Supplementary Materials for “Identification and Estimation of Causal Peer Effects Using Double Negative Controls for Unmeasured Network Confounding”

Naoki Egami Eric J. Tchetgen Tchetgen

A Identification

A.1 Proof of Lemma 4

Here, we prove the existence of the outcome confounding bridge.

A.1.1 Setup

To make the discussion general, we use Y to denote the outcome and A to denote the treatment instead of (Y_{12}, Y_{21}) and (Y_{i2}, A_i), which we use in Section 2 and Section 3, respectively. To provide rigorous discussion on the existence of a solution to a Fredholm integral equation of the first kind, we rely on Picard's theorem (Kress, 1989, Theorem 15.18).

Lemma 3 (Picard's theorem (Kress, 1989, Theorem 15.18)) *Given Hilbert spaces $\mathcal{S}_1$ and $\mathcal{S}_2$, let $K : \mathcal{S}_1 \to \mathcal{S}_2$ be a compact operator with singular system $(\nu_p, \upsilon_p, \kappa_p)_{p=1}^{+\infty}$. Define its adjoint to be $K^* : \mathcal{S}_2 \to \mathcal{S}_1$. Then, for $h \in \mathcal{S}_1$ and $\tilde{h} \in \mathcal{S}_2$, there exists a solution to a Fredholm integral equation of the first kind $Kh = \tilde{h}$ if and only if (1) $\tilde{h} \in \mathrm{Null}(K^*)^{\perp}$ and (2) $\sum_{p=1}^{+\infty} \frac{1}{\nu_p^2} |\langle \tilde{h}, \kappa_p \rangle|^2 < +\infty$, where the inner product is defined for a Hilbert space $\mathcal{S}_2$, $\mathrm{Null}(K^*) = \{\tilde{h} : K^* \tilde{h} = 0\}$ is the null space of K^*, and $\perp$ represents the orthogonal complement to a subset.*

To apply Picard's theorem, we need to provide some additional notation. We use F and dF to denote the cumulative distribution function and the Radon-Nikodym derivative of F. We define $L^2\{F(t)\}$ to be the space of all square integrable functions of t with respect to a cumulative distribution function $F(t)$, which is a Hilbert space with the inner product

$$\langle h_1, h_2 \rangle \coloneqq \int_{-\infty}^{+\infty} h_1(t) h_2(t) dF(t) \quad \text{for all } h_1, h_2 \in L^2\{F(t)\}.$$

We define a kernel

$$K(w, u, a, x) = \frac{dF(w, u \mid a, x)}{dF(w \mid a, x) dF(u \mid a, x)}.$$

We then define the linear operators $K_{a,x} : L^2\{F(w \mid a,x)\} \to L^2\{F(u \mid a,x)\}$ by

$$K_{a,x}h = \int_{-\infty}^{+\infty} K(w,u,a,x)h(w)dF(w \mid a,x) = \mathbb{E}\{h(w) \mid u,a,x\}$$

for $h \in L^2\{F(w \mid a,x)\}$. The adjoint of this linear operator $K^*_{a,x} : L^2\{F(u \mid a,x)\} \to L^2\{F(w \mid a,x)\}$ is given by

$$K^*_{a,x}\tilde{h} = \int_{-\infty}^{+\infty} K(w,u,a,x)\tilde{h}(u)dF(u \mid a,x) = \mathbb{E}\{\tilde{h}(u) \mid w,a,x\}$$

for $\tilde{h} \in L^2\{F(u \mid a,x)\}$.

A.1.2 Main Results

We first assume that W is relevant for U.

Assumption 5 (Relevance of W for U) *For any square integrable function f and any a and x, if $\mathbb{E}\{f(U) \mid W = w, A = a, X = x\} = 0$ for almost all w, then $f(U) = 0$ almost surely.*

This is formally known as a completeness condition, and can be interpreted similarly to Assumption 1.4. We also introduce regularity conditions related to the singular value decomposition.

Assumption 6 (Regularity Conditions)

$$\int_{-\infty}^{+\infty} dF(u \mid w,a,x)dF(w \mid u,a,x)dwdu \; < \; +\infty \tag{A.1}$$

$$\int_{-\infty}^{+\infty} \mathbb{E}(Y \mid a,u,x)^2 dF(u \mid a,x)du \; < \; +\infty \tag{A.2}$$

$$\sum_{p=1}^{+\infty} \frac{1}{\nu^2_{a,x,p}} |\langle \mathbb{E}(Y \mid a,u,x), \kappa_{a,x,p}\rangle|^2 \; < \; +\infty \tag{A.3}$$

where $\nu_{a,x,p}$ is the p-th singular value of $K_{a,x}$, and $\kappa_{a,x,p} \in L^2\{F(u \mid a,x)\}$ is an orthogonal sequence.

Lemma 4 *Under Assumptions 5 and 6, there exists a function $h(W, Y_{21}, X)$ such that for all $y_{21} \in \mathcal{Y}_{21}$, equation (5) holds.*

A.1.3 Proof

Under Assumptions 5 and 6, we prove the existence of a solution to the following Fredholm integral equation of the first kind.

$$\mathbb{E}(Y \mid A = a, U = u, X = x) = \mathbb{E}\{h(W,a,x) \mid A = a, U = u, X = x\}. \tag{A.4}$$

First, we can re-write equation (A.4) as follows using notation introduced above.

$$K_{a,x}h = \mathbb{E}(Y \mid A = a, U = u, X = x). \tag{A.5}$$

Therefore, to evoke Picard's theorem, we need to prove (i) $K_{a,x}$ is a compact operator, (ii) $\mathbb{E}(Y \mid A = a, U = u, X = x) \in L^2\{F(u \mid a, x)\}$, (iii) $\mathbb{E}(Y \mid A = a, U = u, X = x) \in \text{Null}(K^*_{a,x})^{\perp}$, and (iv) $\sum_{p=1}^{+\infty} \frac{1}{\nu^2_{a,x,p}} |\langle \mathbb{E}(Y \mid A = a, U = u, X = x), \kappa_{a,x,p}\rangle|^2 < +\infty$, where $\nu_{a,x,p}$ is the p-th singular value of $K_{a,x}$, and $\kappa_{a,x,p} \in L^2\{F(u \mid a, x)\}$ is an orthogonal sequence.

Proof of (i): We note that $K_{a,x}$ and $K^*_{a,x}$ are compact operators under equation (A.1) (Carrasco *et al.*, 2007, Example 2.3 on page 5659). Therefore, there exists a singular system $(\nu_{a,x,p}, \upsilon_{a,x,p}, \kappa_{a,x,p})$ of $K_{a,x}$ according to Kress (1989, Theorem 15.16) where $\nu_{a,x,p}$ is the p-th singular value of $K_{a,x}$, and $\upsilon_{a,x,p} \in L^2\{F(w \mid a, x)\}$ and $\kappa_{a,x,p} \in L^2\{F(u \mid a, x)\}$ are orthogonal sequences.

Proof of (ii): Under equation (A.2), we have $\mathbb{E}(Y \mid a, u, x) \in L^2\{F(u \mid a, x)\}$.

Proof of (iii): We show that $\text{Null}(K^*_{a,x})^{\perp} = L^2\{F(u \mid a, x)\}$. For any $\tilde{h} \in \text{Null}(K^*_{a,x})$, we have $K^*_{a,x}\tilde{h} = \mathbb{E}\{\tilde{h}(u) \mid w, a, x\} = 0$ almost surely by the definition of the null space. Under Assumption 5 (Relevance of W for U), we have $\tilde{h}(u) = 0$ almost surely. Therefore, $\text{Null}(K^*_{a,x})^{\perp} = L^2\{F(u \mid a, x)\}$. Based on (ii), we have $\mathbb{E}(Y \mid a, u, x) \in L^2\{F(u \mid a, x)\}$ under equation (A.2), and therefore, $\mathbb{E}(Y \mid a, u, x) \in \text{Null}(K^*_{a,x})^{\perp}$.

Proof of (iv): Finally, this key condition for Picard's theorem is directly implied by equation (A.3), which completes the proof. □

A.2 Details on Completeness Conditions

In this section, to simplify the discussion, we only focus on two random variables W and Z. We say that Z is complete with respect to W if $\forall f(W) \in L^2\{F(W)\}$,

$$\mathbb{E}\{f(W) \mid Z\} = 0 \text{ almost surely} \implies f(W) = 0 \text{ almost surely}. \tag{A.6}$$

This completeness condition, also known as L^2-completeness, requires that the conditional expectation projection operator $K : L^2\{F(W)\} \to L^2\{F(Z)\}$ be injective (i.e., $\text{Null}(K) = \{0\}$). Intuitively, this means that no information has been lost through projection of W on Z. A necessary and sufficient condition of completeness is given by the following lemma.

Lemma 5 (Severini and Tripathi (2006); Andrews (2017)) *Z is complete with respect* *to W if and only if every non-constant random variable $\lambda(W) \in L^2\{F(W)\}$ is correlated with some random variable $\tilde{\lambda}(Z) \in L^2\{F(Z)\}$.*

This formally captures the notion that completeness ensures that there is no loss of information through projection of W on Z.

As explained in Section 2.3, the completeness condition has been long used in statistics and econometrics. Originally in statistics, Lehmann and Scheffé (2012a,b) introduced the concept of completeness and used it to define estimators with minimal risk within unbiased estimators. They defined completeness as $\mathbb{E}_\theta(f(V)) = 0$ for any $\theta \in \Theta$ implying $f(V) = 0$ a.s. with respect to some parameter space Θ parameterizing the distribution space. Shao (2003) defined completeness with respect to a family of distributions, i.e., $\mathbb{E}_P(f(V)) = 0$ for any $P \in \mathcal{P}$ implying $f(V) = 0$ a.s. with respect to some family of $\mathcal{P}$. In our definition of the completeness (Assumption 1.4 and Assumption 2.4), we set $\mathcal{P}$ to be the conditional distribution. If we define a family of distributions to be $\mathcal{P} = \{F(W \mid Z) : Z \in \mathcal{Z}\}$ of random variable W, the connection between our definition of completeness and the traditional completeness condition given in Lehmann and Scheffé (2012a,b) becomes clear. In particular, we say that a family of distributions $\mathcal{P} = \{F(W \mid Z) : Z \in \mathcal{Z}\}$ of random variable W is complete with respect to Z if $\forall f(W) \in L^2\{F(W)\}$, $\mathbb{E}_{F(W|Z)}\{f(W)\} = \mathbb{E}\{f(W) \mid Z\} = 0$ for almost all Z implies that $f(W) = 0$ almost surely. This is equivalent to our definition given in equation (A.6).

Recently, completeness conditions have been extensively applied in the econometrics literature to obtain identification for a variety of nonparametric and semi-parametric models, most famously, in nonparametric models with instrumental variables (e.g., Ai and Chen, 2003; Newey and Powell, 2003; Chernozhukov *et al.*, 2007; Darolles *et al.*, 2011). Other examples include measurement error models (e.g., Hu and Schennach, 2008) and panel or dynamic models (e.g., Hu and Shum, 2012; Freyberger, 2018).

Finally, as in our paper, completeness conditions have been essential in the literature of negative controls and proximal causal learning (Tchetgen Tchetgen *et al.*, 2020a). Miao *et al.* (2018b) make two completeness conditions (a) the completeness of W with respect to Z, (b) the completeness of Z with respect to U (see Conditions 2 and 3 in their paper). Deaner (2018); Shi *et al.* (2020); Kallus *et al.* (2021) make alternative two completeness conditions (a) the completeness of W with respect to U, (b) the completeness of Z with respect to U (see

Assumption 3 in Deaner (2018), Assumption 4 in Shi *et al.* (2020), and Example 6 in Kallus *et al.* (2021)). Miao *et al.* (2018a) make one completeness condition (the completeness of Z with respect to W; see Assumption 5 in their paper) along with the assumption of the existence of an outcome confounding bridge function, which can be justified by another completeness condition (the completeness of W with respect to U).

In Sections 2 and 3, we followed Miao *et al.* (2018a) and made Assumptions 1.3 and 1.4 and Assumptions 2.3 and 2.4, respectively. We prove nonparametric identification of the ACPE under those assumptions in Appendix A.5 below. We briefly note however that analogous developments can be established under one of the aforementioned alternative completeness conditions in the literature. This is further explored below in Appendix A.6.

