2 + 2\gamma_2\theta_{xu}\theta_{xv})\sigma^2}} = \beta_{y\to x}\frac{\sqrt{\gamma_1\beta_{x\to y}^2 + 1 + 2\gamma_2\beta_{x\to y}}}{\sqrt{\gamma_1 + \beta_{y\to x}^2 + 2\gamma_2\beta_{y\to x}}}.$$

Then squaring both ratios yields:

$$\begin{cases} k_1^2 = \beta_{x\to y}^2 \dfrac{\gamma_1 + 2\gamma_2\beta_{y\to x} + \beta_{y\to x}^2}{\gamma_1\beta_{x\to y}^2 + 2\gamma_2\beta_{x\to y} + 1}, \\ \\ k_2^2 = \beta_{y\to x}^2 \dfrac{\gamma_1\beta_{x\to y}^2 + 2\gamma_2\beta_{x\to y} + 1}{\gamma_1 + 2\gamma_2\beta_{y\to x} + \beta_{y\to x}^2}. \end{cases}$$

Thus, we have the constraint $k_1k_2 = \beta_{x\to y}\beta_{y\to x}$. Using this constraint, we can obtain a quadratic equation about $\beta_{x\to y}$:

$$\begin{aligned} &k_1^2\left(\gamma_1\beta_{x\to y}^2 + 2\gamma_2\beta_{x\to y} + 1\right) = \beta_{x\to y}^2\gamma_1 + 2\gamma_2\beta_{x\to y}^2\beta_{y\to x} + k_1^2k_2^2 \\ \Rightarrow &k_1^2\gamma_1\beta_{x\to y}^2 + 2\gamma_2k_1^2\beta_{x\to y} + k_1^2 = \beta_{x\to y}^2\gamma_1 + 2\gamma_2k_1k_2\beta_{x\to y} + k_1^2k_2^2 \\ \Rightarrow &\left(k_1^2\gamma_1 - \gamma_1\right)\beta_{x\to y}^2 + \left(2\gamma_2k_1^2 - 2k_1k_2\gamma_2\right)\beta_{x\to y} + k_1^2 - k_1^2k_2^2 = 0. \end{aligned}$$

Therefore, according to the root formula, we can obtain:

$$\beta_{x\to y} = \frac{-r_2 \pm \sqrt{r_2^2 - 4r_1r_3}}{2r_1},$$

where $r_1 = k_1^2\gamma_1 - \gamma_1$, $r_2 = 2\gamma_2k_1^2 - 2k_1k_2\gamma_2$, $r_3 = k_1^2 - k_1^2k_2^2$. So we finally obtain:

$$\beta_{x\to y} = \frac{\gamma_2k_1(k_2 - k_1) \pm |k_1|\sqrt{\Delta_1}}{\gamma_1(k_1^2 - 1)}, \quad \beta_{y\to x} = \frac{k_1k_2}{\beta_{x\to y}} = \frac{\gamma_1k_1k_2(k_1^2 - 1)}{\gamma_2k_1(k_2 - k_1) \pm |k_1|\sqrt{\Delta_1}},$$

equivalently,

$$\begin{cases} \beta_{x\to y} = \dfrac{\gamma_2k_1(k_2 - k_1) + |k_1|\sqrt{\Delta_1}}{\gamma_1(k_1^2 - 1)}, \\ \\ \beta_{y\to x} = \dfrac{\gamma_1k_1k_2(k_1^2 - 1)}{\gamma_2k_1(k_2 - k_1) + |k_1|\sqrt{\Delta_1}}. \end{cases} \quad \text{or} \quad \begin{cases} \beta_{x\to y} = \dfrac{\gamma_2k_1(k_2 - k_1) - |k_1|\sqrt{\Delta_1}}{\gamma_1(k_1^2 - 1)}, \\ \\ \beta_{y\to x} = \dfrac{\gamma_1k_1k_2(k_1^2 - 1)}{\gamma_2k_1(k_2 - k_1) - |k_1|\sqrt{\Delta_1}}. \end{cases}$$

where $\Delta_1 = \gamma_1 k_1^2 k_2^2 - \gamma_1 k_1^2 - \gamma_1 k_2^2 + \gamma_1 + \gamma_2^2 k_1^2 - 2\gamma_2^2 k_1 k_2 + \gamma_2^2 k_2^2$.

The two components obtained using the same sign in the paired expressions above must be taken together. For each $j \in \{1, 2\}$, let $(\beta_{x\to y,j}, \beta_{y\to x,j})$ denote the corresponding candidate solution pair. The jth pair is admissible only if $\mathrm{sgn}(\beta_{x\to y,j}) = \mathrm{sgn}(k_1)$ and $\mathrm{sgn}(\beta_{y\to x,j}) = \mathrm{sgn}(k_2)$. The sign restrictions are therefore checked separately for the two candidate pairs. If exactly one pair satisfies both restrictions, the bidirectional causal effects are uniquely identified; otherwise, all admissible pairs constitute the identification results. $\square$

We also provide here the identification procedure from the perspective of $\beta_{y\to x}$. Symmetrically, using the definition of k_2, we obtain a quadratic equation in $\beta_{y\to x}$:

$$(k_2^2 - 1)\beta_{y\to x}^2 + (2\gamma_2 k_2^2 - 2\gamma_2 k_1 k_2)\beta_{y\to x} + \gamma_1 k_2^2 - \gamma_1 k_1^2 k_2^2 = 0.$$

Using the quadratic formula, we obtain

$$\beta_{y\to x} = \frac{\gamma_2 k_2(k_1 - k_2) \pm |k_2|\sqrt{\Delta_2}}{k_2^2 - 1}, \quad \beta_{x\to y} = \frac{k_1 k_2}{\beta_{y\to x}} = \frac{k_1 k_2(k_2^2 - 1)}{\gamma_2 k_2(k_1 - k_2) \pm |k_2|\sqrt{\Delta_2}},$$

equivalently,

$$\begin{cases} \beta_{y\to x} = \dfrac{\gamma_2 k_2(k_1 - k_2) + |k_2|\sqrt{\Delta_2}}{k_2^2 - 1}, \\[2ex] \beta_{x\to y} = \dfrac{k_1 k_2(k_2^2 - 1)}{\gamma_2 k_2(k_1 - k_2) + |k_2|\sqrt{\Delta_2}}. \end{cases} \quad \text{or} \quad \begin{cases} \beta_{y\to x} = \dfrac{\gamma_2 k_2(k_1 - k_2) - |k_2|\sqrt{\Delta_2}}{k_2^2 - 1}, \\[2ex] \beta_{x\to y} = \dfrac{k_1 k_2(k_2^2 - 1)}{\gamma_2 k_2(k_1 - k_2) - |k_2|\sqrt{\Delta_2}}. \end{cases}$$

where $\Delta_2 = \gamma_2^2 k_1^2 - 2k_1k_2\gamma_2^2 + \gamma_2^2 k_2^2 - \gamma_1 k_2^2 + \gamma_1 k_1^2 k_2^2 + \gamma_1 - \gamma_1 k_1^2$.

For each of the two paired expressions above, the resulting candidate pair is admissible only if the candidate values satisfy $\mathrm{sgn}(\beta_{y\to x,j}) = \mathrm{sgn}(k_2)$ and $\mathrm{sgn}(\beta_{x\to y,j}) = \mathrm{sgn}(k_1)$.

C. Proof of theoretic results in Section 3.2

Given Assumption 2, we conduct a sensitivity analysis based on the following model, which incorporates the sensitivity parameters η and δ:

$$\begin{aligned} X^\circ &= \mu_{x0} + \beta_{y\to x} Y^\circ + \mu_{xz} Z + \delta W + U, \quad X = \mathbb{I}(X^\circ > 0), \\ Y^\circ &= \mu_{y0} + \beta_{x\to y} X^\circ + \eta Z + \mu_{yw} W + V, \quad Y = \mathbb{I}(Y^\circ > 0). \end{aligned} \tag{S2}$$

C.1 Proposition S2 and proof of Proposition S2

Proposition S2. *Under Model (3.7) and Assumption 2, for any given sensitivity parameters* $(\gamma_1, \gamma_2, \eta_0, \delta_0)$*, the bidirectional causal effect* $\beta_{y\to x}$ *can be partially identified by solving the following univariate quartic equation:* $\mathrm{F}(\beta_{y\to x}; \xi_{xz}, \xi_{xw}, \xi_{yz}, \xi_{yw}, \eta_0, \delta_0, \gamma_1, \gamma_2) = 0$*, where the detail expression of* $\mathrm{F}(\cdot)$ *is provided in the following proof. Once* $\beta_{y\to x}$ *is obtained, the corresponding causal effect* $\beta_{x\to y}$ *can be identified through the following explicit formula:*

$$\beta_{x\to y} = \frac{\xi_{xz}\xi_{yw}\delta_0\eta_0 - \xi_{yz}\xi_{xw} - \xi_{xz}\xi_{yw}\eta_0\beta_{y\to x} + \xi_{yz}\xi_{xw}\eta_0\beta_{y\to x}}{\xi_{xz}\xi_{yw}\beta_{y\to x} - \xi_{yz}\xi_{xw}\eta_0\delta_0\beta_{y\to x} + \xi_{xz}\xi_{yw}\delta_0 - \xi_{yz}\xi_{xw}\delta_0}.$$

Building upon Proposition 2, Proposition S2 incorporates the direct effect parameters η and δ to conduct a more comprehensive sensitivity analysis under the Model (3.7). By utilizing the univariate quartic equation provided in Proposition S2, we can obtain corresponding estimates of the causal effects under any combination of sensitivity parameters $(\gamma_1, \gamma_2, \eta_0, \delta_0)$. This enables us to evaluate the impact of different confounding scenarios induced by unobserved confounders U and V, as well as violations of instrumental variable assumptions related to Z and W, on the causal effect estimates.

Under Assumption 2 and Model (S2), we consider the general case by simultaneously examining the influence of variations in the sensitivity parameters γ_1, γ_2, η, and δ on bidirectional causal effects. Below, we present the detailed process of the sensitivity analysis and the derivation of the causal effect identification expressions. From the structural model, we can obtain its reduced form:

$$\begin{aligned} X^\circ &= \theta_{x0} + \theta^*_{xz} Z + \theta^*_{xw} W + \theta_{xu} U + \theta_{xv} V, \quad X = \mathbb{I}(X^\circ > 0), \\ Y^\circ &= \theta_{y0} + \theta^*_{yz} Z + \theta^*_{yw} W + \theta_{yu} U + \theta_{yv} V, \quad Y = \mathbb{I}(Y^\circ > 0), \end{aligned} \tag{S3}$$

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

Proof. Under Assumption 2, we have $V \sim N(0, \sigma^2)$, $U \sim N(0, \gamma_1\sigma^2)$, and

$\mathrm{Cov}(U,V) = \gamma_2\sigma^2$, which implies that the error terms in the reduced form follow normal distributions:

$$\theta_{xu}U + \theta_{xv}V \sim N(0, (\theta_{xu}^2\gamma_1 + \theta_{xv}^2 + 2\gamma_2\theta_{xu}\theta_{xv})\sigma^2),$$

$$\theta_{yu}U + \theta_{yv}V \sim N(0, (\theta_{yu}^2\gamma_1 + \theta_{yv}^2 + 2\gamma_2\theta_{yu}\theta_{yv})\sigma^2).$$

Therefore, according to Model (S3), we can derive the Probit model as follows:

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

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

where $\xi_{x0} = \theta_{x0}/\sqrt{\lambda_1}$, $\xi_{xz} = \theta_{xz}^*/\sqrt{\lambda_1}$, $\xi_{xw} = \theta_{xw}^*/\sqrt{\lambda_1}$; $\xi_{y0} = \theta_{y0}/\sqrt{\lambda_2}$, $\xi_{yz} = \theta_{yz}^*/\sqrt{\lambda_2}$, $\xi_{yw} = \theta_{yw}^*/\sqrt{\lambda_2}$, $\lambda_1 = \mathrm{Var}(\theta_{xu}U + \theta_{xv}V)$, $\lambda_2 = \mathrm{Var}(\theta_{yu}U + \theta_{yv}V)$. The specific expressions for the four key parameters (ξ_{xz}, ξ_{xw}, ξ_{yz}, ξ_{yw}) after introducing sensitivity parameters are as follows:

$$\xi_{xz} = \frac{c(\mu_{xz} + \eta\beta_{y\to x})}{|c|\sigma\sqrt{\gamma_1 + \beta_{y\to x}^2 + 2\gamma_2\beta_{y\to x}}}, \quad \xi_{xw} = \frac{c(\mu_{yw}\beta_{y\to x} + \delta)}{|c|\sigma\sqrt{\gamma_1 + \beta_{y\to x}^2 + 2\gamma_2\beta_{y\to x}}},$$
$$\xi_{yz} = \frac{c(\mu_{xz}\beta_{x\to y} + \eta)}{|c|\sigma\sqrt{\gamma_1 + \beta_{x\to y}^2 + 2\gamma_2\beta_{x\to y}}}, \quad \xi_{yw} = \frac{c(\delta\beta_{x\to y} + \mu_{yw})}{|c|\sigma\sqrt{\gamma_1 + \beta_{x\to y}^2 + 2\gamma_2\beta_{x\to y}}}.$$