A.3 Proof of Lemma 1

First, equation (10) implies that

$$\begin{aligned} & C_i \perp\!\!\!\perp (\{C_j : j \neq i\}, A_i) \ \mid \ U_i, X_i, && \text{(A.7)} \\ \Longrightarrow \quad & W_i \perp\!\!\!\perp A_i \ \mid \ U_i, X_i, && \text{(A.8)} \end{aligned}$$

as we define $W_i = C_i$ and $A_i = \phi(\{Y_{j1} : j \in \mathcal{N}(i;1)\}) \in \mathbb{R}$. Then, equation (A.7) also implies that

$$\begin{aligned} & C_i \perp\!\!\!\perp \{C_j : j \neq i\} \ \mid \ A_i, U_i, X_i, \\ \Longrightarrow \quad & W_i \perp\!\!\!\perp Z_i \ \mid \ A_i, U_i, X_i, && \text{(A.9)} \end{aligned}$$

as we define $W_i = C_i$ and $Z_i = \{C_j : j \neq i\}$. Finally, equation (11) implies that

$$\begin{aligned} & Y_{i2} \perp\!\!\!\perp \{C_j : j \neq i\} \ \mid \ A_i, U_i, X_i, \\ \Longrightarrow \quad & Y_{i2} \perp\!\!\!\perp Z_i \ \mid \ A_i, U_i, X_i, && \text{(A.10)} \end{aligned}$$

where $Z_i = \{C_j : j \neq i\}$. Therefore, equations (A.8)–(A.10) are equivalent to Assumption 2.2, which completes the proof. $\square$

A.4 Proof of Lemma 2

First, equation (12) implies that

$$Y_{i1} \perp\!\!\!\perp (A_i, \{Y_{j1} : j \in \mathcal{N}(i;s), s \geq 2\}) \ \mid \ U_i, X_i, \tag{A.11}$$

$$\Longrightarrow \quad W_i \perp\!\!\!\perp A_i \mid U_i, X_i, \tag{A.12}$$

as we define $W_i = Y_{i1}$. Then, equation (A.11) also implies that

$$\begin{aligned} & Y_{i1} \perp\!\!\!\perp \{Y_{j1} : j \in \mathcal{N}(i;s), s \geq 2\} \mid A_i, U_i, X_i, \\ \Longrightarrow \quad & W_i \perp\!\!\!\perp Z_i \mid A_i, U_i, X_i, \end{aligned} \tag{A.13}$$

as we define $W_i = Y_{i1}$ and $Z_i = \{Y_{j1} : j \in \mathcal{N}(i;s), s \geq \tilde{s}\}$ where $\tilde{s} \geq 2$.

Finally, equation (13) states that

$$\begin{aligned} & Y_{i2} \perp\!\!\!\perp \{Y_{j1} : j \in \mathcal{N}(i;s), s \geq \tilde{s}\} \mid A_i, U_i, X_i, \\ \Longrightarrow \quad & Y_{i2} \perp\!\!\!\perp Z_i \mid A_i, U_i, X_i, \end{aligned} \tag{A.14}$$

where $Z_i = \{Y_{j1} : j \in \mathcal{N}(i;s), s \geq \tilde{s}\}$. Therefore, equations (A.12)–(A.14) are equivalent to Assumption 2.2, which completes the proof. □

A.5 Proof of Theorem 2

Here, we prove identification of $\mathbb{E}\{Y_{i2}(a)\}$ for $a \in \mathcal{A}$ and a given unit $i \in N_n$, which is sufficient for proving identification of the ACPE. The proof of Theorem 1 is a special case of the proof we provide below.

This proof adopts the proof by Miao *et al.* (2018a) to our network setting. First, we prove that the mean potential outcomes can be identified as the mean of the outcome confounding bridge function.

$$\mathbb{E}\{Y_{i2}(a)\} = \mathbb{E}\{h(W_i, a, X_i)\}.$$

Proof: Under Assumption 2.1,

$$\begin{aligned} \int \mathbb{E}(Y_{i2} \mid A_i = a, U_i = u, X_i = x) dF(U_i = u, X_i = x) &= \int \mathbb{E}\{Y_{i2}(a) \mid U_i = u, X_i = x\} dF(U_i = u, X_i = x) \\ &= \mathbb{E}\{Y_{i2}(a)\}. \end{aligned}$$

Under Assumption 2.2,

$$\begin{aligned} & \int \mathbb{E}\{h(W_i, a, X_i) \mid A_i = a, U_i = u, X_i = x) dF(U_i = u, X_i = x) \\ = \ & \int \mathbb{E}\{h(W_i, a, X_i) \mid U_i = u, X_i = x) dF(U_i = u, X_i = x) \\ = \ & \mathbb{E}\{h(W_i, a, X_i)\}. \end{aligned}$$

Under Assumption 2.3, $\mathbb{E}(Y_{i2} \mid A_i = a, U_i = u, X_i = x) = \mathbb{E}\{h(W_i, a, X_i) \mid A_i = a, U_i = u, X_i = x\}$, and therefore,

$$\mathbb{E}\{Y_{i2}(a)\} = \mathbb{E}\{h(W_i, a, X_i)\},$$

which completes the proof. $\square$

Next, we prove that the confounding bridge function is identified as follows.

$$\mathbb{E}(Y_{i2} \mid Z_i, A_i, X_i) = \mathbb{E}\{h(W_i, A_i, X_i) \mid Z_i, A_i, X_i\}. \tag{A.15}$$

Proof: Under Assumption 2.2,

$$\begin{aligned}
& \int \mathbb{E}(Y_{i2} \mid A_i, U_i = u, X_i) dF(U_i = u \mid Z_i, A_i, X_i) \\
= & \int \mathbb{E}(Y_{i2} \mid Z_i, A_i, U_i = u, X_i) dF(U_i = u \mid Z_i, A_i, X_i) \\
= & \ \mathbb{E}(Y_{i2} \mid Z_i, A_i, X_i).
\end{aligned}$$

Under Assumption 2.2,

$$\begin{aligned}
& \int \mathbb{E}\{h(W_i, A_i, X_i) \mid A_i, U_i = u, X_i\} dF(U_i = u \mid Z_i, A_i, X_i) \\
= & \int \mathbb{E}\{h(W_i, A_i, X_i) \mid Z_i, A_i, U_i = u, X_i\} dF(U_i = u \mid Z_i, A_i, X_i) \\
= & \ \mathbb{E}\{h(W_i, A_i, X_i) \mid Z_i, A_i, X_i\}.
\end{aligned}$$

Under Assumption 2.3, $\mathbb{E}(Y_{i2} \mid A_i, U_i, X_i) = \mathbb{E}\{h(W_i, A_i, X_i) \mid A_i, U_i, X_i\}$, and therefore,

$$\mathbb{E}(Y_{i2} \mid Z_i, A_i, X_i) = \mathbb{E}\{h(W_i, A_i, X_i) \mid Z_i, A_i, X_i\}.$$

We finally demonstrate that the solution to equation (A.15) is unique and identifies the outcome confounding bridge function h under Assumption 2.4. Suppose there are two functions $h(W_i, A_i, X_i)$ and $h'(W_i, A_i, X_i)$ that satisfy equation (A.15). Then,

$$\mathbb{E}\{h(W_i, A_i, X_i) - h'(W_i, A_i, X_i) \mid Z_i = z, A_i = a, X_i = x\} = 0$$

for all a, x, and almost all z. Then, under Assumption 2.4, $h(W_i, A_i, X_i) = h'(W_i, A_i, X_i)$ almost surely. Thus, the solution to equation (A.15) identifies the outcome confounding bridge function. $\square$

A.6 Identification of the ACPE under Alternative Assumptions

Here, we show that the same identification formula for the ACPE can be proven based on an alternative set of assumptions. The main difference is that we first define an outcome bridge function as a solution to the following Fredholm integral equation of the first kind.

$$\mathbb{E}(Y_{i2} \mid Z_i, A_i, X_i) = \mathbb{E}\{h^\dagger(W_i, A_i, X_i) \mid Z_i, A_i, X_i\}. \tag{A.16}$$

Then, we show, under some assumptions, this outcome bridge function satisfies

$$\mathbb{E}(Y_{i2} \mid A_i = a, U_i, X_i) = \mathbb{E}\{h^\dagger(W_i, a, X_i) \mid A_i = a, U_i, X_i\}. \tag{A.17}$$

This approach is in contrast with the approach we used in the main paper and proved in Appendix A.5 where we defined an outcome bridge function as a solution to equation (A.17) and then showed that it satisfies equation (A.16). The approach used in this section is similar to the one used in Deaner (2018); Miao *et al.* (2018b); Shi *et al.* (2020).

We see below that although this difference leads to a different set of assumptions, they both result in the same identification formula for the ACPE.

In particular, while we maintain Assumption 2.1 and Assumption 2.2, we replace Assumption 2.3 and Assumption 2.4 with two different assumptions below (Assumptions 7 and 8).

Assumption 7 (Outcome Confounding Bridge $h^\dagger$) *There exists some function $h^\dagger(W_i, A_i, X_i)$ such that for all $a \in \mathcal{A}$, and all $i \in N_n$,*

$$\mathbb{E}(Y_{i2} \mid A_i = a, Z_i = z, X_i) = \mathbb{E}\{h^\dagger(W_i, a, X_i) \mid A_i = a, Z_i = z, X_i\}. \tag{A.18}$$

Assumption 8 (Relevance of Z for U) *For any square integrable function f and for any a and x, if $\mathbb{E}\{f(U_i) \mid Z_i = z, A_i = a, X_i = x\} = 0$ for almost all z, then $f(U_i) = 0$ almost surely.*

Theorem 5 *Under Assumptions 2.1, 2.2, 7 and 8, an outcome confounding bridge function $h^\dagger$ (defined in equation (A.18)) satisfies the following equality for all $a \in \mathcal{A}$, and all $i \in N_n$,*

$$\mathbb{E}(Y_{i2} \mid U_i, A_i = a, X_i) = \mathbb{E}\{h^\dagger(W_i, a, X_i) \mid U_i, A_i = a, X_i\},$$

and, the ACPE is identified by

$$\tau(a, a') = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\{h^\dagger(W_i, a, X_i) - h^\dagger(W_i, a', X_i)\}.$$

Finally, like Lemma 4, we can also prove Assumption 7 under a completeness condition and associated regularity conditions (Assumption 9 defined below in Appendix A.6.2).

Lemma 6 *Under Assumptions 5 and 9, there exists a function $h^\dagger(W_i, a, X_i)$ such that for all $a \in \mathcal{A}$ and all $i \in N_n$, equation* (A.18) *holds.*

A.6.1 Proof of Theorem 5

First, we show that an outcome confounding bridge function $h^\dagger$ defined in equation (A.18) satisfies the following equality.

$$\mathbb{E}\{h^\dagger(W_i, a, X_i) \mid A_i = a, U_i = u, X_i = x) = \mathbb{E}\{Y_{i2}(a) \mid A_i = a, U_i = u, X_i = x\} \tag{A.19}$$

We have

$$\begin{aligned}
&\mathbb{E}\{h^\dagger(W_i, a, X_i) \mid A_i = a, Z_i = z, X_i = x) \\
= &\int \mathbb{E}\{h^\dagger(W_i, a, X_i) \mid A = a, U_i = u, Z_i = z, X_i = x) dF(U_i = u \mid A = a, Z_i = z, X_i = x) \\
= &\int \mathbb{E}\{h^\dagger(W_i, a, X_i) \mid A = a, U_i = u, X_i = x) dF(U_i = u \mid A = a, Z_i = z, X_i = x) \qquad \text{(A.20)}
\end{aligned}$$

where the first equality follows from iterated expectations, and the second from Assumption 2.2. We also have

$$\begin{aligned}
&\mathbb{E}(Y_{i2} \mid A_i = a, Z_i = z, X_i = x) \\
= &\int \mathbb{E}(Y_{i2} \mid A_i = a, U_i = u, Z_i = z, X_i = x) dF(U_i = u \mid A_i = a, Z_i = z, X_i = x) \\
= &\int \mathbb{E}(Y_{i2} \mid A_i = a, U_i = u, X_i = x) dF(U_i = u \mid A_i = a, Z_i = z, X_i = x) \qquad \text{(A.21)}
\end{aligned}$$

where the first equality follows from iterated expectations, and the second from Assumption 2.2.

Under Assumption 7, we have

$$\begin{aligned}
&\mathbb{E}(Y_{i2} \mid A_i = a, Z_i = z, X_i) = \mathbb{E}\{h^\dagger(W_i, a, X_i) \mid A_i = a, Z_i = z, X_i\} \\
\iff &\int \{\mathbb{E}(Y_{i2} \mid A_i = a, U_i = u, X_i = x) - \mathbb{E}\{h^\dagger(W_i, a, X_i) \mid A_i = a, U_i = u, X_i = x)\} \\
&\qquad \times dF(U_i = u \mid A_i = a, Z_i = z, X_i = x) = 0 \\
\implies &\mathbb{E}(Y_{i2}(a) \mid A_i = a, U_i = u, X_i = x) = \mathbb{E}\{h^\dagger(W_i, a, X_i) \mid A_i = a, U_i = u, X_i = x\}
\end{aligned}$$

where the first equivalence comes from equations (A.20) and (A.21), and the final line follows from Assumption 8 and the consistency of the potential outcomes.

Next, by using equation (A.19), we prove that

$$\mathbb{E}\{Y_{i2}(a)\} = \mathbb{E}\{h^{\dagger}(W_i, a, X_i)\}.$$

Under Assumption 2.1, we have

$$\begin{aligned}
& \int \mathbb{E}(Y_{i2}(a) \mid A_i = a, U_i = u, X_i = x) dF(U_i = u, X_i = x) \\
= & \int \mathbb{E}\{Y_{i2}(a) \mid U_i = u, X_i = x\} dF(U_i = u, X_i = x) \\
= & \ \mathbb{E}\{Y_{i2}(a)\}.
\end{aligned}$$

Under Assumption 2.2, we have

$$\begin{aligned}
& \int \mathbb{E}\{h^{\dagger}(W_i, a, X_i) \mid A_i = a, U_i = u, X_i = x) dF(U_i = u, X_i = x) \\
= & \int \mathbb{E}\{h^{\dagger}(W_i, a, X_i) \mid U_i = u, X_i = x) dF(U_i = u, X_i = x) \\
= & \ \mathbb{E}\{h^{\dagger}(W_i, a, X_i)\}.
\end{aligned}$$

Equation (A.19) states that $\mathbb{E}(Y_{i2}(a) \mid A_i = a, U_i = u, X_i = x) = \mathbb{E}\{h^{\dagger}(W_i, a, X_i) \mid A_i = a, U_i = u, X_i = x\}$, and therefore,

$$\mathbb{E}\{Y_{i2}(a)\} = \mathbb{E}\{h^{\dagger}(W_i, a, X_i)\},$$

which completes the proof. □

A.6.2 Proof of Lemma 6

In this proof, as done above, we use Y to denote the outcome and A to denote the treatment instead of (Y_{i2}, A_i), which we use in Section 3. To provide rigorous discussion on the existence of a solution to a Fredholm integral equation of the first kind, we keep using notation introduced in Appendix A.1.

Using general notation, we re-state Lemma 6 as follows. Under Assumptions 5 and 9, there exists a function $h(W, a, X)$ such that for all $a \in \mathcal{A}$, a solution to the following Fredholm integral equation of the first kind exists.

$$\mathbb{E}(Y \mid Z = z, A = a, X = x) = \mathbb{E}\{h^{\dagger}(W, a, x) \mid Z = z, A = a, X = x\}. \qquad \text{(A.22)}$$

We also introduce regularity conditions related to the singular value decomposition.

Assumption 9 (Regularity Conditions II)

$$\int_{-\infty}^{+\infty} dF(z \mid w,a,x)dF(w \mid z,a,x)dwdz \;<\; +\infty \tag{A.23}$$

$$\int_{-\infty}^{+\infty} \mathbb{E}(Y \mid a,z,x)^2 dF(z \mid a,x)dz \;<\; +\infty \tag{A.24}$$

$$\sum_{p=1}^{+\infty} \frac{1}{\tilde{\nu}_{a,x,p}^2} |\langle \mathbb{E}(Y \mid a,z,x), \tilde{\kappa}_{a,x,p}\rangle|^2 \;<\; +\infty \tag{A.25}$$

where $\tilde{\nu}_{a,x,p}$ is the p-th singular value of $\widetilde{K}_{a,x}$, and $\tilde{\kappa}_{a,x,p} \in L^2\{F(z \mid a,x)\}$ is an orthogonal sequence.

Proof: We start by defining a kernel

$$\widetilde{K}(w,z,a,x) = \frac{dF(w,z \mid a,x)}{dF(w \mid a,x)dF(z \mid a,x)}.$$

We then define the linear operators $\widetilde{K}_{a,x} : L^2\{F(w \mid a,x)\} \to L^2\{F(z \mid a,x)\}$ by

$$\widetilde{K}_{a,x}h = \int_{-\infty}^{+\infty} \widetilde{K}(w,z,a,x)h(w)dF(w \mid a,x) = \mathbb{E}\{h(w) \mid z,a,x\}$$

for $h \in L^2\{F(w \mid a,x)\}$.

The adjoint of this linear operator $\widetilde{K}^*_{a,x} : L^2\{F(z \mid a,x)\} \to L^2\{F(w \mid a,x)\}$ is given by

$$\widetilde{K}^*_{a,x}\tilde{h} = \int_{-\infty}^{+\infty} \widetilde{K}(w,z,a,x)\tilde{h}(z)dF(z \mid a,x) = \mathbb{E}\{\tilde{h}(z) \mid w,a,x\}$$

for $\tilde{h} \in L^2\{F(z \mid a,x)\}$.

Using the introduced notations, we can re-write equation (A.22) as follows using the notations introduced above.

$$\widetilde{K}_{a,x}h = \mathbb{E}(Y \mid A=a, Z=z, X=x). \tag{A.26}$$

Therefore, to use Picard's theorem (Lemma 3), we need to prove (i) $\widetilde{K}_{a,x}$ is a compact operator, (ii) $\mathbb{E}(Y \mid A=a, Z=z, X=x) \in L^2\{F(z \mid a,x)\}$, (iii) $\mathbb{E}(Y \mid A=a, Z=z, X=x) \in \mathrm{Null}(\widetilde{K}^*_{a,x})^\perp$, and (iv) $\sum_{p=1}^{+\infty} \frac{1}{\widetilde{\nu}_{a,x,p}^2} |\langle \mathbb{E}(Y \mid A=a, Z=z, X=x), \widetilde{\kappa}_{a,x,p}\rangle|^2 \;<\; +\infty$, where $\widetilde{\nu}_{a,x,p}$ is the p-th singular value of $\widetilde{K}_{a,x}$, and $\widetilde{\kappa}_{a,x,p} \in L^2\{F(z \mid a,x)\}$ is an orthogonal sequence.

Proof of (i): We note that $\widetilde{K}_{a,x}$ and $\widetilde{K}^*_{a,x}$ are compact operators under equation (A.23) (Carrasco *et al.*, 2007, Example 2.3 on page 5659). Therefore, there exists a singular system $(\widetilde{\nu}_{a,x,p}, \widetilde{\upsilon}_{a,x,p}, \widetilde{\kappa}_{a,x,p})$ of $\widetilde{K}_{a,x}$ according to Kress (1989, Theorem 15.16) where $\widetilde{\nu}_{a,x,p}$ is the p-th singular value of $\widetilde{K}_{a,x}$, and $\widetilde{\upsilon}_{a,x,p} \in L^2\{F(w \mid a,x)\}$ and $\widetilde{\kappa}_{a,x,p} \in L^2\{F(z \mid a,x)\}$ are orthogonal sequences.

Proof of (ii): Under equation (A.24), we have $\mathbb{E}(Y \mid a,z,x) \in L^2\{F(z \mid a,x)\}$.

Proof of (iii): To show that $\mathbb{E}(Y \mid A=a, Z=z, X=x) \in \text{Null}(\widetilde{K}^*_{a,x})^{\perp}$, we first define $\tilde{h} \in \text{Null}(\widetilde{K}^*_{a,x})$. Then, we show below that, for any $\tilde{h} \in \text{Null}(\widetilde{K}^*_{a,x})$,

$$\langle \mathbb{E}(Y \mid A=a, Z=z, X=x), \tilde{h}(z) \rangle = 0.$$

We begin by showing $\mathbb{E}\{\tilde{h}(z) \mid U,a,x\} = 0$. First, we have

$$\begin{aligned} \mathbb{E}\{\tilde{h}(z) \mid w,a,x\} &= \mathbb{E}\{\mathbb{E}\{\tilde{h}(z) \mid U,w,a,x\} \mid w,a,x\} \\ &= \mathbb{E}\{\mathbb{E}\{\tilde{h}(z) \mid U,a,x\} \mid w,a,x\} \end{aligned} \tag{A.27}$$

where the first equality follows from iterated expectations, and the second from Assumption 2.2. By definition of the null space, we have $\widetilde{K}^*_{a,x}\tilde{h} = \mathbb{E}\{\tilde{h}(z) \mid w,a,x\} = 0$ almost surely. Therefore,

$$\begin{aligned} \mathbb{E}\{\tilde{h}(z) \mid w,a,x\} = 0 &\iff \mathbb{E}\{\mathbb{E}\{\tilde{h}(z) \mid U,a,x\} \mid w,a,x\} = 0 \\ &\implies \mathbb{E}\{\tilde{h}(z) \mid U,a,x\} = 0 \quad \text{almost surely.} \end{aligned} \tag{A.28}$$

where the first equivalence follows from equation (A.27), and the second line follows from the relevance of W for U (Assumption 5). Finally, we now show $\langle \mathbb{E}(Y \mid A=a, Z=z, X=x), \tilde{h}(z) \rangle = 0$.

$$\begin{aligned} & \langle \mathbb{E}(Y \mid A=a, Z=z, X=x), \tilde{h}(z) \rangle \\ := \; & \mathbb{E}\{\mathbb{E}(Y \mid A=a, Z=z, X=x)\tilde{h}(z) \mid A=a, X=x\} \\ = \; & \mathbb{E}\{\mathbb{E}\{\mathbb{E}(Y \mid U, A=a, Z=z, X=x) \mid A=a, Z=z, X=x\}\tilde{h}(z) \mid A=a, X=x\} \\ = \; & \mathbb{E}\{\mathbb{E}\{\mathbb{E}(Y \mid U, A=a, X=x) \mid A=a, Z=z, X=x\}\tilde{h}(z) \mid A=a, X=x\} \\ = \; & \mathbb{E}\{\mathbb{E}\{\mathbb{E}\{\mathbb{E}(Y \mid U, A=a, X=x) \mid A=a, Z=z, X=x\}\tilde{h}(z) \mid U, A=a, X=x\} \mid A=a, X=x\} \\ = \; & \mathbb{E}\{\mathbb{E}(Y \mid U, A=a, X=x)\mathbb{E}\{\tilde{h}(z) \mid U, A=a, X=x\} \mid A=a, X=x\} \\ = \; & 0, \end{aligned}$$

where the first line follows from the definition of the inner product in a Hilbert space, the second from iterated expectations applied to $\mathbb{E}(Y \mid A=a, Z=z, X=x)$, the third from Assumption 2.2, the fourth from iterated expectations, the fifth from conditioning on $(U, A =$

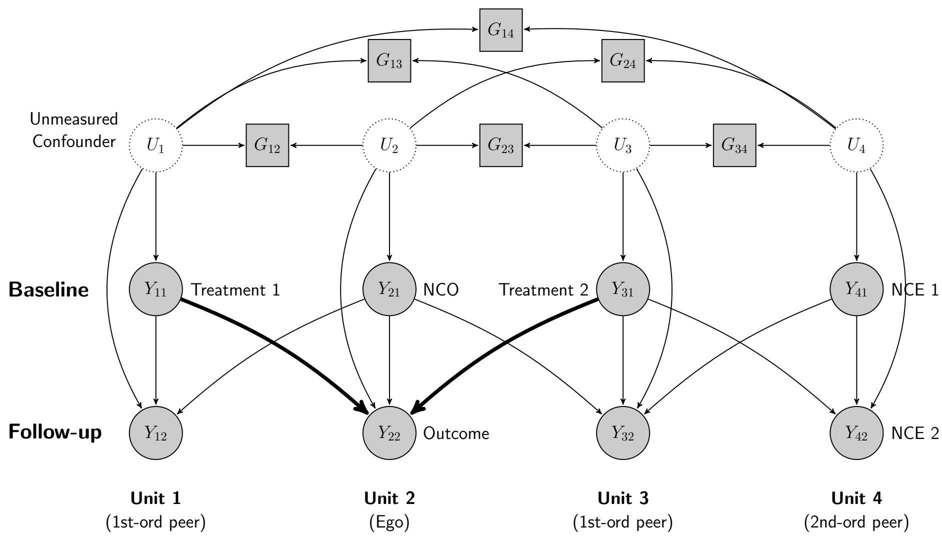

Figure A1: Example of DAGs with Unobserved Homophily. *Note:* For concreteness, we show Unit 2 as the ego. The thick arrows from Y_{11} to Y_{22} and from Y_{31} to Y_{22} indicate the causal peer effects of interest. We use shaded (dotted) nodes to denote observed (unobserved) variables. For simplicity, we suppressed observed covariates X. Here, the square box around G represents that we observe variables conditional on network $\mathcal{G}$.