Thus we obtain two key ratio relationships:

$$\frac{\xi_{xz}}{\xi_{yz}} = \frac{\mu_{xz} + \eta\beta_{y\to x}}{\sqrt{\gamma_1 + \beta_{y\to x}^2 + 2\gamma_2\beta_{y\to x}}} \times \frac{\sqrt{\gamma_1 + \beta_{x\to y}^2 + 2\gamma_2\beta_{x\to y}}}{\mu_{xz}\beta_{x\to y} + \eta},$$
$$\frac{\xi_{xw}}{\xi_{yw}} = \frac{\mu_{yw}\beta_{y\to x} + \delta}{\sqrt{\gamma_1 + \beta_{y\to x}^2 + 2\gamma_2\beta_{y\to x}}} \times \frac{\sqrt{\gamma_1 + \beta_{x\to y}^2 + 2\gamma_2\beta_{x\to y}}}{\mu_{yw} + \delta\beta_{x\to y}}.$$

Given $\eta_0 = \eta/\mu_{xz}$ and $\delta_0 = \delta/\mu_{yw}$, we can obtain the equation:

$$\frac{\xi_{xz}}{\xi_{yz}} \Big/ \frac{\xi_{xw}}{\xi_{yw}} = \frac{1 + \eta_0 \beta_{y \to x}}{\beta_{y \to x} + \delta_0} \times \frac{1 + \delta_0 \beta_{x \to y}}{\beta_{x \to y} + \eta_0}.$$

Through algebraic manipulation, this leads to the expression of $\beta_{x \to y}$:

$$\beta_{x \to y} = \frac{\xi_{xz}\xi_{yw}\eta_0\beta_{y \to x} - \xi_{yz}\xi_{xw}\eta_0\beta_{y \to x} - \xi_{xz}\xi_{yw}\delta_0\eta_0 + \xi_{yz}\xi_{xw}}{\xi_{xz}\xi_{yw}\beta_{y \to x} - \xi_{yz}\xi_{xw}\eta_0\delta_0\beta_{y \to x} + \xi_{xz}\xi_{yw}\delta_0 - \xi_{yz}\xi_{xw}\delta_0}.$$

Additionally, we can establish another constraint by considering the squared ratio:

$$\left(\frac{\xi_{xz}}{\xi_{yz}}\right)^2 = \frac{(1 + \eta_0 \beta_{y \to x})^2}{\gamma_1 + \beta_{y \to x}^2 + 2\gamma_2 \beta_{y \to x}} \times \frac{\gamma_1 + \beta_{x \to y}^2 + 2\gamma_2 \beta_{x \to y}}{(\beta_{x \to y} + \eta_0)^2}.$$

Define $q_0 = (\xi_{xz}/\xi_{yz})^2$, we derive a system of polynomial equations in the causal effect parameters and the complete system yields:

$$\begin{aligned}
&(q_0\gamma_1 - 1)\beta_{x \to y}^2 + (2\eta_0\gamma_1 q_0 2\gamma_2)\beta_{x \to y} + (q_0\gamma_1\eta_0^2 - \gamma_1) \\
&(q_0 - 1)\beta_{x \to y}^2\beta_{y \to x}^2 + (2\eta_0 q_0 - 2\gamma_2)\beta_{y \to x}^2\beta_{x \to y} + (q_0\eta_0^2 - \gamma_1)\beta_{y \to x}^2 \\
&(2\gamma_2 q_0 - 2\eta_0)\beta_{x \to y}^2 + (4\eta_0\gamma_2 q_0 - 4\eta_0\gamma_2)\beta_{x \to y}\beta_{y \to x} + (2q_0\gamma_2\eta_0^2 - 2\gamma_1\eta_0)\beta_{y \to x} = 0.
\end{aligned}$$

By simultaneously solving these constraint equations, we obtain a quartic equation in one unknown with respect to the causal effect parameters:

$$\begin{aligned}
&a_4\beta_{y \to x}^4 + a_3\beta_{y \to x}^3 + a_2\beta_{y \to x}^2 + a_1\beta_{y \to x} + a_0 = 0 \\
\Leftrightarrow\ &\mathrm{F}(\beta_{y \to x}; \xi_{xz}, \xi_{xw}, \xi_{yz}, \xi_{yw}, \gamma_1, \gamma_2, \eta_0, \delta_0) = 0,
\end{aligned}$$

where the coefficients a_0, a_1, a_2, a_3, a_4 are complex functions of the observable parameters and sensitivity parameters.

This quartic equation provides the foundation for identification of bidirectional causal effects. The approach allows us to examine how variations in the sensitivity parameters γ_1, γ_2, η_0, and δ_0 within the identification expressions influence the estimated causal effects, thereby enabling evaluation of the robustness of our proposed method when underlying assumptions are violated. While the complete analytical form of the quartic equation is algebraically complex, it can be solved numerically for any given sensitivity parameters, making this approach practical for empirical sensitivity analysis.

$\square$

In what follows, we present formal proofs for the three sensitivity analysis settings outlined in Corollaries 1–3.

C.2 Proof of Corollary 1

In Corollary 1, we set $\delta_0 = 0$, $\gamma_1 = 1$, $\gamma_2 = 0$, and let η_0 be an arbitrary variable sensitivity parameter. Consequently, given Assumption 2 and Model (S2), we obtain the reduced form:

$$
\begin{aligned}
X^\circ &= \theta_{x0} + \theta_{xz,\eta} Z + \theta_{xw,\eta} W + \theta_{xu} U + \theta_{xv} V, \quad X = \mathbb{I}(X^\circ > 0), \\
Y^\circ &= \theta_{y0} + \theta_{yz,\eta} Z + \theta_{yw,\eta} W + \theta_{yu} U + \theta_{yv} V, \quad Y = \mathbb{I}(Y^\circ > 0),
\end{aligned}
\tag{S4}
$$

where $\theta_{xz,\eta} = c(\mu_{xz} + \eta\beta_{y\to x})$, $\theta_{xw,\eta} = c\beta_{y\to x}\mu_{yw}$, $\theta_{yz,\eta} = c(\beta_{x\to y}\mu_{xz} + \eta)$, $\theta_{yw,\eta} = c\mu_{yw}$, the remaining parameters θ_{x0}, θ_{xu}, θ_{xv}, θ_{y0}, θ_{yu}, θ_{yv}, and c have been defined in Section C.1.

Proof. According to Model (S4), we can derive the Probit model as follows:

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

where $\xi_{xz} = \theta_{xz,\eta}/\sqrt{\lambda_1}$, $\xi_{xw} = \theta_{xw,\eta}/\sqrt{\lambda_1}$, $\xi_{yz} = \theta_{yz,\eta}/\sqrt{\lambda_2}$, $\xi_{yw} = \theta_{yw,\eta}/\sqrt{\lambda_2}$. The coefficient ξ_{x0}, ξ_{y0}, λ_1, and λ_2 have been defined in Section C.1. However, note that we have defined $\gamma_1 = 1$ and $\gamma_2 = 0$ here, so the error variances simplify to: $\lambda_1 = (\theta_{xu}^2 + \theta_{xv}^2)\sigma^2$ and $\lambda_2 = (\theta_{yu}^2 + \theta_{yv}^2)\sigma^2$.

For the parameters ξ_{xz}, ξ_{xw}, ξ_{yz}, and ξ_{yw}, after parameterization via $\eta_0 = \eta/\mu_{xz}$, the following ratio relationships can be obtained:

$$
\begin{aligned}
t_1 &= \xi_{xw}/\xi_{xz} = \frac{\mu_{yw}\beta_{y\to x}}{\mu_{xz}(1 + \eta_0\beta_{y\to x})}, \qquad t_2 = \xi_{yz}/\xi_{yw} = \frac{\mu_{xz}(\beta_{x\to y} + \eta_0)}{\mu_{yw}}, \\
t_3 &= \xi_{xz}/\xi_{yz} = \frac{(1 + \eta_0\beta_{y\to x})\sqrt{1 + \beta_{x\to y}^2}}{(\beta_{x\to y} + \eta_0)\sqrt{1 + \beta_{y\to x}^2}}, \qquad t_4 = \xi_{xw}/\xi_{yw} = \frac{\beta_{y\to x}\sqrt{1 + \beta_{x\to y}^2}}{\sqrt{1 + \beta_{y\to x}^2}}.
\end{aligned}
$$

From the definitions of t_1, t_2, t_3, and t_4, we can derive two key relationships:

$$\begin{cases} \beta_{x\to y} = \dfrac{t_1 t_2(1+\eta_0\beta_{y\to x})}{\beta_{y\to x}} - \eta_0, \\ \\ t_1t_2t_3t_4 = t_4^2 = \beta_{y\to x}^2 \dfrac{1+\beta_{x\to y}^2}{1+\beta_{y\to x}^2}. \end{cases}$$

Through algebraic manipulation of the constraint system, we obtain a quadratic equation in $\beta_{y\to x}$:

$$s_1\beta_{y\to x}^2 + s_2\beta_{y\to x} + s_3 = 0,$$

where the coefficients are $s_1 = \eta_0^2(t_1t_2-1)^2 - t_4^2 + 1$, $s_2 = 2t_1t_2\eta_0(t_1t_2-1)$, and $s_3 = t_1^2t_2^2 - t_4^2$. Therefore, the quadratic formula yields two sets of identification expressions for bidirectional causal effects:

$$\beta_{y\to x} = \frac{-s_2 \pm \sqrt{s_2^2 - 4s_1s_3}}{2s_1}, \quad \beta_{x\to y} = \eta_0(t_1t_2-1) + \frac{2s_1t_1t_2}{-s_2 \pm \sqrt{s_2^2-4s_1s_3}},$$

where $s_2^2 - 4s_1s_3 \geq 0$. For each $j \in \{1,2\}$, the two expressions obtained using the same sign form the candidate pair $(\beta_{x\to y,j}, \beta_{y\to x,j})$. The jth candidate pair is admissible only if $\mathrm{sgn}(\beta_{y\to x,j}) = \mathrm{sgn}(t_4)$. $\square$

C.3 Proof of Corollary 2

In Corollary 2, we set $\eta_0 = 0$, $\gamma_1 = 1$, $\gamma_2 = 0$, and let δ_0 be an arbitrary variable sensitivity parameter. Consequently, given Assumption 2 and Model (S2), we obtain the reduced form:

$$\begin{aligned} X^\circ &= \theta_{x0} + \theta_{xz,\delta} Z + \theta_{xw,\delta} W + \theta_{xu} U + \theta_{xv} V, \quad X = \mathbb{I}(X^\circ > 0), \\ Y^\circ &= \theta_{y0} + \theta_{yz,\delta} Z + \theta_{yw,\delta} W + \theta_{yu} U + \theta_{yv} V, \quad Y = \mathbb{I}(Y^\circ > 0), \end{aligned} \tag{S5}$$

where $\theta_{xz,\delta} = c\mu_{xz}$, $\theta_{xw,\delta} = c(\delta + \beta_{y\to x}\mu_{yw})$, $\theta_{yz,\delta} = c\beta_{x\to y}\mu_{xz}$, $\theta_{yw,\delta} = c(\delta\beta_{x\to y} + \mu_{yw})$, the remaining parameters θ_{x0}, θ_{xu}, θ_{xv}, θ_{y0}, θ_{yu}, θ_{yv}, and c have been defined in Section C.1.

Proof. According to Model (S5), we can derive the Probit model as follows:

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

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

where $\xi_{x0} = \theta_{x0}/\sqrt{\lambda_1}$, $\xi_{xz} = \theta_{xz,\delta}/\sqrt{\lambda_1}$, $\xi_{xw} = \theta_{xw,\delta}/\sqrt{\lambda_1}$, $\xi_{y0} = \theta_{y0}/\sqrt{\lambda_2}$, $\xi_{yz} = \theta_{yz,\delta}/\sqrt{\lambda_2}$, $\xi_{yw} = \theta_{yw,\delta}/\sqrt{\lambda_2}$, and the definitions of λ_1 and λ_2 here are the same as those in Section C.2.