$a, X = x)$, and finally, the sixth follows from equation (A.28). Therefore, this shows that $\mathbb{E}(Y \mid A = a, Z = z, X = x) \in \text{Null}(\widetilde{K}^*_{a,x})^{\perp}$.

Proof of (iv): Finally, this key condition for Picard's theorem is directly implied by equation (A.25), which completes the proof. $\square$

A.7 Using Focal Behaviors as Negative Controls in Network Settings

In certain settings, we may also use focal behaviors Y_{it} of peers and those measured at baseline as plausible candidates for negative controls. In particular, ego's focal behavior at baseline may serve as valid NCO and focal behaviors of peers-of-peers $\{Y_{jt} : j \in \mathcal{N}(i; 2), t \in \{1, 2\}\}$ may constitute valid NCEs. Figure A1 represents a causal graph illustrating an instance of the causal model where Y_{21} qualifies as NCO and variables $\{Y_{41}, Y_{42}\}$ qualify as NCE. More generally, when focal behaviors of peers-of-peers constitute valid negative control exposures, focal behaviors of units at least of network distance 2 from node i may be credible negative control exposures. A hybrid approach might entail combining the auxiliary variables and focal behaviors as negative controls. In Figure 3 of the main paper, we define focal behavior measured at baseline Y_{21} as NCO (instead of C_2) and auxiliary variables of peers $\{C_1, C_3, C_4\}$ as NCEs.

This selection of negative controls is particularly plausible when focal behaviors do not have direct causal relationships with peers' focal behaviors measured concurrently. This absence of causal simultaneity has previously been assumed in the literature of causal peer effects (Shalizi and Thomas, 2011; Ogburn and VanderWeele, 2014; Egami, 2018; Liu and Tchetgen Tchetgen, 2020; McFowland III and Shalizi, 2021).

In practice, this assumption is most credible when researchers a priori know that the focal behaviors of units are indeed measured concurrently. For example, in Add Health data, students' GPA are likely to be measured at the same time for students within a school, and thus, a student's GPA cannot be affected by peers' GPA in the same semester. Importantly, a student's GPA can be affected by peers' GPA in the last semester, and a student's study habit might be affected by peer's study habits within the same semester. These, however, do not invalidate the use of GPA of peers-of-peers as NCEs as long as students' GPA within the same semester do not have direct causal relationships with each other. In some applications, analysts can directly measure focal behaviors of interest. In such cases, by virtue of survey/study design, researchers can ensure that the focal behaviors measured at each wave do not affect peers' focal behaviors within the same wave by conducting surveys concurrently. This assumption is less credible when measurements of focal behaviors are aggregated over long periods of time, such as the number of political tweets over a year, which is likely to be affected by peers' tweets within the same year. This is often called the temporal aggregation problem, which invalidates not only peer effect analysis but also a large class of panel data analyses (Granger, 1988).

This particular selection strategy of negative controls has two advantages when valid. First, when the NCO entails focal behaviors measured at baseline, the confounding bridge assumption (Assumption 2.3) is often more likely to hold because the NCO and the main outcome are measured on the same scale (Sofer *et al.*, 2016). Second, if researchers can leverage focal behaviors of peers and those measured at baseline as negative controls, researchers do not need to collect additional auxiliary variables, which lowers data collection requirements and improves applicability of the double negative control approach. However, selection of valid negative control variables must always be based on reliable domain knowledge because Assumption 2.1 – Assumption 2.4 must be met.

A.8 Proof: Identification under Linear Confounding Bridge

We offer an example of a linear confounding bridge with binary treatment and NCE in Section 2.3. Here, we provide a proof of the exact expression. For notational simplicity, we remove conditioning on $S = 1$ in this section. Because this part of discussions is general about the double negative control approach, we use A to denote the binary treatment and Y to denote the outcome variable.

We assume a linear confounding bridge, $h(W, A; \gamma) = \gamma_0 + \gamma_1 W + \gamma_2 A$. In this case, the ACPE is equal to γ_2. From Theorem 1, we have the following equality under Assumptions 1.1-1.4.

$$\mathbb{E}(Y \mid Z, A) = \mathbb{E}\{h(W, A) \mid Z, A\}.$$

Conditioning on unmeasured confounder U shows up explicitly in the proof of Theorem 1. Please see Appendix A.5 where we prove Theorem 1 as a special case of Theorem 2.

Using a linear confounding bridge, the right term is

$$\mathbb{E}\{h(W, A) \mid Z, A\} = \gamma_0 + \gamma_1 \times \mathbb{E}(W \mid Z, A) + \gamma_2 A.$$

Therefore, from Theorem 1, we have

$$\mathbb{E}(Y \mid Z, A) = \gamma_0 + \gamma_1 \times \mathbb{E}(W \mid Z, A) + \gamma_2 A \tag{A.29}$$

Using equation (A.29), we have

$$\begin{aligned}
&\mathbb{E}\{\mathbb{E}(Y \mid Z = 1, A) - \mathbb{E}(Y \mid Z = 0, A)\} = \gamma_1 \times \mathbb{E}\{\mathbb{E}(W \mid Z = 1, A) - \mathbb{E}(W \mid Z = 0, A)\} \\
\Longrightarrow \quad & \gamma_1 = \frac{\mathbb{E}\{\mathbb{E}(Y \mid Z = 1, A) - \mathbb{E}(Y \mid Z = 0, A)\}}{\mathbb{E}\{\mathbb{E}(W \mid Z = 1, A) - \mathbb{E}(W \mid Z = 0, A)\}}
\end{aligned}$$

Again from equation (A.29), we have

$$\begin{aligned}
& \mathbb{E}\{\mathbb{E}(Y \mid Z, A = 1) - \mathbb{E}(Y \mid Z, A = 0)\} = \gamma_2 + \gamma_1 \times \mathbb{E}\{\mathbb{E}(W \mid Z, A = 1) - \mathbb{E}(W \mid Z, A = 0)\} \\
\Longrightarrow \quad \gamma_2 &= \mathbb{E}\{\mathbb{E}(Y \mid Z, A = 1) - \mathbb{E}(Y \mid Z, A = 0)\} - \gamma_1 \times \mathbb{E}\{\mathbb{E}(W \mid Z, A = 1) - \mathbb{E}(W \mid Z, A = 0)\} \\
\Longrightarrow \quad \gamma_2 &= \mathbb{E}\{\mathbb{E}(Y \mid Z, A = 1) - \mathbb{E}(Y \mid Z, A = 0)\} \\
& \quad - \mathbb{E}\{\mathbb{E}(W \mid Z, A = 1) - \mathbb{E}(W \mid Z, A = 0)\} \times \frac{\mathbb{E}\{\mathbb{E}(Y \mid Z = 1, A) - \mathbb{E}(Y \mid Z = 0, A)\}}{\mathbb{E}\{\mathbb{E}(W \mid Z = 1, A) - \mathbb{E}(W \mid Z = 0, A)\}},
\end{aligned}$$

which completes the proof. $\square$

B DNC Estimator for Dyadic Data

We now propose a strategy for estimation and inference of the ACPE. Because we observe n independent and identically distributed samples of dyads, we observe independent and identically distributed samples on $(Y_{12}, Y_{21}, W, Z, X)$ given $S = 1$ where Y_{12}, Y_{21}, W, Z, X are the outcome of interest, treatment, NCO, NCE, and observed pre-treatment covariates, respectively.

Suppose that an analyst has specified a parametric or semiparametric model for the confounding bridge $h(W, Y_{21}, X; \gamma)$ with parameter γ. Then, based on Theorem 1, we can estimate γ by solving the following empirical moment equations.

$$\frac{1}{n}\sum_{i=1}^{n}\{Y_{i12} - h(W_i, Y_{i21}, X_i; \gamma)\} \times \eta(Z_i, Y_{i21}, X_i) = 0,$$

where η is a user-specified vector function with dimension equal to that of γ. For example, if a linear confounding bridge function is used, i.e., $h(W, Y_{21}, X; \gamma) = (1, W, Y_{21}, X)^\top \gamma$, we can use $\eta(Z, Y_{21}, X) = (1, Z, Y_{21}, X)^\top$.

Once the bridge function h is estimated, we can estimate the ACPE by

$$\frac{1}{n}\sum_{i=1}^{n}\{h(W_i, y_{21}, X_i; \widehat{\gamma}) - h(W_i, y'_{21}, X_i; \widehat{\gamma})\}.$$

To appropriately account for uncertainty of the estimated bridge function and for the possibility that dimension of η might be larger than that of γ, we combine the two moments into generalized method of moments (GMM) with parameter $\theta = (\tau, \gamma)$ (Hansen, 1982; Miao *et al.*, 2018a). We define a moment for dyad i to be

$$m(Y_{i12}, Y_{i21}, W_i, Z_i, X_i; \theta) = \left\{\begin{array}{c} \tau - \{h(W_i, y_{21}, X_i; \gamma) - h(W_i, y'_{21}, X_i; \gamma)\} \\ \{Y_{i12} - h(W_i, Y_{i21}, X_i; \gamma)\} \times \eta(Z_i, Y_{i21}, X_i) \end{array}\right\}.$$

Then, the GMM estimator is

$$\widehat{\theta} = \operatorname*{argmin}_{\theta} \overline{m}(\theta)^\top \Omega\ \overline{m}(\theta) \tag{A.30}$$

where $\overline{m}(\theta) = \frac{1}{n}\sum_{i=1}^{n} m(Y_{i12}, Y_{i21}, W_i, Z_i, X_i; \theta)$ and Ω is a user-specified positive-definite weight matrix. Asymptotic properties described below hold for any positive-definite weight matrix Ω and for alternative GMM estimators, such as the two-step GMM and continuously updating GMM.

The proposed double negative control (DNC) estimator $\widehat{\tau}(y_{21}, y'_{21})$ for $\tau(y_{21}, y'_{21})$ is the first element of $\widehat{\theta}$ defined in equation (A.30). Because we consider i.i.d. samples of dyads in this section, the moment $m(Y_{i12}, Y_{i21}, W_i, Z_i, X_i; \theta)$ is also i.i.d., and thus, under the standard regularity conditions for GMM (Hansen, 1982; Newey and McFadden, 1994), the DNC estimator is consistent:

$$\widehat{\tau}(y_{21}, y'_{21}) \xrightarrow{p} \tau(y_{21}, y'_{21}),$$

and asymptotically normal:

$$\frac{\widehat{\tau}(y_{21}, y'_{21}) - \tau(y_{21}, y'_{21})}{\sqrt{\sigma^2/n}} \xrightarrow{d} \mathsf{Normal}(0, 1),$$

where $\xrightarrow{p}$ denotes convergence in probability, and $\xrightarrow{d}$ denotes convergence in distribution. Moreover, the asymptotic variance σ^2 can be consistently estimated by $\widehat{\sigma}^2 = (\widehat{\Gamma}\widehat{\Lambda}\widehat{\Gamma}^\top)_{11}$, which is the $(1,1)$ th element of matrix $\widehat{\Gamma}\widehat{\Lambda}\widehat{\Gamma}^\top$, and

$$\begin{aligned}
\widehat{\Lambda} &= \frac{1}{n}\sum_{i=1}^{n} m(Y_{i12}, Y_{i21}, W_i, Z_i, X_i; \widehat{\theta})\ m(Y_{i12}, Y_{i21}, W_i, Z_i, X_i; \widehat{\theta})^\top, \\
\widehat{\Gamma} &= (\widehat{M}^\top \Omega \widehat{M})^{-1}\widehat{M}^\top \Omega, \quad \text{and} \quad \widehat{M} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \theta} m(Y_{i12}, Y_{i21}, W_i, Z_i, X_i; \widehat{\theta}).
\end{aligned}$$

Therefore, an asymptotically valid $(1-\alpha)$ confidence interval for $\tau(y_{21}, y'_{21})$ is given by $[\widehat{\tau}(y_{21}, y'_{21}) - \Phi(1-\alpha/2) \times \widehat{\sigma}/\sqrt{n},\ \widehat{\tau}(y_{21}, y'_{21}) + \Phi(1-\alpha/2) \times \widehat{\sigma}/\sqrt{n}]$ where $\Phi(\cdot)$ denotes the quantile function for the standard normal distribution.