For the parameters ξ_{xz}, ξ_{yz}, ξ_{xw}, and ξ_{yw}, after parameterization via $\delta_0 = \delta/\mu_{yw}$, the following ratio relationships can be obtained:

$$t_1 = \xi_{xw}/\xi_{xz} = \frac{\mu_{yw}(\beta_{y\to x} + \delta_0)}{\mu_{xz}}, \qquad t_2 = \xi_{yz}/\xi_{yw} = \frac{\mu_{xz}\beta_{x\to y}}{\mu_{yw}(\delta_0\beta_{x\to y} + 1)},$$

$$t_3 = \xi_{xz}/\xi_{yz} = \frac{\sqrt{1+\beta_{x\to y}^2}}{\beta_{x\to y}\sqrt{1+\beta_{y\to x}^2}}, \qquad t_4 = \xi_{xw}/\xi_{yw} = \frac{(\beta_{y\to x} + \delta_0)\sqrt{1+\beta_{x\to y}^2}}{(\delta_0\beta_{x\to y} + 1)\sqrt{1+\beta_{y\to x}^2}}.$$

From the definitions of t_1, t_2, t_3, and t_4, we can derive two key relationships:

$$\begin{cases} \beta_{y\to x} = \dfrac{t_1 t_2(\delta_0\beta_{x\to y} + 1)}{\beta_{x\to y}} - \delta_0, \\ t_3 t_4 / t_1 t_2 = t_3^2 = \dfrac{1+\beta_{x\to y}^2}{\beta_{x\to y}^2(1+\beta_{y\to x}^2)}. \end{cases}$$

Through algebraic manipulation of the constraint system, we obtain a quadratic equation in $\beta_{x\to y}$:

$$s_4\beta_{x\to y}^2 + s_5\beta_{x\to y} + s_6 = 0,$$

where the coefficients are $s_4 = t_3t_4 + t_3t_4\delta_0^2(t_1t_2-1)^2 - t_1t_2,\ s_5 = 2t_4^2\delta_0(t_1t_2-1)$, and $s_6 = t_1t_2(t_4^2-1)$. Therefore, we obtain the identification expressions for bidirectional causal effects:

$$\beta_{x\to y} = \frac{-s_5 \pm \sqrt{s_5^2 - 4s_4s_6}}{2s_4}, \quad \beta_{y\to x} = \delta_0(t_1t_2-1) + \frac{2s_4t_1t_2}{-s_5 \pm \sqrt{s_5^2 - 4s_4s_6}},$$

where $s_5^2 - 4s_4s_6 \geq 0$. For each $j \in \{1,2\}$, the two expressions obtained using the same sign form the candidate pair $(\beta_{x\to y,j}, \beta_{y\to x,j})$. The jth candidate pair is admissible only if $\mathrm{sgn}(\beta_{x\to y,j}) = \mathrm{sgn}(t_3)$.

□

C.4 Proof of Corollary 3

Corollary 3. (Complete version.) *Under Assumption 2 and Model* (S2), *set $\gamma_1 = \gamma_2 = 1$ and use the fixed sensitivity-parameter values $\eta_0 = \eta/\sigma$ and $\delta_0 = \delta/\sigma$. Let $\xi_{xz}, \xi_{xw}, \xi_{yz}$, and ξ_{yw} denote the identifiable slope coefficients in the reduced-form Probit models for X and Y. Assume $\sigma > 0$, $\beta_{x\to y} \neq -1$, $\beta_{y\to x} \neq -1$, $\beta_{x\to y}\beta_{y\to x} \neq 1$, and that every displayed denominator is nonzero, and write $c = (1 - \beta_{x\to y}\beta_{y\to x})^{-1}$. For fixed (η_0, δ_0), the* *complete nondegenerate solution set consists of all candidate pairs below whose evaluated values satisfy the constraint conditions stated for the corresponding case.*

When $(c > 0,\ \beta_{y\to x} > -1,\ \beta_{x\to y} > -1)$ *or* $(c < 0,\ \beta_{y\to x} < -1,\ \beta_{x\to y} < -1)$*:*

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

When $(c > 0,\ \beta_{y\to x} < -1,\ \beta_{x\to y} > -1)$ *or* $(c < 0,\ \beta_{y\to x} > -1,\ \beta_{x\to y} < -1)$*:*

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

When $(c > 0,\ \beta_{y\to x} > -1,\ \beta_{x\to y} < -1)$ *or* $(c < 0,\ \beta_{y\to x} < -1,\ \beta_{x\to y} > -1)$*:*

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

When $c < 0,\ \beta_{y\to x} > -1,\ \beta_{x\to y} > -1$*:*

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

Note that the combination $c > 0$, $\beta_{y\to x} < -1$, $\beta_{x\to y} < -1$ *is impossible because it would imply* $\beta_{x\to y}\beta_{y\to x} > 1$, *contradicting* $c > 0$.

The main-text statement of Corollary 3 corresponds to the first and last cases above, both of which include the situation where both effects exceed -1. Given some prior information about the causal effects $\beta_{x\to y}$

and $\beta_{y\to x}$, one can select the appropriate identification expression among the four algebraic candidate pairs for computation. Cases in which an effect equals -1, the product equals 1, or a displayed denominator is zero lie outside this closed-form statement and require separate analysis of the unreduced coefficient equations.

Proof. For Corollary 3 only, we set $\gamma_1 = \gamma_2 = 1$ and use the signal-to-noise parameterization $\eta_0 = \eta/\sigma$ and $\delta_0 = \delta/\sigma$. Thus, η_0 and δ_0 are treated as fixed sensitivity-parameter values in each identification calculation. Under Assumption 2, $\gamma_1 = \gamma_2 = 1$ makes the covariance matrix of (U, V) positive semidefinite but singular; in particular, $U = V$ almost surely. The main-text statement of Corollary 3 concerns the common domain $\beta_{x\to y} > -1$ and $\beta_{y\to x} > -1$. The remaining sign configurations are recorded below only as supplementary cases. Consequently, given Assumption 2 and Model (S2), we obtain its reduced form:

$$
\begin{aligned}
X^\circ &= \theta_{x0} + \theta^*_{xz} Z + \theta^*_{xw} W + \theta_{xu} U + \theta_{xv} V, \quad X = \mathbb{I}(X^\circ > 0),\\
Y^\circ &= \theta_{y0} + \theta^*_{yz} Z + \theta^*_{yw} W + \theta_{yu} U + \theta_{yv} V, \quad Y = \mathbb{I}(Y^\circ > 0),
\end{aligned}
\tag{S6}
$$

where the parameters θ_{x0}, θ^*_{xz}, θ^*_{xw}, θ_{xu}, θ_{xv}, θ_{y0}, θ^*_{yz}, θ^*_{yw}, θ_{yu}, and θ_{yv} have been provided in Section C.1.

According to Model (S6), we can derive the Probit model as follows:

$$\begin{aligned}\mathrm{pr}(X=1\mid Z,W)&=\mathrm{pr}(X^{\circ}>0\mid Z,W)=\Phi(\xi_{x0}+\xi_{xz}Z+\xi_{xw}W),\\ \mathrm{pr}(Y=1\mid Z,W)&=\mathrm{pr}(Y^{\circ}>0\mid Z,W)=\Phi(\xi_{y0}+\xi_{yz}Z+\xi_{yw}W),\end{aligned}\tag{S7}$$

where ξ_{x0}, ξ_{xz}, ξ_{xw}, ξ_{y0}, ξ_{yz}, and ξ_{yw} have been provided in Section C.1. Since $U=V$ almost surely, the two reduced-form error variances simplify to

$$\lambda_1=c^2(1+\beta_{y\to x})^2\sigma^2,\qquad \lambda_2=c^2(1+\beta_{x\to y})^2\sigma^2.$$

Here $\sigma>0$, and the Probit standardization is well defined only when $\beta_{x\to y}\neq -1$ and $\beta_{y\to x}\neq -1$. Moreover, $c=\{1-\beta_{x\to y}\beta_{y\to x}\}^{-1}$ requires $\beta_{x\to y}\beta_{y\to x}\neq 1$. Throughout this derivation, every denominator displayed in an identification formula is also assumed to be nonzero.

Set

$$a=\beta_{x\to y},\qquad b=\beta_{y\to x},\qquad m=\mu_{xz}/\sigma,\qquad n=\mu_{yw}/\sigma,$$

and let $s_c=\mathrm{sgn}(c)$, $r_y=\mathrm{sgn}(1+b)$, $r_x=\mathrm{sgn}(1+a)$, and $(A,B)=s_c(r_y,r_x)$. Because $A,B\in\{-1,1\}$, the four standardized coefficient relations can be written without dividing by either causal effect:

$$\begin{aligned}m+\eta_0 b&=A\xi_{xz}(1+b), &\qquad am+\eta_0&=B\xi_{yz}(1+a),\\ bn+\delta_0&=A\xi_{xw}(1+b), &\qquad n+\delta_0 a&=B\xi_{yw}(1+a).\end{aligned}$$

These equations also cover zero causal effects. Eliminating m, n and a gives

$$(1+b)\big\{(B\xi_{yw}-A\xi_{xw})(A\xi_{xz}-\eta_0)b-(A\xi_{xw}-\delta_0)(A\xi_{xz}-B\xi_{yz})\big\}=0.$$

It is clear that $b=-1$ is a root of this quadratic equation, but it is excluded since it yields $\lambda_1=0$. Using $A^2=B^2=1$, the remaining root and the corresponding value of a are

$$a=\frac{(\xi_{xw}-AB\xi_{yw})(\xi_{yz}-B\eta_0)}{(AB\xi_{yz}-\xi_{xz})(\xi_{yw}-B\delta_0)},\quad b=\frac{(AB\xi_{yz}-\xi_{xz})(\xi_{xw}-A\delta_0)}{(\xi_{xw}-AB\xi_{yw})(\xi_{xz}-A\eta_0)}. \tag{S8}$$

Table S1 shows that the seven mutually exclusive sign-and-product cases correspond to the following values of (A,B).

Beyond the seven configurations discussed above, the omitted configuration $ab<1$, $a<-1$, and $b<-1$ is impossible because it implies $ab>1$, which contradicts $ab<1$. Substituting the seven pairs of values into Equation (S8) yields exactly the candidate pairs listed in Corollary 3; the reduction to four candidate pairs is due to the fact that the identifying expressions coincide across different cases, and thus we present them in a consolidated form. Figure S1 depicts the algebraic partition induced by the above constraints. For each fixed (η_0,δ_0), the full set of nondegenerate solutions consists of all candidate pairs from Corollary 3 that satisfy the corresponding constraints; further, with prior information, one can use this set to screen the complete nondegenerate solution set.

Table S1: Values of (A, B) under sign and product constraints. The first three columns list the constraints, and the last column gives the corresponding values. $a = \beta_{x\to y}$, $b = \beta_{y\to x}$, $c = (1 - ab)^{-1}$.

a	b	c	(A, B)
> -1	> -1	> 0	$(1, 1)$
> -1	< -1	> 0	$(-1, 1)$
< -1	> -1	> 0	$(1, -1)$
> -1	> -1	< 0	$(-1, -1)$
< -1	< -1	< 0	$(1, 1)$
< -1	> -1	< 0	$(-1, 1)$
> -1	< -1	< 0	$(1, -1)$

$\square$

The statement and derivation of Corollary 3 concern the setting $\gamma_1 = 1$ and $\gamma_2 = 1$. The case $\gamma_2 = -1$ requires a separate standardization involving $1 - \beta_{y\to x}$ and $1 - \beta_{x\to y}$ and is not covered by the identification expressions stated here, but the corresponding identification results can still be obtained by a derivation analogous to that of Corollary 3.

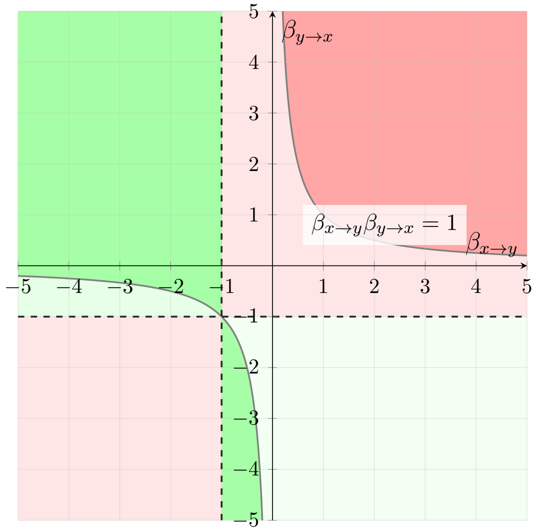


Figure S1: Partition of the plane of $(\beta_{x\to y}, \beta_{y\to x})$ into four regions based on the constraints corresponding to the candidate pairs in Corollary 3. Different colors distinguish the extent of each region. The solid curve represents $\beta_x\beta_y = 1$, and the dashed lines represent $\beta_x = -1$ and $\beta_y = -1$.

C.5 Parameterization using $\eta_0 = \eta/\mu_{xz}$ and $\delta_0 = \delta/\mu_{yw}$ under $\gamma_1 = \gamma_2 = 1$

This subsection returns to the original main-text parameterization $\eta_0 = \eta/\mu_{xz}$ and $\delta_0 = \delta/\mu_{yw}$. It is distinct from the Corollary 3-specific signal-to-noise parameterization $\eta_0 = \eta/\sigma$ and $\delta_0 = \delta/\sigma$ used in Section C.4; the same symbols are retained here only to match the corresponding main-text notation. In this section, we discuss the IV sensitivity analysis under the parameter combinations specified in Corollary 3, using the same parameterization approach as in Section C.1. Specifically, we define $\eta_0 = \eta/\mu_{xz}$ and $\delta_0 = \delta/\mu_{yw}$. Therefore, under this parameterization, the coefficients of Z and W in Equation (S7) have the following relationship:

$$\begin{cases} \dfrac{\xi_{xz}}{\xi_{yz}} = \dfrac{1+\beta_{y\to x}\eta_0}{1+\beta_{y\to x}} \times \dfrac{1+\beta_{x\to y}}{\beta_{x\to y}+\eta_0}, \\ \\ \dfrac{\xi_{xw}}{\xi_{yw}} = \dfrac{\delta_0+\beta_{y\to x}}{1+\beta_{y\to x}} \times \dfrac{1+\beta_{x\to y}}{\beta_{x\to y}\delta_0+1}. \end{cases} \tag{S9}$$

Similar to the calculation process in Section C.1, we can obtain a quadratic equation in one unknown with respect to $\beta_{y\to x}$:

$$p_4\beta_{y\to x}^2 + p_5\beta_{y\to x} + p_6 = 0,$$

where

$$\begin{cases} p_4 &= \{(\xi_{xz}/\xi_{yz}) - 1\}\{1 - (\xi_{xw}/\xi_{yw})\delta_0\}\eta_0 + \{(\xi_{xw}/\xi_{yw}) - 1\}\{(\xi_{xz}/\xi_{yz}) - \eta_0\}, \\ p_5 &= \{1 - (\xi_{xw}/\xi_{yw})\}\{(\xi_{xz}/\xi_{yz}) - 1\}\eta_0\delta_0 + (\delta_0 - \eta_0)\{(\xi_{xw}/\xi_{yw}) - (\xi_{xz}/\xi_{yz})\} \\ & \quad +\{1 - (\xi_{xz}/\xi_{yz})(\xi_{xw}/\xi_{yw})\}\delta_0\eta_0 + 2(\xi_{xz}/\xi_{yz})(\xi_{xw}/\xi_{yw}) - (\xi_{xz}/\xi_{yz}) - (\xi_{xw}/\xi_{yw}), \\ p_6 &= \{(\xi_{xz}/\xi_{yz}) - (\xi_{xz}/\xi_{yz})(\xi_{xw}/\xi_{yw})\}\eta_0\delta_0 + (\delta_0 - 1)(\xi_{xw}/\xi_{yw}) + (\xi_{xz}/\xi_{yz})\{(\xi_{xw}/\xi_{yw}) - \delta_0\}. \end{cases}$$

By solving the quadratic equation, we obtain two causal effect expressions:

$$\beta_{y\to x,1} = \frac{-p_5 + \sqrt{\Delta}}{2p_4}, \quad \beta_{y\to x,2} = \frac{-p_5 - \sqrt{\Delta}}{2p_4},$$

where $\Delta = p_5^2 - 4p_4p_6$. Furthermore, substituting the expression for $\beta_{y\to x}$

into the relationship Equation (S9) yields the corresponding expression for $\beta_{x \to y}$.

As can be seen from the above, the causal effect identification expressions under this parameterization are relatively complex, and for the two solutions derived from the quadratic equation, we need to obtain certain prior information to make a selection.

D. Additional simulation results

D.1 Additional simulation results in Section 4.1

Table S2 presents estimation results for both methods across different scenarios and sample sizes, supplementing Table 1 in the main text.

D.2 Additional simulation results in Section 4.3

In this section, we present the additional simulation results for the sensitivity analysis discussed in Section 4.3.

Figures S2 to S7 present the simulation results for three specific sensitivity parameter combinations, demonstrating the corresponding performance across different sample sizes and data generation scenarios. The results show that the point estimates consistently remain stable around the given true values. Although some variations in errors are observed across differ-

D.2 Additional simulation results in Section 4.3

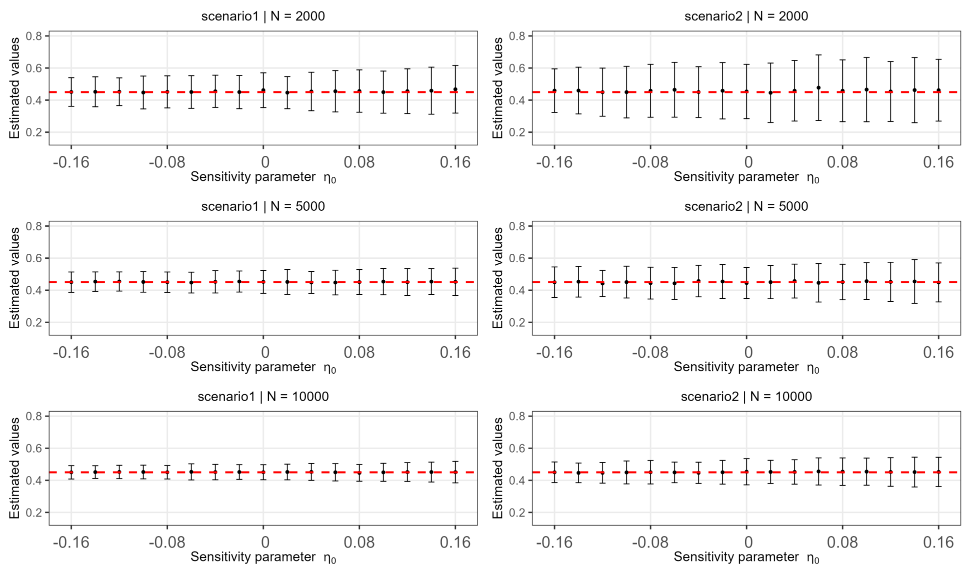


Figure S2: Estimation results of $\beta_{x\to y}$ under the setting of Corollary 1.

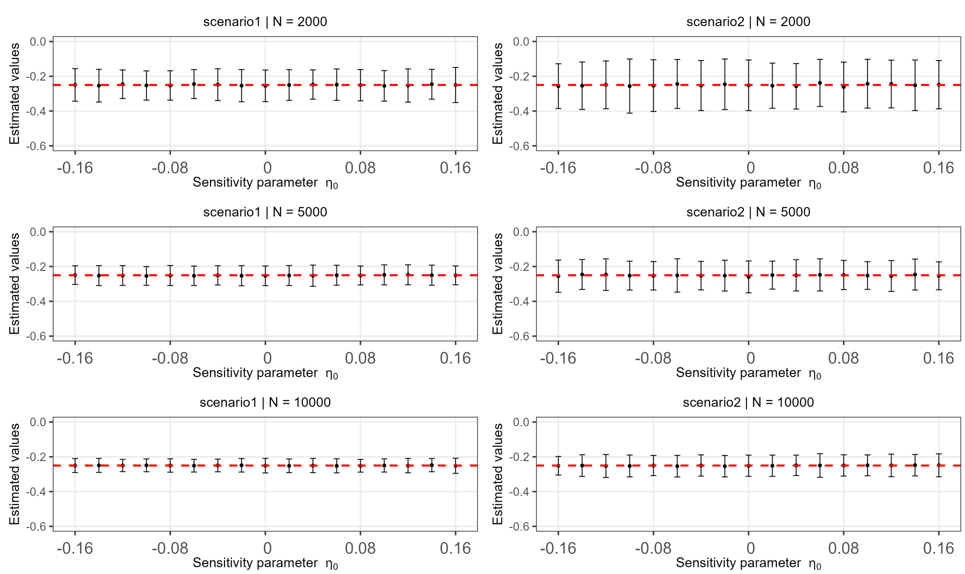


Figure S3: Estimation results of $\beta_{y\to x}$ under the setting of Corollary 1.

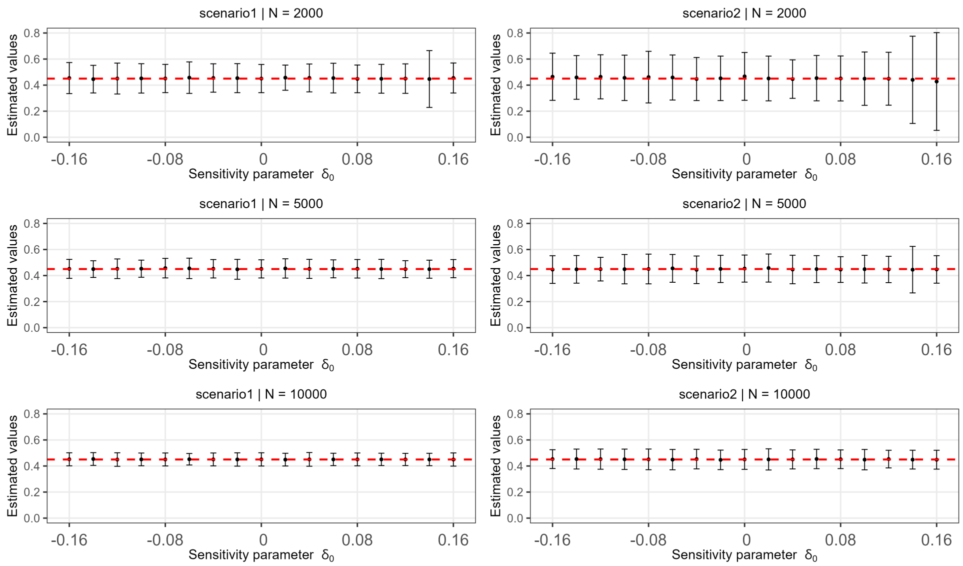


Figure S4: Estimation results of $\beta_{x\to y}$ under the setting of Corollary 2.

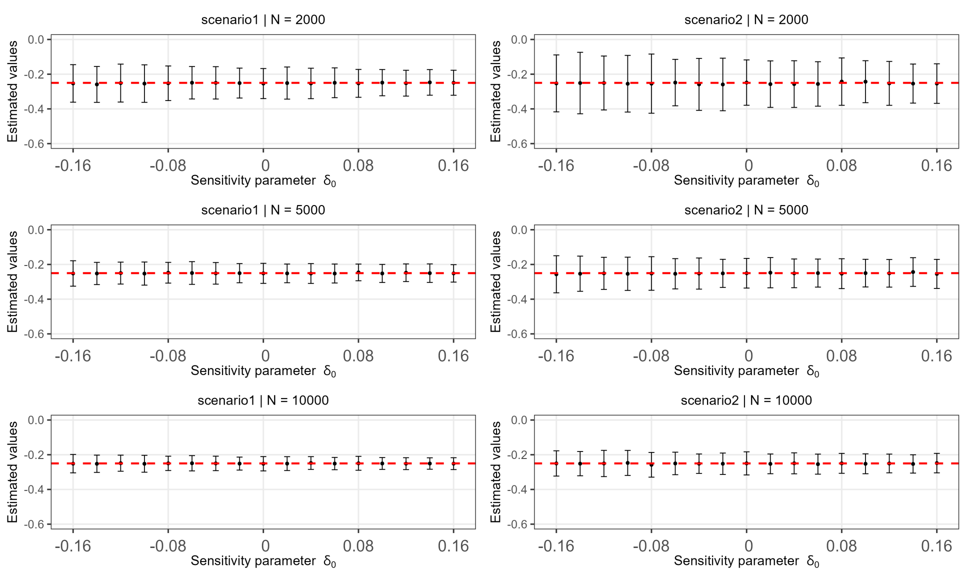


Figure S5: Estimation results of $\beta_{y\to x}$ under the setting of Corollary 2.