To minimize the asymptotic variance within the GMM class, we can use the two-step GMM to estimate the optimal $\widehat{\Omega}$. In the first step, we choose an identity matrix as Ω or some other positive-definite matrix, and compute preliminary GMM estimate $\widehat{\theta}_{(1)}$. This estimator is consistent, but not efficient. In the second step, we compute $\widehat{\Lambda}$ based on $\widehat{\theta}_{(1)}$, which is denoted by $\widehat{\Lambda}_{(1)}$. Then, we can get the final estimate by solving equation (A.30) with $\Omega = \widehat{\Lambda}_{(1)}^{-1}$. The resulting estimator $\widehat{\theta}$ is consistent and asymptotically normal, and is asymptotically efficient within the GMM class (Hansen, 1982). The asymptotic variance also simplifies to $(\widehat{M}^\top \widehat{\Lambda}^{-1} \widehat{M})^{-1}$. To further improve finite sample performance, researchers can consider alternative GMM estimators, such as continuously updating GMM (Hansen *et al.*, 1996).

C Sensitivity Analysis with A Linear Confounding Bridge

Building on sensitivity analysis proposed in Cobzaru *et al.* (2022), we propose a sensitivity analysis approach for the causal peer effect analysis with double negative controls. In particular, the proposed sensitivity analysis allows researchers to investigate the violation of the negative control relevance assumption. We extend the original sensitivity analysis in two ways: (1) we take into account network dependent errors, and (2) we allow for a continuous treatment, which is common in the causal peer analysis.

In this section, we relax the independent and identically distributed error assumption to allow for network-dependent (non-independent) errors, but we still keep the assumption of identical expectation of variables across units used in Cobzaru *et al.* (2022).

C.1 Setup

First, without loss of generality, we assume that unobserved and observed confounders have mean zero, and unobserved confounders have variances equal to one. We use ρ to denote covariance of U_i and X_i. For all i, we have

$$\mathbb{E}(U_i) = \mathbf{0}_{d_u}, \quad \mathrm{Var}(U_i) = \mathbf{I}_{d_u}$$

$$\mathbb{E}(X_i) = \mathbf{0}_{d_x}, \quad \mathrm{Var}(X_i) = \Sigma_X \quad \text{where } \Sigma_x \text{ is a } (d_x \times d_x) \text{ matrix.}$$

$$\mathrm{Cov}(U_i, X_i) = \rho \quad \text{where } \rho \text{ is a } (d_u \times d_x) \text{ matrix with each element between } -1 \text{ and } 1.$$

In this section, we use $d.$ to denote the dimension of random variable, e.g., d_u denotes the number of unobserved confounders U.

We consider cases where the number of NCE and that of NCO is the same, while it is smaller than the number of unobserved confounders U, i.e., $d_u > d_z = d_w$. This represents a scenario where the negative control relevance assumption is violated.

To simplify notations, we use Y to denote the outcome variable. Following Cobzaru *et al.* (2022), we consider the following linear data generating process for NCE Z_i, NCO W_i, and the potential outcome $Y_i(a)$. For all i, we have

$$Z_i = \mu_{0z} + \mu_{az} A_i + \mu_{uz}^\top U_i + \mu_{xz}^\top X_i + \epsilon_{iz}$$

$$W_i = \mu_{0w} + \mu_{uw}^\top U_i + \mu_{xw}^\top X_i + \epsilon_{iw}$$

$$Y(a)_i = \mu_{0y} + \mu_{ay} a + \mu_{uy}^\top U_i + \mu_{xy}^\top X_i + \epsilon_{iy}$$

where $(\mu_{0z}, \mu_{az}, \epsilon_{iz})$ are d_z dimensional vectors, μ_{uz} is a $(d_u \times d_z)$ matrix, and μ_{xz} is a $(d_x \times d_z)$ matrix. Similarly, $(\mu_{0w}, \epsilon_{iw})$ are d_w dimensional vectors, μ_{uw} is a $(d_u \times d_w)$ matrix, and μ_{xw} is a $(d_x \times d_w)$ matrix. Finally, $(\mu_{0y}, \mu_{ay}, \epsilon_{iy})$ are scalar, μ_{uy} is a d_u dimensional vector, and

μ_{xy} is a d_x dimensional vector. Here, following Cobzaru *et al.* (2022), we have $\epsilon_{iz} \perp\!\!\!\perp (A_i, U_i, X_i)$, $\epsilon_{iw} \perp\!\!\!\perp (U_i, X_i)$, and $\epsilon_{iy} \perp\!\!\!\perp (A_i, U_i, X_i)$. Importantly, in contrast to Cobzaru *et al.* (2022), we allow for network-dependent errors, e.g., $\epsilon_{iz} \not\perp\!\!\!\perp \epsilon_{jz}$, $\epsilon_{iw} \not\perp\!\!\!\perp \epsilon_{jw}$, and $\epsilon_{iy} \not\perp\!\!\!\perp \epsilon_{jy}$ where i and j are two different units.

C.2 Sensitivity Analysis

Suppose we use a linear confounding bridge, i.e., $h(W_i, A_i, X_i; \gamma) = \gamma_0 + \gamma_a A_i + \gamma_x^\top X_i + \gamma_w^\top W_i$, and we identify parameters with the following moment conditions.

$$\mathbb{E}\left\{\{Y_i - h(W_i, A_i, X_i; \gamma)\} \times \eta(A_i, Z_i, X_i)\right\} = 0$$

where $\eta(A_i, Z_i, X_i) = (1, A_i, Z_i, X_i)$. Under this setup, the true ACPE is μ_{ay} and the DNC estimator is γ_a. When the negative control relevance assumption is violated, the ACPE is not identified. In particular, we get

$$\gamma_a \quad = \quad \mu_{ay} + \underbrace{b^\top (\mathbf{I}_{d_u} - \mu_{uw}(B^\top \mu_{uw})^{-1} B^\top) \mu_{uy}}_{\text{bias}}, \tag{A.31}$$

where b is a d_u dimensional vector

$$b \quad = \quad \frac{\mathbb{E}(A_i U_i) - \rho \Sigma_x^{-1} \mathbb{E}(A_i X_i)}{\mathbb{E}(A_i^2) - \mathbb{E}(A_i)^2 - \mathbb{E}(A_i X_i)^\top \Sigma_x^{-1} \mathbb{E}(A_i X_i)},$$

and B is a $(d_u \times d_z)$ matrix

$$B = \left(\mathbf{I}_{d_u} - \rho \Sigma_x^{-1} \rho^\top - \frac{\left(\mathbb{E}(A_i U_i) - \rho \Sigma_x^{-1} \mathbb{E}(A_i X_i)\right)\left(\mathbb{E}(A_i U_i)^\top - \mathbb{E}(A_i X_i)^\top \Sigma_x^{-1} \rho^\top\right)}{\mathbb{E}(A_i^2) - \mathbb{E}(A_i)^2 - \mathbb{E}(A_i X_i)^\top \Sigma_x^{-1} \mathbb{E}(A_i X_i)}\right) \mu_{uz}.$$

We provide the proof in the next subsection (Appendix C.3).

It is worth emphasizing that this bias formula contains two special cases of no bias as well. First, this expression contains a special case of the conditional ignorability. When unobserved confounders U has no effect on the primary outcome, $\mu_{uy} = 0$ and thus, $\gamma_a = \mu_{ay}$ and there exists no bias. Second, this derivation also contains a special case when the negative control relevance assumption holds. When the negative control relevance assumption holds, e.g., $d_u = d_z$, $B^\top$ is invertible, and in this case, $\mu_{uw}\gamma_w = \mu_{uy}$, and therefore, $\mu_{uw}(B^\top \mu_{uw})^{-1} B^\top = \mathbf{I}_{d_u}$, and $\gamma_a = \mu_{ay}$, which means there exists no bias.

To use equation (A.31) for sensivity analysis, there are five unknown parameters (μ_{uw}, μ_{uz}, μ_{uy}, $\mathbb{E}(A_iU_i)$, $\mathbb{E}(U_iX_i)$) because b is a function of $\mathbb{E}(A_iU_i)$ and $\mathbb{E}(U_iX_i)$ and B is a function of $\mathbb{E}(A_iU_i)$, $\mathbb{E}(U_iX_i)$, and μ_{uz}.

One option for sensitivity analysis is for researchers to specify $(\mu_{uw}, \mu_{uz}, \mu_{uy}, \mathbb{E}(A_iU_i), \mathbb{E}(U_iX_i))$ and investigate how causal conclusions change depending on different values of these sensitivity parameters. This is the approach similar to the one proposed in Cobzaru *et al.* (2022). One disadvantage of this approach is that researchers have to specify a large number of parameters, i.e., $d_u \times d_w + d_u \times d_z + d_u + d_u + d_u \times d_x$.

Another option is to consider the upper and lower bound of the bias in order to simplify the choice of sensitivity parameters. Importantly, (μ_{uw}, μ_{uy}, μ_{uz}) represent the effects of unobserved confounders on NCO, NCE, and the primary outcome, conditional on observed confounders X and the treatment A, respectively. Note that, for NCO, $W_i \perp\!\!\!\perp A_i \mid U_i, X_i$, so it does not matter whether we condition on the treatment. To simplify sensitivity analysis, we can define $-\mu_u \leq \mu_{uw}, \mu_{uy}, \mu_{uz} \leq \mu_u$ where μ_u is scalar and $\mu_u > 0$.

Similarly, $\mathbb{E}(A_iU_i)$ and $\mathbb{E}(U_iX_i)$ represent covariances of unobserved confounders with the treatment and observed confounders. We then define $-\varsigma_u \leq \mathbb{E}(A_iU_i), \mathbb{E}(U_iX_i) \leq \varsigma_u$ where ς_u is scalar and $\varsigma_u > 0$.

Finally, we can obtain the bound of the bias by the following optimization problem.

$$\begin{aligned}
\text{bias}_{max}(\mu_u, \varsigma_u) \quad &:= \quad \max_{\mu_{uw}, \mu_{uy}, \mu_{uz}, \mathbb{E}(A_iU_i), \mathbb{E}(U_iX_i)} b^\top (\mathbf{I}_{d_u} - \mu_{uw}(B^\top \mu_{uw})^{-1} B^\top)\mu_{uy} \\
& \qquad \text{s.t.} \quad -\mu_u \leq \mu_{uw}, \mu_{uy}, \mu_{uz} \leq \mu_u, \quad -\varsigma_u \leq \mathbb{E}(A_iU_i), \mathbb{E}(U_iX_i) \leq \varsigma_u.
\end{aligned}$$

$$\begin{aligned}
\text{bias}_{min}(\mu_u, \varsigma_u) \quad &:= \quad \min_{\mu_{uw}, \mu_{uy}, \mu_{uz}, \mathbb{E}(A_iU_i), \mathbb{E}(U_iX_i)} b^\top (\mathbf{I}_{d_u} - \mu_{uw}(B^\top \mu_{uw})^{-1} B^\top)\mu_{uy} \\
& \qquad \text{s.t.} \quad -\mu_u \leq \mu_{uw}, \mu_{uy}, \mu_{uz} \leq \mu_u, \quad -\varsigma_u \leq \mathbb{E}(A_iU_i), \mathbb{E}(U_iX_i) \leq \varsigma_u.
\end{aligned}$$

In this way, we can reduce the number of sensitivity parameters to two from $d_u \times d_w + d_u \times d_z + d_u + d_u + d_u \times d_x$. One key limitation of this approach is that the sensitivity analysis might be too conservative because we only consider the bound of the bias.