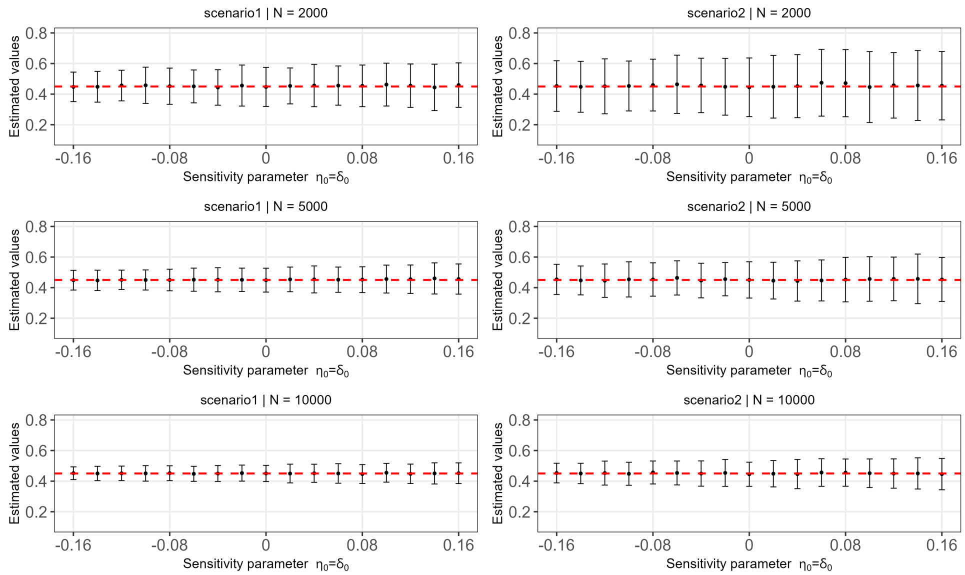


Figure S6: Estimation results of $\beta_{x\to y}$ under the setting of Corollary 3.

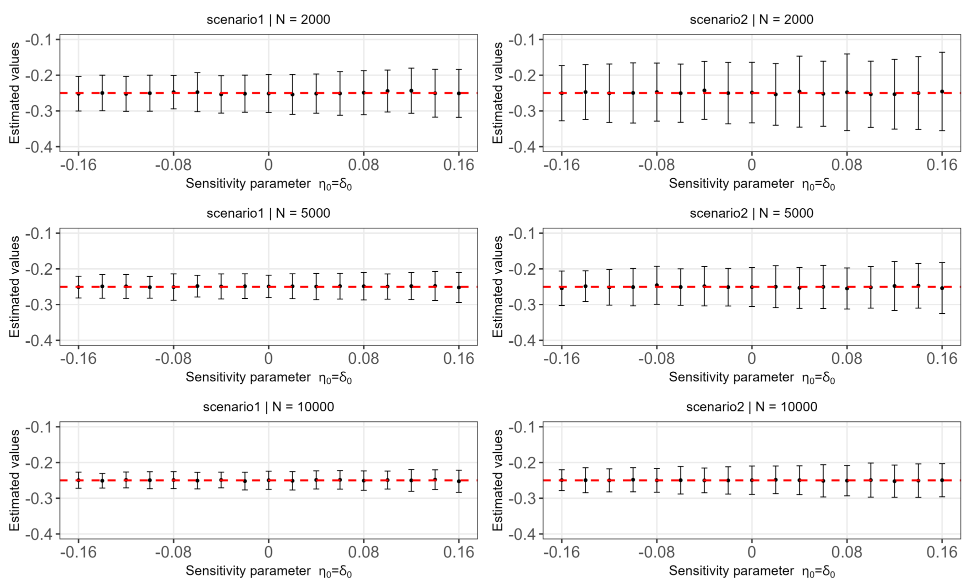


Figure S7: Estimation results of $\beta_{y\to x}$ under the setting of Corollary 3.

Table S2: Simulation study results: standard deviation and bias of bidirectional causal effect estimators across scenarios and methods. All values are multiplied by 10^4.

			Scenario 1			Scenario 2		
	Sample Size		2000	5000	10000	2000	5000	10000
GLS	$\hat{\beta}_{x\to y}$	bias	-1510	-1550	-1540	-1640	-1590	-1570
		sd	724	475	358	635	379	288
	$\hat{\beta}_{y\to x}$	bias	5490	5450	5460	5370	5420	5440
		sd	718	480	355	637	383	289
IV	$\hat{\beta}_{x\to y}$	bias	-2	-9	10	80	25	-8
		sd	535	355	242	941	504	407
	$\hat{\beta}_{y\to x}$	bias	-45	-9	-7	10	-29	-31
		sd	500	265	208	755	433	354

ent data generation scenarios, the overall performance remains relatively stable, with estimation bias decreasing as the sample size increases. The simulation results under these three parameter combinations indicate that the proposed method exhibits robustness against potential fluctuations resulting from different instrumental variable selections.

D.3 List of variables in Section 5.1

Table S3: Descriptive statistics for heart disease and diabetes variables.

Variable	Type	Mean/Rate	Variable	Type	Mean/Rate
BMI	Continuous	28.325	Race	Categorical	1.564
Smoking	Binary	41.25%	AgeCategory	Categorical	7.515
AlcoholDrinking	Binary	6.81%	PhysicalActivity	Binary	77.54%
Stroke	Binary	3.77%	GenHealth	Categorical	2.595
PhysicalHealth	Categorical	3.372	SleepTime	Categorical	7.097
MentalHealth	Categorical	3.898	Asthma	Binary	13.41%
DiffWalking	Binary	13.89%	KidneyDisease	Binary	3.68%
Sex	Binary	47.53%	SkinCancer	Binary	9.32%

Table S3 presents the variable names and their means or proportions in the heart disease dataset.

E. Interpretations of latent continuous variables

We clarify the setting of the latent continuous variables in this paper from two aspects:

1. Falconer (1965) proposed a statistical framework that transforms binary disease data into quantifiable genetic parameters. The core assumption of this framework is that each individual possesses a continuous, unobservable latent variable—liability. Liability integrates

a large number of genetic predispositions and environmental exposures and follows a normal distribution in the population. Meanwhile, there exists a fixed threshold: when an individual's liability exceeds this threshold, the disease manifests as "affected" (1); otherwise, it is "unaffected" (0). Therefore, within this framework, the latent continuous variables X° and Y° in our structural equation model (2.1) represent the liability of heart disease and diabetes, respectively. Once these liabilities exceed certain inherent thresholds, the individual is considered diseased.

2. Heart disease and diabetes have well-established clinical diagnostic criteria. For example, an important diagnostic indicator for coronary heart disease is coronary artery stenosis $\geq 50\%$ (Gould et al., 1974); the diagnostic criteria for diabetes consist of either fasting plasma glucose (FPG) ≥ 126 mg/dL or glycated hemoglobin (HbA1c) $\geq 6.5\%$ (The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus, 1997; The International Expert Committee, 2009). However, due to data collection constraints—for instance, in retrospective studies, the dataset may not contain these continuous diagnostic indicators but only record binary disease status—the diagnostic criteria can be regarded as unobservable latent variables in such cases.

Nonetheless, these latent variables still map to the observed binary disease indicators through the thresholds defined by the diagnostic criteria.

In summary, the binary bidirected causal model (2.1) developed in this paper implies a data-generating process in which the target continuous variables are mapped to the target binary variables via threshold constraints. From an observational perspective, we evaluate causal effects using only the binary outcome data, without requiring more detailed data on the underlying latent continuous variables.

F. Interpretations of binary bidirectional model

The structural equation model (2.2) is given by:

$$X^{\circ} = \mu_{x0} + \beta_{y\to x}Y^{\circ} + \mu_{xz}Z + U, \quad X = \mathbb{I}(X^{\circ} > 0),$$

$$Y^{\circ} = \mu_{y0} + \beta_{x\to y}X^{\circ} + \mu_{yw}W + V, \quad Y = \mathbb{I}(Y^{\circ} > 0).$$

This model describes the bidirectional causal relationship in an equilibrium state. Let $\boldsymbol{O} = (X^{\circ}, Y^{\circ})^{\mathrm{T}}, \boldsymbol{\pi} = (Z, W)^{\mathrm{T}}, \ \ \boldsymbol{\varepsilon} = (U, V)^{\mathrm{T}}, \ \ \boldsymbol{\mu}_0 = (\mu_{x0}, \ \ \mu_{y0})^{\mathrm{T}}$,

$$\boldsymbol{\mu}_{zw} = \begin{pmatrix} \mu_{xz} & 0 \\ 0 & \mu_{yw} \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} 1 & -\beta_{y\to x} \\ -\beta_{x\to y} & 1 \end{pmatrix}.$$

When the coefficient matrix $\boldsymbol{B}$ is nonsingular, i.e., $\beta_{x\to y}\beta_{y\to x} \neq 1$, we can obtain the matrix form (S10) corresponding to the reduced form (2.3):

$$\boldsymbol{O} = B^{-1}\boldsymbol{\mu}_0 + B^{-1}\boldsymbol{\mu}_{zw}\boldsymbol{\pi} + B^{-1}\boldsymbol{\varepsilon}, \quad X = \mathbb{I}(X^\circ > 0), Y = \mathbb{I}(Y^\circ > 0). \tag{S10}$$

When both $\beta_{x\to y}$ and $\beta_{y\to x}$ are non-zero, model (2.2) characterizes a bidirectional causal relationship in equilibrium and also corresponds to a cyclic graph. When $\beta_{y\to x}$ is zero, model (2.2) reduced to a unidirectional causal framework.

Referring to the discussion in Li and Ye (2024) on a potential dynamic process for causal interpretation of cyclic graphs in linear structural equation models, we provide a corresponding explanation for the bidirectional causal framework with binary outcomes in this paper. Let $\boldsymbol{O}_t = (X_t^\circ, Y_t^\circ)^{\mathrm{T}}$ denote the observed traits at time t, and consider the structural causal equation:

$$\boldsymbol{O}_0 = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_{zw}\boldsymbol{\pi} + \boldsymbol{\varepsilon}_{\mathbf{0}} \tag{S11}$$

where $\boldsymbol{\varepsilon}_{\mathbf{0}} = (U_0, V_0)^{\mathrm{T}}$ and for all $t \geq 1$,

$$\boldsymbol{O}_t = \boldsymbol{\mu}_{\mathbf{0}} + \boldsymbol{C}\boldsymbol{O}_{t-1} + \boldsymbol{\mu}_{zw}\boldsymbol{\pi} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{C} = \begin{pmatrix} 0 & \beta_{y\to x} \\ \beta_{x\to y} & 0 \end{pmatrix}, \tag{S12}$$

where $\beta_{y\to x}$ and $\beta_{x\to y}$ are the causal effects of Y_{t-1}° on X_t° and X_{t-1}° on Y_t°, respectively. As t goes to infinity, we have:

$$\lim_{t\to\infty} \boldsymbol{O}_t = \lim_{t\to\infty} \sum_{s=0}^{t} \boldsymbol{C}^{\boldsymbol{s}}(\boldsymbol{\mu_0} + \boldsymbol{\mu}_{zw}\boldsymbol{\pi} + \boldsymbol{\varepsilon_s}) = B^{-1}(\boldsymbol{\mu_0} + \boldsymbol{\mu}_{zw}\boldsymbol{\pi}) + \tilde{\boldsymbol{\varepsilon}}$$

where $\tilde{\boldsymbol{\varepsilon}} = \lim_{t\to\infty} \sum_{s=0}^{t} \boldsymbol{C}^s \boldsymbol{\varepsilon}_s$ and the last step is from $\lim_{t\to\infty} \sum_{s=0}^{t} \boldsymbol{C}^s = (\boldsymbol{I}_2 - \boldsymbol{C})^{-1} = \boldsymbol{B}^{-1}$, which requires that the spectral radius of $\boldsymbol{C}$ is less than 1, i.e., $|\beta_{x\to y}\beta_{y\to x}| < 1$. Meanwhile, for $\lim_{t\to\infty} \boldsymbol{O}_t$ to exist, $\tilde{\boldsymbol{\varepsilon}}$ must have a well-defined distribution. For example, consider a first-order moving average (MA) process for $\boldsymbol{\varepsilon}_t$, i.e., $\boldsymbol{\varepsilon}_t = \boldsymbol{\alpha}\boldsymbol{\epsilon}_{t-1} + \boldsymbol{\epsilon}_t$, where $\boldsymbol{\epsilon}_t$ is white noise with covariance matrix Σ_ϵ. Then $\mathrm{E}(\tilde{\boldsymbol{\varepsilon}}) = 0$ and

$$\mathrm{Var}(\tilde{\boldsymbol{\varepsilon}}) = \boldsymbol{\alpha}\Sigma_\epsilon(\boldsymbol{\alpha})^{\mathrm{T}} + (I + \boldsymbol{C}\boldsymbol{\alpha}) \sum_{s=0}^{\infty} \boldsymbol{C}^s \Sigma_\epsilon (\boldsymbol{C}^s)^{\mathrm{T}} (I + \boldsymbol{C}\boldsymbol{\alpha})^{\mathrm{T}},$$

which is a positive definite matrix given that $|\beta_{x\to y}\beta_{y\to x}| < 1$.