C.3 Proof

In general, by Theorem 2, parameters of the confounding bridge function are identified by

$$\mathbb{E}(Y_i \mid Z_i, A_i, X_i) = \mathbb{E}\{h(W_i, A_i, X_i) \mid Z_i, A_i, X_i\}.$$

In particular, when we assume a linear confounding bridge, i.e., $h(W_i, A_i, X_i; \gamma) = \gamma_0 + \gamma_a A_i + \gamma_x^\top X_i + \gamma_w^\top W_i$, we can identify parameters with the following moment conditions.

$$\mathbb{E}\left\{\{Y_i - h(W_i, A_i, X_i; \gamma)\} \times \eta(A_i, Z_i, X_i)\right\} = 0$$

where $\eta(A_i, Z_i, X_i) = (1, A_i, Z_i, X_i)$. We define $\mathbb{E}\left\{\{Y_i - h(W_i, A_i, X_i; \gamma)\} \times \eta(A_i, Z_i, X_i)\right\} := (m_{i1}, m_{ia}, m_{iz}, m_{ix})$ where m_{i1} and m_{ia} are scalar, m_{iz} is a d_z dimensional vector, and m_{ix} is a d_x dimensional vector. Then, we have

$$m_{i1} := \mathbb{E}\left\{Y_i - (\gamma_0 + \gamma_a A_i + \gamma_x^\top X_i + \gamma_w^\top W_i)\right\} = -\gamma_0 - \mathbb{E}(A_i)\gamma_a - \mu_{0w}^\top \gamma_w + \mathbb{E}(A_i)\mu_{ay},$$

$$\begin{aligned} m_{ia} &:= \mathbb{E}\left\{\{Y_i - (\gamma_0 + \gamma_a A_i + \gamma_x^\top X_i + \gamma_w^\top W_i)\} \times A_i\right\} \\ &= -\mathbb{E}(A_i)\gamma_0 - \mathbb{E}(A_i^2)\gamma_a - \left(\mathbb{E}(A_i)\mu_{0w}^\top + \mathbb{E}(A_i U_i)^\top \mu_{uw} + \mathbb{E}(A_i X_i)^\top \mu_{xw}\right)\gamma_w - \mathbb{E}(A_i X_i)^\top \gamma_x \\ &\quad + \mathbb{E}(A_i)\mu_{0y} + \mathbb{E}(A_i^2)\mu_{ay} + \mathbb{E}(A_i U_i)^\top \mu_{uy} + \mathbb{E}(A_i X_i)^\top \mu_{xy}, \end{aligned}$$

$$\begin{aligned} m_{iz} &:= \mathbb{E}\left\{\{Y_i - (\gamma_0 + \gamma_a A_i + \gamma_x^\top X_i + \gamma_w^\top W_i)\} \times Z_i\right\} \\ &= -\left(\mu_{0z} + \mu_{az}\mathbb{E}(A_i)\right)\gamma_0 - \left(\mu_{0z}\mathbb{E}(A_i) + \mu_{az}\mathbb{E}(A_i^2) + \mathbb{E}(A_i U_i)^\top \mu_{uz} + \mathbb{E}(A_i X_i)^\top \mu_{xz}\right)\gamma_a \\ &\quad - \left((\mathbb{E}(A_i) + \mu_{az}\mathbb{E}(A_i))\mu_{0w}^\top + (\mu_{az}\mathbb{E}(A_i U_i)^\top + \mu_{uz}^\top + \mu_{xz}^\top \rho^\top)\mu_{uw} + \mathbb{E}(A_i X_i)^\top \mu_{xu}\right)\gamma_w \\ &\quad - \left(\mu_{az}\mathbb{E}(A_i X_i)^\top + \mu_{uz}^\top \rho + \mu_{xz}^\top \Sigma_x\right)\gamma_x \\ &\quad + \left(\mu_{0z} + \mathbb{E}(A_i)\mu_{az}\right)\mu_{0y} + \left(\mathbb{E}(A_i)\mu_{0z} + \mathbb{E}(A_i^2)\mu_{az} + \mu_{uz}^\top \mathbb{E}(A_i U_i) + \mu_{xz}^\top \mathbb{E}(A_i X_i)\right)\mu_{ay} \\ &\quad + \left(\mu_{az}\mathbb{E}(A_i U_i)^\top + \mu_{uz}^\top + \mu_{xz}^\top \rho^\top\right)\mu_{uy} + \left(\mu_{az}\mathbb{E}(A_i X_i)^\top + \mu_{uz}^\top \rho + \mu_{xz}^\top \Sigma_x\right)\mu_{xy}, \end{aligned}$$

$$\begin{aligned} m_{ix} &:= \mathbb{E}\left\{\{Y_i - (\gamma_0 + \gamma_a A_i + \gamma_x^\top X_i + \gamma_w^\top W_i)\} \times X_i\right\} \\ &= -\mathbb{E}(A_i X_i)\gamma_a - \left(\Sigma_x \mu_{xw} + \rho^\top \mu_{uw}\right)\gamma_w - \Sigma_x \gamma_x + \mathbb{E}(A_i X_i)\mu_{ay} + \Sigma_x \mu_{xy} + \rho^\top \mu_{uy}, \end{aligned}$$

where we expanded Y_i, W_i and Z_i with respect to (A_i, U_i, X_i) according to the data generating process. Therefore, we can solve for parameters $(\gamma_0, \gamma_a, \gamma_x, \gamma_w)$ by setting $(m_{i1}, m_{ia}, m_{iz}, m_{ix}) = \mathbf{0}_{2+d_z+d_x}$. By solving the moment conditions above, we obtain

$$B^\top \mu_{uw}\gamma_w = B^\top \mu_{uy}$$

where B is a $(d_u \times d_z)$ matrix and

$$B = \left(\mathbf{I}_{d_u} - \rho\Sigma_x^{-1}\rho^\top - \frac{\left(\mathbb{E}(A_iU_i) - \rho\Sigma_x^{-1}\mathbb{E}(A_iX_i)\right)\left(\mathbb{E}(A_iU_i)^\top - \mathbb{E}(A_iX_i)^\top\Sigma_x^{-1}\rho^\top\right)}{\mathbb{E}(A_i^2) - \mathbb{E}(A_i)^2 - \mathbb{E}(A_iX_i)^\top\Sigma_x^{-1}\mathbb{E}(A_iX_i)} \right)\mu_{uz}.$$

When $d_u > d_z$, $B^\top$ is not invertible. Following Cobzaru *et al.* (2022), when we assume $B^\top\mu_{uw}$ is invertible (it is a $(d_z \times d_w)$ matrix), we obtain

$$\gamma_w = (B^\top\mu_{uw})^{-1}B^\top\mu_{uy}.$$

By solving the moment condition, we can write γ_a using γ_w as follows.

$$\gamma_a \quad = \quad \mu_{ay} + b^\top(\mu_{uy} - \mu_{uw}\gamma_w)$$

where b is a d_u dimensional vector and

$$b \quad = \quad \frac{\mathbb{E}(A_iU_i) - \rho\Sigma_x^{-1}\mathbb{E}(A_iX_i)}{\mathbb{E}(A_i^2) - \mathbb{E}(A_i)^2 - \mathbb{E}(A_iX_i)^\top\Sigma_x^{-1}\mathbb{E}(A_iX_i)}.$$

Therefore, plugging in the expression of γ_w, we obtain

$$\gamma_a \quad = \quad \mu_{ay} + b^\top(\mathbf{I}_{d_u} - \mu_{uw}(B^\top\mu_{uw})^{-1}B^\top)\mu_{uy},$$

which completes the proof. □

D Asymptotic Properties of the DNC Estimator

D.1 Setup and Regularity Conditions

To derive asymptotic properties of our estimator, we assume the standard GMM regularity conditions (Hansen, 1982; Newey and McFadden, 1994).

The GMM regularity conditions:

- Parameter space Θ is compact.
- $m(L_{n,i};\theta)$ is differentiable in $\theta \in \Theta$ with probability one.
- $m(L_{n,i};\theta)$ and $\frac{\partial}{\partial\theta}m(L_{n,i};\theta)$ are continuous at each $\theta \in \Theta$ with probability one.
- $\mathbb{E}\{m(L_{n,i};\theta)\} = 0$ only when $\theta = \theta_0$, and θ_0 is in the interior of Θ.
- $\mathbb{E}\{m(L_{n,i};\theta)\}$ and $\mathbb{E}\{\frac{\partial}{\partial\theta}m(L_{n,i};\theta)\}$ are continuous in θ.
- $\frac{1}{n}\sum_{i=1}^n m(L_{n,i};\theta)$ is stochastically equicontinuous on Θ.

- $\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta} m(L_{n,i};\theta)$ is stochastically equicontinuous on Θ.
- $M_0 = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left\{\frac{\partial}{\partial\theta} m(L_{n,i};\theta_0)\right\}$ is full rank.
- For p that satisfies Assumption 3, $\sup_{n\geq 1}\max_{i\in N_n}\mathbb{E}\{|c^\top m(L_{n,i};\theta)|^p\} < \infty$ for any c with $||c||_2 = \sqrt{c^\top c} = 1$ for all $\theta \in \Theta$.
- For p that satisfies Assumption 3, $\sup_{n\geq 1}\max_{i\in N_n}\mathbb{E}\{|\tilde{c}^\top \frac{\partial}{\partial\theta} m(L_{n,i};\theta)|^p\} < \infty$ for any $\tilde{c}$ with $||\tilde{c}||_2 = \sqrt{\tilde{c}^\top \tilde{c}} = 1$ for all $\theta \in \Theta$.

We first define the GMM objective function:

$$\begin{aligned}
Q_n(\theta) &= \left\{\frac{1}{n}\sum_{i=1}^n \mathbb{E}\{m(L_{n,i};\theta)\}\right\}^\top \Omega \left\{\frac{1}{n}\sum_{i=1}^n \mathbb{E}\{m(L_{n,i};\theta)\}\right\}, \\
\widehat{Q}_n(\theta) &= \left\{\frac{1}{n}\sum_{i=1}^n m(L_{n,i};\theta)\right\}^\top \Omega \left\{\frac{1}{n}\sum_{i=1}^n m(L_{n,i};\theta)\right\}.
\end{aligned}$$

Then, the GMM estimator of θ can be written as:

$$\widehat{\theta} = \text{argmin}_{\theta\in\Theta}\widehat{Q}_n(\theta). \tag{A.32}$$

D.2 Proof of Theorem 3

Given that our DNC estimator $\widehat{\tau}(a,a')$ for $\tau(a,a')$ corresponds to the first element of $\widehat{\theta}$, we state theoretical properties in terms of $\widehat{\theta}$, which imply Theorem 3.

Consistency. We first want to show consistency of the GMM estimator:

$$\widehat{\theta} \xrightarrow{p} \theta_0.$$

Proof: Under Assumption 3, Proposition 3.1 by Kojevnikov *et al.* (2020) implies point-wise convergence of $\frac{1}{n}\sum_{i=1}^n m(L_{n,i};\theta)$. That is, for all $\theta \in \Theta$,

$$\frac{1}{n}\sum_{i=1}^n \{m(L_{n,i};\theta) - \mathbb{E}\{m(L_{n,i};\theta)\}\} \xrightarrow{p} 0. \tag{A.33}$$

Under the stochastic equicontinuity, the compactness of the parameter space, and the continuity of moment, we establish the uniform convergence (Newey and McFadden, 1994).

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^n \{m(L_{n,i};\theta) - \mathbb{E}\{m(L_{n,i};\theta)\}\}\right| \xrightarrow{p} 0. \tag{A.34}$$

Therefore, under the GMM regularity conditions described above,

$$\sup_{\theta\in\Theta}\left|\widehat{Q}_n(\theta) - Q_n(\theta)\right| \xrightarrow{p} 0. \tag{A.35}$$

Finally, under the GMM regularity conditions described above, we have (i) $Q_n(\theta)$ is uniquely minimized at θ_0, (ii) parameter space Θ is compact, (iii) $Q_n(\theta)$ is continuous, and (iv) the uniform convergence (equation (A.35)). Therefore, Theorem 2.1 of Newey and McFadden (1994) implies

$$\widehat{\theta} \xrightarrow{p} \theta_0,$$

which completes the proof of consistency. □

Asymptotic Normality. Next, we show asymptotic normality.