The above derivation shows that, under the conditions $|\beta_{x\to y}\beta_{y\to x}| \leq 1$ and $\tilde{\boldsymbol{\varepsilon}}$ being well-defined, the causal effects captured by the parameters $\beta_{x\to y}$ and $\beta_{y\to x}$ can be interpreted from a time-series perspective (for more details, see Li and Ye (2024)).

G. Data generation process and simulation

In this section, we conduct a simulation experiment to achieve the theoretical convergence discussed in Section F. This simulation experiment is a more detailed version of the simulation experiment carried out in Section 4 of the main text, demonstrating the bidirectional causal mechanism, and it more clearly clarifies why $\beta_{x\to y}$ and $\beta_{y\to x}$ are interpreted as bidirectional causal parameters.

G.1 Data Generating Process

1. **The bidirectional iterative process.** For each individual i, based on the iterative mechanism given by model (S13), the main variables (X, Y) are generated as follows:

$$\begin{aligned} X^{\circ}_{t,i} &= \mu_{x0} + \beta_{y\to x} Y^{\circ}_{t-1,i} + \mu_{xz} Z_i + U_i, \\ Y^{\circ}_{t,i} &= \mu_{y0} + \beta_{x\to y} X^{\circ}_{t-1,i} + \mu_{yw} W_i + V_i, \end{aligned} \tag{S13}$$

   where t denotes the iteration step ($t = 1, 2, \ldots, 5000$) and i indexes the sample individuals. This process reflects the bidirectional causal relationship between the main variables X° and Y°: in each iteration, $X^{\circ}_{t,i}$ depends on $Y^{\circ}_{t-1,i}$ from the previous iteration (i.e., the causal effect of $Y \to X$), and $Y^{\circ}_{t,i}$ depends on $X^{\circ}_{t-1,i}$ from the previous iteration (i.e., the causal effect of $X \to Y$). Furthermore, this iterative process starts with $X^{\circ}_{0,i}$ and $Y^{\circ}_{0,i}$ independently drawn from $N(0, 1)$, and iterates 5000 times to allow the system to reach equilibrium. The final iteration yields $(X^{\circ}_{5000,i}, Y^{\circ}_{5000,i})$ as the observed values of $(X^{\circ}_i, Y^{\circ}_i)$ for individual i, i.e., $(X^{\circ}_i, Y^{\circ}_i) = (X^{\circ}_{5000,i}, Y^{\circ}_{5000,i})$. Then, according to the threshold restrictions $X = \mathbb{I}(X^\circ > 0)$ and $Y = \mathbb{I}(Y^\circ > 0)$, we obtain the observed values of the binary variables (X_i, Y_i) for individual i.

2. **Fixed components that ensure convergence.** In the iterative process for each individual i, we keep the background variables (Z_i, W_i, Q_i) (as in Section 4.1, where Q_i denotes omitted observable covariates) and the unobserved confounding variables (U_i, V_i) fixed, so that only $X^\circ_{t,i}$ and $Y^\circ_{t,i}$ vary with t in each iteration. The sample size, the scenarios for generating background variables, and the true parameter values in the simulation experiment are all set to be consistent with those in Section 4.1, and the distribution of the unobserved confounding variables (U_i, V_i) is set to the normal distribution under Assumption 1.

3. **Why the iterative system converges.** In the bidirectional causal model discussed in Section F, we consider the unobserved confounders $\boldsymbol{\varepsilon}_t$ that vary with t, as shown in models (S11) and (S12). However, in the iterative data generating process based on model (S13), we fix the values of the unobserved confounders (U, V) for each individual, so that they do not vary with t. This simplification streamlines the data generating process while still satisfying the basic conditions required for the system to converge to a reduced form. When the condition

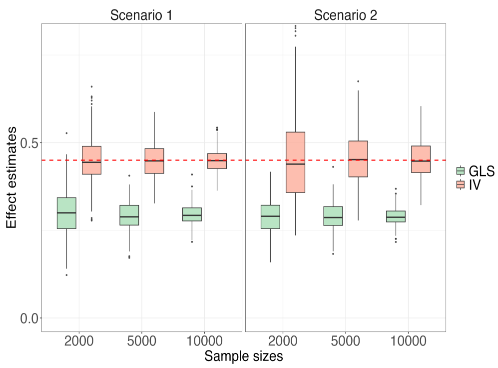
$|\beta_{x\to y}\beta_{y\to x}| < 1$ holds, the system converges to a unique equilibrium point:

$$\lim_{t\to\infty}(X_{t,i}, Y_{t,i})^{\mathrm{T}} = B^{-1}(\boldsymbol{\mu_0} + \boldsymbol{\mu}_{zw}\boldsymbol{\pi}) + \tilde{\boldsymbol{\varepsilon}},$$

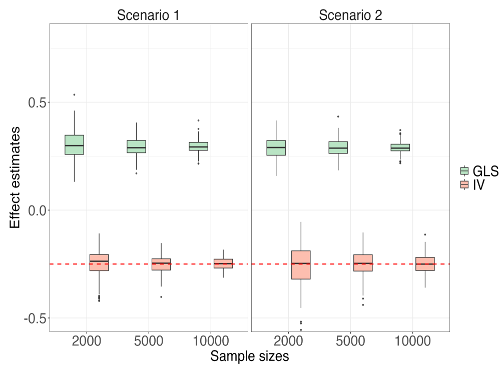
where $\tilde{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon}\sum_{s=0}^{\infty}\boldsymbol{C}^s$ and $\mathrm{Var}(\tilde{\boldsymbol{\varepsilon}}) = \sigma^2\sum_{s=0}^{\infty}C^s(C^s)^T$ is finite under the condition $|\beta_{x\to y}\beta_{y\to x}| < 1$(under Assumption 1, $\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}_2$). Therefore, the convergence as described in Section F is ensured. In the simulation experiment, we set the number of iterations to 5000 to ensure that the system reaches equilibrium. The data set (X, Y, Z, W, Q) obtained after 5000 iterations then correspond to the equilibrium state of Model (2.2). Therefore, the parameters $\beta_{x\to y}$ and $\beta_{y\to x}$ in Model (2.2) represent the causal effect of X on Y and the causal effect of Y on X, respectively.

G.2 Simulation results

We conducted the same numerical simulation experiment on the simulated dataset (X, Y, Z, W, Q) generated by the above process as in Section 4.1.



(a) Simulation results of $\beta_{x\to y}$.



(b) Simulation results of $\beta_{y\to x}$.

Figure S8: Boxplot of simulation results based on 5000 iterations of data.

Figure S8 presents boxplots of the simulation results for $\beta_{x\to y}$ and $\beta_{y\to x}$, respectively, and Table S4 summarizes the corresponding numerical simulation results. The estimation results shown in Figure S8 and Table S4 demonstrate, on the one hand, the reliability of Algorithm 1 and its robustness under finite sample sizes; on the other hand, when compared with the results shown in Figure 3 and Table 1 of the main text, the results of the two simulation experiments are essentially consistent, further verifying the convergence of the above iterative process and the eventual attainment of the equilibrium state described in Model (2.2).

Table S4: Numerical results based on 5000 iterations of data. All values are multiplied by 10^4.

			Scenario 1			Scenario 2		
	Sample Size		2000	5000	10000	2000	5000	10000
GLS	$\hat{\beta}_{x\to y}$	bias	-1507	-1579	-1550	-1619	-1602	-1605
		sd	668	401	295	515	398	252
	$\hat{\beta}_{y\to x}$	bias	5494	5423	5453	5380	5401	5397
		sd	662	401	296	514	398	254
IV	$\hat{\beta}_{x\to y}$	bias	19	-24	-7	100	43	39
		sd	702	439	327	1307	723	569
	$\hat{\beta}_{y\to x}$	bias	17	-33	21	-66	16	-2
		sd	628	418	269	953	589	441

H. Discussion of Assumption 1

H.1 Stable distribution

Here we propose Assumption S1 to replace Assumption 1 in the main text, and discuss the identifiability of bidirectional causal effects under Assumption S1.

Assumption S1. *(i)* $U \sim \mathrm{S}_\alpha(\sigma_s, 0, 0)$ *and* $V \sim \mathrm{S}_\alpha(\sigma_s, 0, 0)$, $\alpha \in (0, 2]$.

α is the stability index and σ_s is an unknown scale parameter in the Stable distribution. (ii) U and V are independent.

The stable distribution family in Assumption S1 includes several important special cases, such as the Gaussian distribution ($\alpha = 2$) and the Cauchy distribution ($\alpha = 1$). A defining feature of this family is its closure under linear combinations: if U and V follow the same Stable distribution, then $aU + bV$ is also Stable with the same stability index α. This stability property plays a central role in our identification arguments, and we further discuss some important distributions in Section H.2.

Under Model (2.2) and Assumption S1 by using the link function induced by the Stable distribution and its stability, we obtain a conditional probability Model (S14):

$$\begin{aligned} \mathrm{pr}(X = 1 \mid Z, W) &= \mathrm{pr}(X^\circ > 0 \mid Z, W) = \mathrm{F_s}(\xi_{x0} + \xi_{xz} Z + \xi_{xw} W), \\ \mathrm{pr}(Y = 1 \mid Z, W) &= \mathrm{pr}(Y^\circ > 0 \mid Z, W) = \mathrm{F_s}(\xi_{y0} + \xi_{yz} Z + \xi_{yw} W), \end{aligned} \tag{S14}$$

where $\mathrm{F_s}(\cdot)$ is the cumulative distribution function of the Stable distribution, $\xi_{x0} = \theta_{x0}/\{\sigma_s(|\theta_{xu}|^\alpha + |\theta_{xv}|^\alpha)^{1/\alpha}\}$, $\xi_{xz} = \theta_{xz}/\{\sigma_s(|\theta_{xu}|^\alpha + |\theta_{xv}|^\alpha)^{1/\alpha}\}$, $\xi_{xw} = \theta_{xw}/\{\sigma_s(|\theta_{xu}|^\alpha + |\theta_{xv}|^\alpha)^{1/\alpha}\}$; $\xi_{y0} = \theta_{y0}/\{\sigma_s(|\theta_{yu}|^\alpha + |\theta_{yv}|^\alpha)^{1/\alpha}\}$, $\xi_{yz} = \theta_{yz}/\{\sigma_s(|\theta_{yu}|^\alpha + |\theta_{yv}|^\alpha)^{1/\alpha}\}$, $\xi_{yw} = \theta_{yw}/\{\sigma_s(|\theta_{yu}|^\alpha + |\theta_{yv}|^\alpha)^{1/\alpha}\}$. Here we still use the same notation as before, but it can be seen that the specific expressions of the series of coefficients $\xi_{\{\cdot\}}$ under Assumption S1 are already different. We summarize the identification and estimation results under Assumption S1 in Proposition S3.

Proposition S3. *Under Model (2.2) and Assumption S1, given $\alpha \in (0,2)$, the bidirectional causal effects are identifiable:*

$$\begin{aligned}\beta_{x\to y} &= \frac{\xi_{yz}}{\xi_{xz}}\left(\frac{|\xi_{xw}|^{\alpha}|\xi_{xz}|^{\alpha}-|\xi_{yw}|^{\alpha}|\xi_{xz}|^{\alpha}}{|\xi_{yz}|^{\alpha}|\xi_{yw}|^{\alpha}-|\xi_{yw}|^{\alpha}|\xi_{xz}|^{\alpha}}\right)^{1/\alpha},\\ \beta_{y\to x} &= \frac{\xi_{xw}}{\xi_{yw}}\left(\frac{|\xi_{yz}|^{\alpha}|\xi_{yw}|^{\alpha}-|\xi_{yw}|^{\alpha}|\xi_{xz}|^{\alpha}}{|\xi_{xw}|^{\alpha}|\xi_{xz}|^{\alpha}-|\xi_{yw}|^{\alpha}|\xi_{xz}|^{\alpha}}\right)^{1/\alpha}.\end{aligned} \tag{S15}$$

where the terms ξ_{xz}, ξ_{xw}, ξ_{yz}, and ξ_{yw} are the coefficients in Model (S14)*. Furthermore, we obtain two bidirectional causal effects estimators:*

$$\begin{aligned}\hat\beta_{x\to y} &= \frac{\hat\xi_{yz}}{\hat\xi_{xz}}\left(\frac{|\hat\xi_{xw}|^{\alpha}|\hat\xi_{xz}|^{\alpha}-|\hat\xi_{yw}|^{\alpha}|\hat\xi_{xz}|^{\alpha}}{|\hat\xi_{yz}|^{\alpha}|\hat\xi_{yw}|^{\alpha}-|\hat\xi_{yw}|^{\alpha}|\hat\xi_{xz}|^{\alpha}}\right)^{1/\alpha},\\ \hat\beta_{y\to x} &= \frac{\hat\xi_{xw}}{\hat\xi_{yw}}\left(\frac{|\hat\xi_{yz}|^{\alpha}|\hat\xi_{yw}|^{\alpha}-|\hat\xi_{yw}|^{\alpha}|\hat\xi_{xz}|^{\alpha}}{|\hat\xi_{xw}|^{\alpha}|\hat\xi_{xz}|^{\alpha}-|\hat\xi_{yw}|^{\alpha}|\hat\xi_{xz}|^{\alpha}}\right)^{1/\alpha}.\end{aligned} \tag{S16}$$

For the identification Equation (S16), the same estimation strategy as in Algorithm S1 can be obtained analogously, and the corresponding asymptotic properties can be established. Proposition S3 actually provides the identification and estimation results of bidirectional causal effects under a class of stable distributions. Therefore, for any $\alpha \in (0,2]$, we can point identify the bidirectional causal effects via Equation (S15) under the corresponding stable distribution assumption. It is worth noting that when $\alpha = 2$, the corresponding stable distribution is exactly the normal distribution. Hence, the identification result under $\alpha = 2$ coincides with that under Assumption 1 in the main text. Below we provide the detailed proof of Proposition S3.

Proof. Given the Model (2.2):

$$X^\circ = \mu_{x0} + \beta_{y\to x} Y^\circ + \mu_{xz} Z + U, \quad X = \mathbb{I}(X^\circ > 0),$$

$$Y^\circ = \mu_{y0} + \beta_{x\to y} X^\circ + \mu_{yw} W + V, \quad Y = \mathbb{I}(Y^\circ > 0),$$

we can derive the reduced form:

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

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

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

Assumption S1 requires that the unobserved confounders U and V are independent and identically distributed as $\mathrm{S}_\alpha(\sigma_s, 0, 0)$. Therefore, according to the relevant properties of the Stable distribution, any linear combination of U and V still follows a Stable distribution, and thus we can obtain:

$$\theta_{xu} U + \theta_{xv} V \sim \mathrm{S}_\alpha(\sigma_s(|\theta_{xu}|^\alpha + |\theta_{xv}|^\alpha)^{1/\alpha}, 0, 0),$$

$$\theta_{yu} U + \theta_{yv} V \sim \mathrm{S}_\alpha(\sigma_s(|\theta_{yu}|^\alpha + |\theta_{yv}|^\alpha)^{1/\alpha}, 0, 0).$$

Therefore, we can obtain the conditional probability Model (S14) and we can identify the ratios ξ_{xz}, ξ_{xw}, ξ_{yz} and ξ_{yw} through the observed data. Here we focus on four key parameters:

$$\xi_{xz} = \frac{c\mu_{xz}}{\sigma_s(|\theta_{xu}|^\alpha + |\theta_{xv}|^\alpha)^{1/\alpha}}, \quad \xi_{xw} = \frac{c\beta_{y\to x}\mu_{yw}}{\sigma_s(|\theta_{xu}|^\alpha + |\theta_{xv}|^\alpha)^{1/\alpha}},$$
$$\xi_{yz} = \frac{c\beta_{x\to y}\mu_{xz}}{\sigma_s(|\theta_{yu}|^\alpha + |\theta_{yv}|^\alpha)^{1/\alpha}}, \quad \xi_{yw} = \frac{c\mu_{yw}}{\sigma_s(|\theta_{yu}|^\alpha + |\theta_{yv}|^\alpha)^{1/\alpha}}.$$

We then similarly define the coefficient ratios: $k_1 = \xi_{yz}/\xi_{xz}$, and $k_2 = \xi_{xw}/\xi_{yw}$. These ratios can be directly estimated from the regression coefficients of Model (S14). After substituting the specific expressions of the coefficients, we obtain:

$$\begin{aligned} k_1 \equiv \frac{\xi_{yz}}{\xi_{xz}} = \beta_{x\to y}\left(\frac{1+|\beta_{y\to x}|^\alpha}{1+|\beta_{x\to y}|^\alpha}\right)^{1/\alpha}, \\ k_2 \equiv \frac{\xi_{xw}}{\xi_{yw}} = \beta_{y\to x}\left(\frac{1+|\beta_{x\to y}|^\alpha}{1+|\beta_{y\to x}|^\alpha}\right)^{1/\alpha}. \end{aligned} \tag{S17}$$

Furthermore, we can derive the relations between the causal effect parameters and the coefficient ratios k_1, k_2:

$$\begin{cases} |k_1| = |\beta_{x\to y}|\left(\frac{1+|\beta_{y\to x}|^\alpha}{1+|\beta_{x\to y}|^\alpha}\right)^{1/\alpha} \Rightarrow |k_1|^\alpha = |\beta_{x\to y}|^\alpha \frac{1+|\beta_{y\to x}|^\alpha}{1+|\beta_{x\to y}|^\alpha}, \\ |k_2| = |\beta_{y\to x}|\left(\frac{1+|\beta_{x\to y}|^\alpha}{1+|\beta_{y\to x}|^\alpha}\right)^{1/\alpha} \Rightarrow |k_2|^\alpha = |\beta_{y\to x}|^\alpha \frac{1+|\beta_{x\to y}|^\alpha}{1+|\beta_{y\to x}|^\alpha}, \\ |k_1|^\alpha |k_2|^\alpha = |\beta_{x\to y}|^\alpha |\beta_{y\to x}|^\alpha. \end{cases}$$

Based on the relations between k_1 and k_2, we can obtain the following through calculation:

$$
\begin{aligned}
&|k_1|^\alpha \cdot (1 + |\beta_{x\to y}|^\alpha) = |\beta_{x\to y}|^\alpha \cdot (1 + |\beta_{y\to x}|^\alpha) \\
\Rightarrow &|k_1|^\alpha + |k_1|^\alpha|\beta_{x\to y}|^\alpha = |\beta_{x\to y}|^\alpha + |k_1|^\alpha|k_2|^\alpha \\
\Rightarrow &|\beta_{x\to y}|^\alpha = \frac{|k_1|^\alpha|k_2|^\alpha - |k_1|^\alpha}{|k_1|^\alpha - 1} = \frac{|k_1|^\alpha(|k_2|^\alpha - 1)}{|k_1|^\alpha - 1}, \\
&|k_2|^\alpha \cdot (1 + |\beta_{y\to x}|^\alpha) = |\beta_{y\to x}|^\alpha \cdot (1 + |\beta_{x\to y}|^\alpha) \\
\Rightarrow &|k_2|^\alpha + |k_2|^\alpha|\beta_{y\to x}|^\alpha = |\beta_{y\to x}|^\alpha + |k_1|^\alpha|k_2|^\alpha \\
\Rightarrow &|\beta_{y\to x}|^\alpha = \frac{|k_1|^\alpha|k_2|^\alpha - |k_2|^\alpha}{|k_2|^\alpha - 1} = \frac{|k_2|^\alpha(|k_1|^\alpha - 1)}{|k_2|^\alpha - 1}.
\end{aligned}
$$

Equation (S17) implies that $\beta_{x\to y}$ has the same sign as k_1, and $\beta_{y\to x}$ has the same sign as k_2. Therefore, the identification expressions for the bidirectional causal effect parameters can be obtained:

$$
\begin{cases}
\beta_{x\to y} & = \operatorname{sgn}(k_1)|k_1|\left(\dfrac{|k_2|^\alpha - 1}{|k_1|^\alpha - 1}\right)^{1/\alpha} = k_1\left(\dfrac{|k_2|^\alpha - 1}{|k_1|^\alpha - 1}\right)^{1/\alpha} \\
& = \dfrac{\xi_{yz}}{\xi_{xz}}\left(\dfrac{|\xi_{xw}|^\alpha|\xi_{xz}|^\alpha - |\xi_{yw}|^\alpha|\xi_{xz}|^\alpha}{|\xi_{yz}|^\alpha|\xi_{yw}|^\alpha - |\xi_{yw}|^\alpha|\xi_{xz}|^\alpha}\right)^{1/\alpha}, \\
\beta_{y\to x} & = \operatorname{sgn}(k_2)|k_2|\left(\dfrac{|k_1|^\alpha - 1}{|k_2|^\alpha - 1}\right)^{1/\alpha} = k_2\left(\dfrac{|k_1|^\alpha - 1}{|k_2|^\alpha - 1}\right)^{1/\alpha} \\
& = \dfrac{\xi_{xw}}{\xi_{yw}}\left(\dfrac{|\xi_{yz}|^\alpha|\xi_{yw}|^\alpha - |\xi_{yw}|^\alpha|\xi_{xz}|^\alpha}{|\xi_{xw}|^\alpha|\xi_{xz}|^\alpha - |\xi_{yw}|^\alpha|\xi_{xz}|^\alpha}\right)^{1/\alpha}.
\end{cases}
$$

Since k_1 and k_2 are identifiable under Model (S14) and Equation (S17), we can identify the bidirectional causal effects $\beta_{x\to y}$ and $\beta_{y\to x}$ through observed data using Equation (S15). □

H.2 Illustration: Cauchy distribution

H.2.1 Identification

In this subsection, we further discuss important special cases of Assumption S1. Beyond the Gaussian case, the stable family includes other heavy-tailed distributions of practical interest. When the stability index α of the Stable distribution imposed in Assumption S1 equals 1, this distribution reduces to the Cauchy distribution. Now we discuss the corresponding identification results under this special case. Thus, we impose Assumption S2.

Assumption S2. *(i)* $U \sim \mathrm{Cauchy}(0, \gamma_c)$ *and* $V \sim \mathrm{Cauchy}(0, \gamma_c)$*, where* γ_c *is an unknown scale parameter in the Cauchy distribution. (ii)* U *and* V *are independent.*

Under Model (2.2) and Assumption S2, by using the cumulative distribution function of the Cauchy distribution (i.e., the cauchit link), we obtain the following conditional probability model (S18) :

$$\begin{aligned} \mathrm{pr}(X = 1 \mid Z, W) &= \mathrm{pr}(X^\circ > 0 \mid Z, W) = \mathrm{F_c}(\xi_{x0} + \xi_{xz} Z + \xi_{xw} W), \\ \mathrm{pr}(Y = 1 \mid Z, W) &= \mathrm{pr}(Y^\circ > 0 \mid Z, W) = \mathrm{F_c}(\xi_{y0} + \xi_{yz} Z + \xi_{yw} W), \end{aligned} \tag{S18}$$

where $\mathrm{F_c}$ is the cauchit link, $\xi_{x0} = \theta_{x0}/\{\gamma_c(|\theta_{xu}|+|\theta_{xv}|)\}$, $\xi_{xz} = \theta_{xz}/\{\gamma_c(|\theta_{xu}|+|\theta_{xv}|)\}$, $\xi_{xw} = \theta_{xw}/\{\gamma_c(|\theta_{xu}| + |\theta_{xv}|)\}$; $\xi_{y0} = \theta_{y0}/\{\gamma_c(|\theta_{yu}| + |\theta_{yv}|)\}$, $\xi_{yz} =$

$\theta_{yz}/\{\gamma_c(|\theta_{yu}|+|\theta_{yv}|)\}$, $\xi_{yw}=\theta_{yw}/\{\gamma_c(|\theta_{yu}|+|\theta_{yv}|)\}$. These coefficient expressions can all be obtained by setting $\alpha=1$ in the corresponding results in Section H.1, and below we summarize the identification and estimation results under Assumption S2 in Proposition S4.