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} \mathsf{Normal}(0, \Sigma)$$

where

$$\Sigma = \Gamma_0 \text{Var}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(L_{n,i}; \theta_0)\right) \Gamma_0^\top,$$

$$\Gamma_0 = (M_0^\top \Omega M_0)^{-1} M_0^\top \Omega, \quad M_0 = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left\{\frac{\partial}{\partial \theta} m(L_{n,i}; \theta_0)\right\}.$$

Proof: By definition, we have

$$\widehat{\theta} = \text{argmin}_{\theta \in \Theta} \widehat{Q}_n(\theta)$$

We take the first order condition.

$$\frac{\partial \widehat{Q}_n(\widehat{\theta})}{\partial \theta} = 0$$

Using the mean-value expansion, we have

$$\begin{aligned}
\sqrt{n}(\widehat{\theta} - \theta_0) &= -\left\{\frac{\partial^2 \widehat{Q}_n(\widetilde{\theta})}{\partial \theta \partial \theta^\top}\right\}^{-1} \times \sqrt{n} \frac{\partial \widehat{Q}_n(\theta_0)}{\partial \theta} \\
&= -\left\{\frac{\partial^2 \widehat{Q}_n(\widetilde{\theta})}{\partial \theta \partial \theta^\top}\right\}^{-1} \times \left\{\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta^\top} m(L_i; \theta_0)\right\}^\top \Omega \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(L_i; \theta_0)
\end{aligned}$$

where $\widetilde{\theta}$ is a mean value, located between $\widehat{\theta}$ and θ_0, and

$$\begin{aligned}
\left[\frac{\partial^2 \widehat{Q}_n(\widetilde{\theta})}{\partial \theta \partial \theta^\top}\right]_{jk} &= \left\{\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta_j} m(L_i; \widetilde{\theta})\right\}^\top \Omega \left\{\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta_k} m(L_i; \widetilde{\theta})\right\} \\
&\quad + \left\{\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial \theta_j \partial \theta_k} m(L_{n,i}; \widetilde{\theta})\right\}^\top \Omega \left\{\frac{1}{n} \sum_{i=1}^{n} m(L_{n,i}; \widetilde{\theta})\right\}.
\end{aligned}$$

Therefore, under the GMM regularity conditions, Assumption 3, and consistency of $\widehat{\theta}$,

$$\left\{\frac{\partial^2 \widehat{Q}_n(\widetilde{\theta})}{\partial \theta \partial \theta^\top}\right\}^{-1} \xrightarrow{p} (M_0^\top \Omega M_0)^{-1},$$

$$\left\{\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta^\top} m(L_i; \theta_0)\right\}^\top \Omega \xrightarrow{p} M_0^\top \Omega.$$

Thus,

$$\sqrt{n}(\widehat{\theta}-\theta_0) = -(M_0^\top \Omega M_0)^{-1} M_0^\top \Omega \times \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(L_{n,i};\theta_0) + o_p(1).$$

Finally, under Assumption 3, the Cramér–Wold device and the network CLT (Theorem 3.2) by Kojevnikov *et al.* (2020) imply

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(L_{n,i};\theta_{0,n}) \xrightarrow{d} \mathcal{N}\left(0, \mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(L_{n,i};\theta_0)\right)\right).$$

By combining the results using the Slutsky's theorem, we obtain the desired result.

$$\sqrt{n}(\widehat{\theta}-\theta_0) \xrightarrow{d} \mathsf{Normal}(0,\Sigma),$$

which completes the proof. □

D.3 Proof of Theorem 4

We consider asymptotic properties of the network HAC variance estimator. In addition to the regularity conditions required to prove Theorem 3, we also require the following regularity conditions for the choice of kernel and bandwidth. With p that satisfies Assumption 3,

$$\lim_{n\to\infty} \sum_{s\geq 0} |\omega(s/b_n) - 1| \rho_n(s) \beta_{n,s}^{1-2/p} = 0 \ \text{ a.s.},$$

where $\rho_n(s)$ measures the average number of network peers at the distance s, $\rho_n(s) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{N}_n(i;s)$.

Proof: Given that $\widehat{\theta}$ is a consistent estimator of θ_0, using the continuous mapping theorem under the GMM regularity condition, we need to prove that

$$\widetilde{\Lambda}_n = \sum_{s\geq 0} \omega(s/b_n) \left\{ \frac{1}{n} \sum_{i\in N_n} \sum_{j\in\mathcal{N}_n(i;s)} m(L_{n,i};\theta_0) m(L_{n,j};\theta_0)^\top \right\}.$$

is a consistent estimator of Λ_0. Because we assume that $m(L_{n,i};\theta_0)$ is ψ-weakly dependent (Assumption 3), under the regularity condition on the choice of kernel and bandwidth (equation (19)), Proposition 4.1 of Kojevnikov *et al.* (2020) implies that $\widetilde{\Lambda}_n$ is a consistent estimator for Λ_0.

Moreover, under Assumption 3 and the GMM regularity conditions, we obtain consistency of $\widehat{M}$: $\widehat{M} - M_0 \xrightarrow{p} 0$, where $\widehat{M} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta} m(L_{n,i};\widehat{\theta})$ and $M_0 = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left\{\frac{\partial}{\partial\theta} m(L_{n,i};\theta_0)\right\}$. Finally, we can combine the results to obtain the desired result.

$$\widehat{\Sigma} - \Sigma \xrightarrow{p} 0$$

where

$$\Sigma = \Gamma_0 \Lambda_0 \Gamma_0^\top, \quad \widehat{\Sigma} = \widehat{\Gamma}\widehat{\Lambda}\widehat{\Gamma}^\top$$
$$\Gamma_0 = (M_0^\top \Omega M_0)^{-1} M_0^\top \Omega, \quad \widehat{\Gamma} = (\widehat{M}^\top \Omega \widehat{M})^{-1} \widehat{M}^\top \Omega,$$

which completes the proof. □

D.4 Choice of Bandwidth

In general settings of network-dependent errors (Section 3.5), one must estimate a bandwidth b_n for the network HAC variance estimator (e.g., Kojevnikov *et al.*, 2020) because how far network dependence persists is a priori unknown. However, when the following assumption holds, we can analytically select the bandwidth.

Assumption 10

1. *The ACPE is equal to a linear function of parameters γ in the confounding bridge function.*

2. *There exists integer s^* such that for units i, j with distance $d_n(i, j) \geq s^*$,*

$$L_{n,j} \perp\!\!\!\perp L_{n,i} \mid A_i, Z_i, X_i, U_i.$$

Assumption 10.1 holds for a linear confounding bridge function as we consider in this section. Assumption 10.2 requires that observed data for unit j, $L_{n,j}$, is conditionally independent of observed data for unit i, $L_{n,i}$, given unit i's treatment, NCEs, observed pre-treatment covariates, and the unmeasured confounder. This conditional independence is required only upon conditioning on latent confounder U_i, and thus, it does not restrict network dependence of the observed data law itself.

Importantly, we emphasize that Assumption 10.2 holds under many relevant scenarios. Figure 3 provides examples of causal graphs where Assumption 10.2 is satisfied. In Figure 3 of the main paper, suppose one uses C_i as the NCO and $Z_i = \{C_j : j \in \mathcal{N}_n(i; 1)\}$ as the NCEs. Then, Assumption 10.2 holds with $s^* = 2$. If one uses auxiliary variables of both peers and peers-of-peers, $Z_i = \{C_j : j \in \{\mathcal{N}_n(i; 1), \mathcal{N}_n(i; 2)\}\}$, Assumption 10.2 holds with $s^* = 3$. Figure A1 represents another example. Suppose one exploits Y_{i1} as the NCO and $Z_i = \{Y_{jt} : j \in \mathcal{N}_n(i; 2),\ t \in \{1, 2\}\}$ as the NCEs. Then, Assumption 10.2 holds with $s^* = 4$.

Under Assumption 10, Lemma 7 below shows that one can analytically select the bandwidth for the network HAC variance estimator.

Lemma 7 *Suppose the conditions given in Theorem 3 hold. Under Assumption 10.1, we can simplify the moment function to $\widetilde{m}(L_{n,i}; \gamma) = \{Y_{i2} - h(W_i, A_i, X_i; \gamma)\} \times \eta(A_i, Z_i, X_i)$ as our target parameter is a linear function of γ. Then, under Assumption 10.2 with integer s^*, we can use the following network HAC variance estimator for $\widehat{\gamma}$, which is the GMM estimator with moment function $\widetilde{m}(L_{n,i}; \gamma)$.*

$$\widehat{\mathrm{Var}}(\widehat{\gamma}) = \frac{1}{n} \widehat{\Gamma}_\gamma \widehat{\Lambda}_{s^*} \widehat{\Gamma}_\gamma^\top \tag{A.36}$$

where

$$\widehat{\Lambda}_{s^*} = \sum_{s=0}^{s^*-1} \omega(s/b_n) \left\{ \frac{1}{n} \sum_{i \in N_n} \sum_{j \in \mathcal{N}_n(i;s)} \widetilde{m}(L_{n,i}; \widehat{\gamma}) \widetilde{m}(L_{n,j}; \widehat{\gamma})^\top \right\},$$

$\widehat{\Gamma}_\gamma = (\widehat{M}_\gamma^\top \Omega \widehat{M}_\gamma)^{-1} \widehat{M}_\gamma^\top \Omega$, *and* $\widehat{M}_\gamma = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \gamma} \widetilde{m}(L_{n,i}; \widehat{\gamma})$.

The key to this result is that, to compute the variance of a sum of products of moments, one only needs to consider moments of units with distance less than s^* (i.e., we added $s \in \{0, 1, \ldots, s^* - 1\}$), as the remaining contributions are null. This is in contrast to the default network HAC variance estimator (equation (17)) where we have to incorporate all products of moments of units with distance less than b_n, which is in general larger than s^*. We provide a proof in Appendix D of the supplementary material.

For the linear DNC estimator, Assumption 10.1 automatically holds, and thus, as long as Assumption 10.2 holds, we can rely on this analytical choice of bandwidth. We evaluate both analytical and default bandwidth selections (equation (20)) in a simulation study (Appendix E).

D.4.1 Proof of Lemma 7

Under Assumption 10.1, the ACPE can be represented as a linear function of parameters γ in the outcome confounding bridge function. Under this setting, it is sufficient to obtain multivariate asymptotic normality and consistent variance estimator for γ. As a result, we can simplify the moment function to be

$$\widetilde{m}(L_{n,i}; \gamma) = \{Y_{i2} - h(W_i, A_i, X_i; \gamma)\} \times \eta(A_i, Z_i, X_i).$$

Under Assumption 10.2, there exists integer s^* such that for units i, j with the distance $d_n(i, j) \geq s^*$,

$$L_{n,j} \perp\!\!\!\perp L_{n,i} \mid A_i, Z_i, X_i, U_i.$$

For such s^* and units i, j, we have

$$\widetilde{m}(L_{n,j}; \gamma_0) \perp\!\!\!\perp \widetilde{m}(L_{n,i}; \gamma_0) \mid A_i, Z_i, X_i, U_i. \tag{A.37}$$

In addition, under Assumptions 2.2 and 2.3, we have

$$\mathbb{E}\{\widetilde{m}(L_{n,i}; \gamma_0) \mid A_i, Z_i, X_i, U_i\} = 0. \tag{A.38}$$

Combining equations (A.37) and (A.38), we obtain

$$\begin{aligned}
&\mathbb{E}\{\widetilde{m}(L_{n,i}; \gamma_0)\widetilde{m}(L_{n,j}; \gamma_0)^\top \mid A_i, Z_i, X_i, U_i\} = 0 \\
\Longrightarrow\quad &\mathbb{E}\{\widetilde{m}(L_{n,i}; \gamma_0)\widetilde{m}(L_{n,j}; \gamma_0)^\top\} = 0.
\end{aligned}$$

for integer s^* and units i, j with the distance $d_n(i, j) \geq s^*$. Therefore,

$$\Lambda_0 = \sum_{s=0}^{s^*-1} \Lambda_0(s)$$

where

$$\Lambda_0(s) = \left\{ \frac{1}{n} \sum_{i \in N_n} \sum_{j \in \mathcal{N}_n(i;s)} \mathbb{E}\{\widetilde{m}(L_{n,i}; \gamma_0)\widetilde{m}(L_{n,j}; \gamma_0)^\top\} \right\}.$$

We can obtain its estimator as follows.

$$\widehat{\Lambda}_{s^*} = \sum_{s=0}^{s^*-1} \omega(s/b_n) \left\{ \frac{1}{n} \sum_{i \in N_n} \sum_{j \in \mathcal{N}_n(i;s)} \widetilde{m}(L_{n,i}; \widehat{\gamma})\widetilde{m}(L_{n,j}; \widehat{\gamma})^\top \right\}.$$

Finally, we obtain the variance estimator for $\widehat{\gamma}$.

$$\widehat{\text{Var}}(\widehat{\gamma}) = \frac{1}{n}\widehat{\Gamma}_{\gamma}\widehat{\Lambda}_{s^*}\widehat{\Gamma}_{\gamma}^{\top}. \tag{A.39}$$

where $\widehat{\Gamma}_{\gamma} = (\widehat{M}_{\gamma}^{\top}\Omega\widehat{M}_{\gamma})^{-1}\widehat{M}_{\gamma}^{\top}\Omega$, and $\widehat{M}_{\gamma} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\gamma}\widetilde{m}(L_{n,i};\widehat{\gamma})$, which completes the proof. $\square$

D.5 Heterogeneous Expectation

In Section 3.5, we assume that the expectation of the causal peer effect, $\mathbb{E}\{Y_{i2}(a)-Y_{i2}(a') \mid \mathcal{G}_n\}$, is constant across units, while we allow for network-dependent (non-independent) errors. Here, to examine the heterogeneous expectation, we explicitly write out the conditioning on $\mathcal{G}_n$. In this section, we allow for heterogeneous expectation across units. As we observe only one sample of interconnected units in a single network, we have to make some assumptions to make progress. In this vein, we assume that $\mathbb{E}\{Y_{i2}(a) - Y_{i2}(a') \mid \mathcal{G}_n\}$ depends only on a summary statistic of network $\mathcal{G}_n$, which we denote by vector $\mathbf{g}_i$. For example, $\mathbf{g}_i$ could be the network-degree of unit i, centrality of unit i, or other network summary statistics. This is a common assumption scholars make in practice, and is similar to the idea of the exposure mapping (Aronow and Samii, 2017), which is used to reduce dimensionality of the potential outcomes.