Proposition S4. *Under Model (2.2) and Assumption S2, the bidirectional causal effects are identifiable:*

$$\beta_{x\to y}=\frac{\xi_{yz}}{\xi_{xz}}\frac{|\xi_{xw}||\xi_{xz}|-|\xi_{yw}||\xi_{xz}|}{|\xi_{yz}||\xi_{yw}|-|\xi_{yw}||\xi_{xz}|},\beta_{y\to x}=\frac{\xi_{xw}}{\xi_{yw}}\frac{|\xi_{yz}||\xi_{yw}|-|\xi_{yw}||\xi_{xz}|}{|\xi_{xw}||\xi_{xz}|-|\xi_{yw}||\xi_{xz}|}.$$

where the terms ξ_{xz}, ξ_{xw}, ξ_{yz}, and ξ_{yw} are the coefficients in Model (S18)*. Furthermore, we obtain two bidirectional causal effects estimators:*

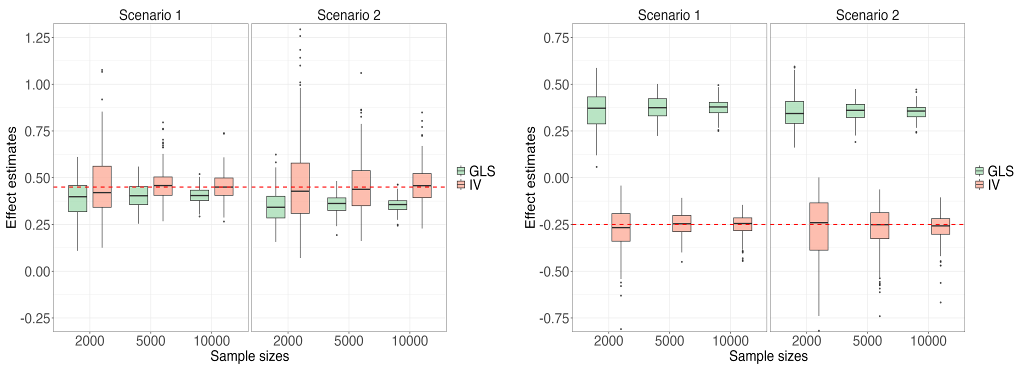
$$\hat{\beta}_{x\to y}=\frac{\hat{\xi}_{yz}}{\hat{\xi}_{xz}}\frac{|\hat{\xi}_{xw}||\hat{\xi}_{xz}|-|\hat{\xi}_{yw}||\hat{\xi}_{xz}|}{|\hat{\xi}_{yz}||\hat{\xi}_{yw}|-|\hat{\xi}_{yw}||\hat{\xi}_{xz}|},\hat{\beta}_{y\to x}=\frac{\hat{\xi}_{xw}}{\hat{\xi}_{yw}}\frac{|\hat{\xi}_{yz}||\hat{\xi}_{yw}|-|\hat{\xi}_{yw}||\hat{\xi}_{xz}|}{|\hat{\xi}_{xw}||\hat{\xi}_{xz}|-|\hat{\xi}_{yw}||\hat{\xi}_{xz}|}. \quad \text{(S19)}$$

For the identification Equation (S19), the same estimation strategy as in Algorithm S1 can be obtained analogously, and the corresponding asymptotic properties can be established. The proof of Proposition S4 is identical to that of Proposition S3 and is therefore omitted here.

H.2.2 Simulation

In this section, we conduct numerical simulation studies on the method proposed in Section H.1 to evaluate its performance under finite sample sizes. Regarding the Cauchy distribution assumption imposed on unobserved confounding in Assumption S2, we set $\gamma_c = 0.75$ in our simulation experiments, i.e., $U, V \sim \text{Cauchy}(0, 0.75)$, while keeping other true values and distribution settings consistent with those in Section 4.1 of the main text. Parallel to the settings in Section 4.1, we consider two estimation methods under different sample sizes and different data-generating scenarios.

We consider two methods for estimation. The first method (GLS) is also a direct regression approach, but unlike Section 4.1, to align with the Cauchy distribution assumption on unobserved confounding, we fit two Cauchit 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 $\hat{\beta}_{x\to y}$ and $\hat{\beta}_{y\to x}$, respectively. The second method (IV) is based on the estimation approach of Proposition S4, i.e., we fit Cauchit binomial regression models of X and Y on Z, W, Q separately and obtain coefficient estimates via maximum likelihood estimation, which are then substituted into Equation (S19) to obtain estimates of the bidirectional causal effects. Furthermore, we adopted a data generating process consistent with Section G to obtain simulated data, and performed 200 replications.



(a) Simulation results of $\beta_{x \to y}$.

(b) Simulation results of $\beta_{y \to x}$.

Figure S9: Boxplot of balance data simulation experiment results under Assumption S2.

Figure S9 presents the boxplots of the estimated bidirectional causal effect parameters under the two methods. As indicated by the horizontal line representing the true value (red line) in Figure S9, the second method (IV) outperforms the first method (GLS) across different data generating scenarios and sample sizes. Table S5 summarizes the corresponding numerical results of the simulation study, presenting the estimation bias and standard deviation under different scenarios and sample sizes. From the numerical values in the table, it can also be seen that the IV method outperforms the GLS method, particularly in terms of bias. Furthermore, the estimated standard deviation decreases as the sample size increases, a trend that can also be observed in Figure S9. Moreover, the simulation results are largely consistent across the two data-generating scenarios, with Scenario 1 showing marginally better overall performance.

Both the simulation results presented in Figure S9 and Table S5 indicate that the proposed method still yields reliable estimates when a Cauchy distribution assumption is imposed on the unobserved confounder, demonstrating that the identification strategy in this paper possesses a certain degree of robustness to the distribution of unmeasured confounding.

H.3 Robustness analysis with respect to the distribution of the unobserved confounding

Regarding the normality assumption imposed in Assumption 1, Section H.1 provides the family of stable distributions, represented by the Cauchy distribution, as an alternative non-Gaussian distribution. Under these non-Gaussian distributions, the proposed identification framework can still identify the bidirectional causal effects. To further discuss the impact of the distributional assumption on unobserved confounding in Assumption 1 on the identifiability of causal effects, we next conduct a robustness analysis of the distribution of unobserved confounding factors. Specifically, we will specify different distributions for the unobserved confounders and apply the identification strategy proposed in Proposition 1 to

Table S5: Simulation study results: standard deviation and bias of bidirectional causal effect estimators across scenarios and methods. All values are multiplied by 10^4.

			Scenario 1			Scenario 2		
	Sample Size		2000	5000	10000	2000	5000	10000
GLS	$\hat{\beta}_{x\to y}$	bias	-609	-478	-455	-993	-911	-957
		sd	1000	635	443	871	532	381
	$\hat{\beta}_{y\to x}$	bias	6106	6240	6259	6005	6077	6028
		sd	1007	635	450	869	537	385
IV	$\hat{\beta}_{x\to y}$	bias	59	181	33	570	101	110
		sd	1605	905	736	4255	1424	1033
	$\hat{\beta}_{y\to x}$	bias	-327	5	-33	-471	-204	-167
		sd	1401	615	544	2612	1201	780

evaluate the bidirectional causal effects.

We selected the t-distribution with 3 and 10 degrees of freedom, the uniform distribution on $[-1, 1]$, the Bernoulli distribution, and the binomial distribution as the distribution types for the unobserved confounders. Figures S10 and S11 present the estimation results of $\beta_{x\to y}$ and $\beta_{y\to x}$ under different distributional assumptions of the unobserved confounders. Two scenarios are considered for generating Z and W: $Z, W \sim N(0,1)$ and $Z, W \sim \text{Unif}(-2,2)$. For each scenario, three sample size settings are given. The true values of the parameters are the same as those in Section 4.1 of the main text. The horizontal red lines in the figures indicate the true values of the causal effect parameters.

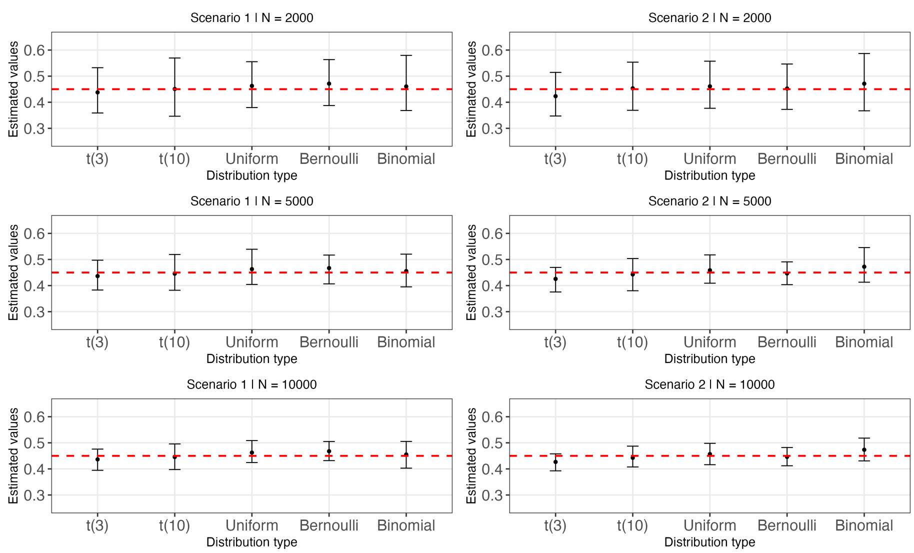


Figure S10: Estimation results of $\beta_{x \to y}$ under different distributions of unobserved confounding.

As shown in Figures S10 and S11, the identification strategy under Assumption 1 still yields satisfactory estimation results when different non-Gaussian distributions are assigned to the unobserved confounders, indicating that the identification method proposed in Proposition 1 is robust to

Table S6: Bias and sd of bidirectional causal effects under different unobserved confounder distributions. All values are multiplied by 10^4.

			Scenario 1			Scenario 2		
	Sample Size		2000	5000	10000	2000	5000	10000
$t(3)$	$\hat{\beta}_{x\to y}$	bias	-122	-137	-134	-268	-244	-233
		sd	460	306	195	430	245	175
	$\hat{\beta}_{y\to x}$	bias	-95	-83	-75	100	15	34
		sd	385	245	170	369	225	162
$t(10)$	$\hat{\beta}_{x\to y}$	bias	7	-45	-45	34	-72	-74
		sd	577	358	263	512	309	209
	$\hat{\beta}_{y\to x}$	bias	-36	-58	-59	-1	2	3
		sd	464	270	194	409	247	180
Uniform	$\hat{\beta}_{x\to y}$	bias	127	130	124	110	88	69
		sd	493	336	216	477	276	201
	$\hat{\beta}_{y\to x}$	bias	43	50	63	-154	-151	-137
		sd	430	245	172	403	234	162
Bernoulli	$\hat{\beta}_{x\to y}$	bias	213	170	177	25	-36	-40
		sd	498	282	189	446	224	177
	$\hat{\beta}_{y\to x}$	bias	100	66	106	-153	-190	-184
		sd	315	213	148	312	216	155
Binomial	$\hat{\beta}_{x\to y}$	bias	104	56	53	211	222	234
		sd	574	359	268	553	345	220
	$\hat{\beta}_{y\to x}$	bias	30	21	40	-91	-79	-53
		sd	448	290	181	390	249	194

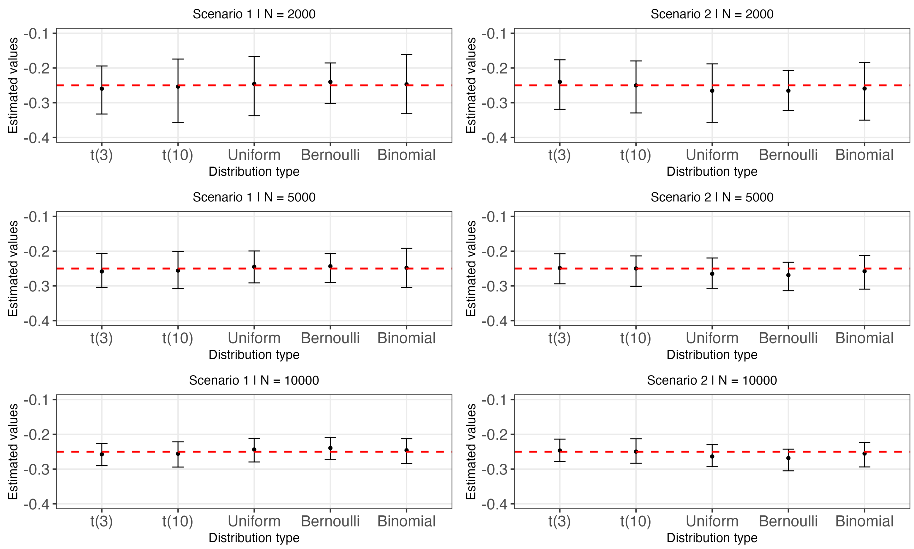


Figure S11: Estimation results of $\beta_{y \to x}$ under different distributions of unobserved confounding.

the distributional assumption on unobserved confounders in Assumption 1. Table S6 summarizes the corresponding numerical results.

References

Falconer, D. S. (1965). The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Ann Hum Genet*, 29(1).

Gould, K. L., Lipscomb, K., and Hamilton, G. W. (1974). Physiologic basis for assessing critical coronary stenosis. instantaneous flow response and regional distribution during coronary hyperemia as measures of coronary flow reserve. *Am J Cardiol*, 33(1):87–94.

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.

The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus (1997). Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care*, 20(7):1183–1197.

The International Expert Committee (2009). International expert committee report on the role of the a1c assay in the diagnosis of diabetes. *Diabetes Care*, 32(7):1327–1334.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*, volume 3. Cambridge University Press.