Formally, we assume $\mathbb{E}\{Y_{i2}(a) - Y_{i2}(a') \mid \mathcal{G}_n\} = \mathbb{E}\{Y_{i2}(a) - Y_{i2}(a') \mid \mathbf{g}_i\}$. We then posit a model for the conditional expectation $\mathbb{E}\{Y_{i2}(a) \mid \mathbf{g}_i\}$ with shared coefficients. This allows us to accommodate heterogeneous expectation across units in the network, while we can still make statistical inference about the target estimand with network-dependent errors.

As a concrete example, consider the following linear model with coefficients φ.

$$\mathbb{E}\{Y_{i2}(a) \mid \mathbf{g}_i\} \quad = \quad \varphi_0 + \varphi_1 \cdot a + \{\ell(\mathbf{g}_i)^{\top}\varphi_2\} \cdot a$$

where $\ell(\mathbf{g}_i)$ is a user-specified function of $\mathbf{g}_i$. Under this model, we can re-write the ACPE as follows.

$$\tau(a,a') := \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\{Y_{i2}(a) - Y_{i2}(a') \mid \mathcal{G}_n\} = \varphi_1 \cdot (a-a') + (\overline{\ell}^{\top}\varphi_2)\cdot(a-a')$$

where $\overline{\ell} = \frac{1}{n}\sum_{i=1}^{n}\ell(\mathbf{g}_i)$. To estimate the ACPE, we first modify the moment function as follows.

$$m^{\dagger}(L_{n,i};\theta) \quad = \quad \left\{\tau + \left(\ell(\mathbf{g}_i) - \overline{\ell}\right)^{\top}\varphi_2\cdot(a-a')\right\} - \{h(W_i,a,X_i;\gamma) - h(W_i,a',X_i;\gamma)\},$$

where $\theta = (\tau,\varphi_2,\gamma)$. We then show that $\mathbb{E}\{m^{\dagger}(L_{n,i};\theta) \mid \mathcal{G}_n\} = 0$ for all $i \in N_n$. We start with the first term.

$$\begin{aligned}\mathbb{E}\left\{\tau + \left(\ell(\mathbf{g}_i) - \overline{\ell}\right)^{\top}\varphi_2\cdot(a-a') \mid \mathcal{G}_n\right\} &= \varphi_1\cdot(a-a') + \ell(\mathbf{g}_i)^{\top}\varphi_2\cdot(a-a') \\ &= \mathbb{E}\{Y_{i2}(a) - Y_{i2}(a') \mid \mathcal{G}_n\}.\end{aligned}$$

We next consider the second term. Under Assumption 2,

$$\mathbb{E}\left(\{h(W_i,a,X_i;\gamma) - h(W_i,a',X_i;\gamma)\} \mid \mathcal{G}_n\right) \quad = \quad \mathbb{E}\{Y_{i2}(a) - Y_{i2}(a') \mid \mathcal{G}_n\},$$

which shows that $\mathbb{E}\{m^{\dagger}(L_{n,i};\theta) \mid \mathcal{G}_n\} = 0$ for all $i \in N_n$. Therefore, we can use the following moment functions to estimate the ACPE $\tau(a, a')$.

$$m^*(L_{n,i};\theta) = \left\{ \begin{array}{l} m^{\dagger}(L_{n,i};\theta) \times \eta^*(\mathbf{g}_i) \\ \{Y_{i2} - h(W_i, A_i, X_i; \gamma)\} \times \eta(A_i, Z_i, X_i) \end{array} \right\},$$

where $\eta^*(\mathbf{g}_i) = (1, \ell(\mathbf{g}_i)^{\top})^{\top}$. Under the same assumption used in Section 3.5, we can consistently estimate the ACPE and construct an asymptotic confidence interval. $\square$

E Simulation Study

We investigate the finite sample performance of the proposed DNC estimator of the ACPE using networks of varying density and size. We also examine the performance of the proposed estimator in settings where key identification assumptions are violated. In Appendix E.2, we consider violation of the negative control assumption (Assumption 2.2), and in Appendix E.3, we examine violation of the outcome confounding bridge assumption (Assumption 2.3) due to violation of the underlying completeness condition. In Appendix E.4, we consider how the performance of the estimator changes as we change the association between NCO W and unobserved confounder U. In Appendix E.5, we consider a binary outcome to show that our nonparametric identification results can accommodate different types of outcomes.

E.1 Finite Sample Performance

Setup. To investigate the performance of the proposed estimator, we consider two different types of networks: the small world network and the real-world network from Add Health data. To generate the small world network, we use `sample_smallworld` with the rewiring probability of 0.15 based on R package `igraph`. We consider two levels of densities: low (the average degree of four) and high (the average degree of eight). Add Health project collected detailed information about friendship networks by an in-school survey. We define friendships as symmetric relationships: the pair of students i and j in the same school are coded as friends if either i lists j as a friend, or j lists i as a friend, or both. While we analyze this data more thoroughly in Section 4, we also use it here as basis for the simulation. For each simulation, we generate a network of size n where we consider sample size $n \in \{500, 1000, 2000, 4000\}$. For the small-world network, we generate a single network of size n. For the Add Health network, we retain the original network characteristics by randomly sampling schools with probability proportional to its size until the total sample size reaches n. The average degree of the Add Health network ranges from 3.82 to 5.95, which are in the middle of the low-density small-world network (average degree = 4) and the high-density small world network (average degree = 8). The density of the Add Health network ranges from 0.15 to 0.77 %, which are close to the density of the low-density small-world network. Thus, these three different types of networks jointly cover a wide range of network density and size. See Table A1 for more details.

Given a network, we simulate data with the following data-generating mechanism: For units $i = 1, \ldots, n$,

(1) Unobserved confounder with network dependence: $U_i = \sum_{s \geq 0} \zeta^s \sum_{j \in \mathcal{N}(i;s)} \widetilde{U}_j / |\mathcal{N}(i;s)|$ where $\zeta = 0.8$ and $\widetilde{U}_j \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0,1)$. This data generating process for a network-dependent variable follows a simulation setup of Kojevnikov *et al.* (2020).

(2) Observed covariates with network dependence: $X_i = (X_{i1}, X_{i2}, X_{i3})$ where, for $k \in \{1,2,3\}$, $X_{ik} = \sum_{s \geq 0} \zeta^s \sum_{j \in \mathcal{N}(i;s)} \widetilde{X}_{jk} / |\mathcal{N}(i;s)|$, $\zeta = 0.8$, and $\widetilde{X}_{jk} \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0,1)$.

(3) Observed auxiliary variable: $C_i = U_i + \beta_c^\top X_i + \epsilon_{i0}$ where $\epsilon_{i0} \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0,1)$.

(4) Focal behavior at the baseline: $Y_{i1} = U_i + 0.05C_i + \beta_1^\top X_i + \epsilon_{i1}$ where $\epsilon_{i1} \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0, 1)$.

(5) Focal behavior at the follow-up: $Y_{i2} = \tau A_i + 0.2Y_{i1} + 3U_i + 0.05C_i + \beta_2^\top X_i + \epsilon_{i2}$ where $\epsilon_{i2} \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0, 1)$. The treatment variable A_i is defined as $A_i = \sum_{j \in \mathcal{N}(i;1)} Y_{j1}/|\mathcal{N}(i; 1)|$,

where $\beta_c = (0.05, 0.05, 0.05), \beta_1 = (-1, -1, -1)$, and $\beta_2 = (-1, -1, -1)$. The above models imply that the ACPE is τ, which we set to be 0.3. We can use $W_i = C_i$ as the NCO, and $Z_i = \sum_{j \in \mathcal{N}(i;1)} C_j/|\mathcal{N}(i; 1)|$ as the NCE. Under this setup, the linear confounding bridge function, $h(W_i, A_i, X_i; \gamma) = \gamma_\alpha + \tau A_i + \gamma_W W_i + \gamma_X^\top X_i$, satisfies Assumption 2.

We evaluate the performance of the proposed DNC estimator and the network HAC variance estimator. We evaluate two choices of bandwidth for the network HAC variance estimator. First, we use the bandwidth of 2, which we analytically derive based on Lemma 7. Second, we also use the default bandwidth b_n (equation (20)) suggested in Kojevnikov *et al.* (2020). We use the Parzen kernel function, i.e., $\omega(x) = 1 - 6x^2 + 6|x|^3$ if $0 \leq |x| \leq 1/2$, $\omega(x) = 2(1 - |x|)^3$ if $1/2 < |x| \leq 1$, and $\omega(x) = 0$ if $1 < |x|$.

For reference, we also report two other estimators. (1) The ordinary least squares estimator where we regress Y_{i2} on the treatment variable A_i and a set of observed variables $(Y_{i1}, C_i, X_{i1}, X_{i2}, X_{i3})$. This estimator is consistent only under conditional ignorability, which is violated due to unmeasured network confounder U_i under this simulation setup. Thus, this OLS estimator quantifies the amount of network confounding that the DNC estimator has to correct for. For coverage, we apply the network HAC variance estimator (Kojevnikov *et al.*, 2020) to residuals in order to make comparison clear. (2) We also report the difference-in-differences style estimator proposed in Egami (2018). This estimator is consistent under the assumption that the confounding effect of U on the primary outcome Y_{i2} is the same as the confounding effect of U on Y_{i1}. This assumption is violated under this simulation setup. For coverage, we apply the network HAC variance estimator (Kojevnikov *et al.*, 2020) to residuals in order to make comparison clear.

Results. We generate 2000 simulations and evaluate estimators in terms of absolute mean bias, standard error (computed as standard deviation of point estimates across simulations), root mean squared error (RMSE), and coverage of 95% confidence intervals based on the network HAC variance estimator. We standardize the first three quantities by the true ACPE to ease interpretation. Table A1 summarizes the results of the simulation study.

Our proposed DNC estimator remained stable with relatively small bias across all scenarios, and the bias reduced as sample size increased. As expected, standard errors of the proposed DNC estimators were larger than the biased OLS estimators, but the RMSE of the DNC estimator was smaller due to smaller bias. Importantly, in the OLS estimator, we include the baseline outcome Y_{i1} as a control variable, and yet, we still see the OLS estimator is heavily biased. This verifies an important point by Shalizi and Thomas (2011) that just controlling for the baseline outcome does not allow researchers to estimate the ACPE. Importantly, our proposed approach uses the baseline outcome as the NCO and combine it with the NCE, and this unique combination of double negative controls allows for identification of the ACPE.

Simulation Design				DNC					OLS				DID			
Network	Sample Size	Average degree	Density	Bias	Standard Error	RMSE	Coverage (Analytical)	Coverage (Default)	Bias	Standard Error	RMSE	Coverage	Bias	Standard Error	RMSE	Coverage
SW-4	500	4.00	0.80	0.14	0.63	0.65	0.96	0.96	0.98	0.26	1.02	0.03	1.09	0.26	1.12	0.08
	1000	4.00	0.40	0.07	0.42	0.42	0.95	0.95	1.00	0.18	1.01	0.00	1.10	0.18	1.12	0.00
	2000	4.00	0.20	0.03	0.28	0.28	0.95	0.94	1.01	0.13	1.01	0.00	1.11	0.13	1.12	0.00
	4000	4.00	0.10	0.02	0.20	0.20	0.95	0.94	1.00	0.09	1.01	0.00	1.12	0.09	1.12	0.00
SW-8	500	8.00	1.60	0.25	1.23	1.26	0.96	0.96	0.88	0.35	0.95	0.26	0.97	0.36	1.03	0.44
	1000	8.00	0.80	0.11	0.58	0.59	0.96	0.96	0.88	0.25	0.91	0.04	0.97	0.25	1.00	0.10
	2000	8.00	0.40	0.05	0.38	0.39	0.94	0.94	0.90	0.17	0.92	0.00	0.99	0.18	1.01	0.00
	4000	8.00	0.20	0.02	0.26	0.26	0.95	0.95	0.89	0.12	0.90	0.00	0.99	0.12	0.99	0.00
Add Health	500	3.82	0.77	0.15	0.72	0.74	0.96	0.87	0.98	0.29	1.02	0.06	1.09	0.30	1.13	0.12
	1000	4.80	0.48	0.06	0.46	0.46	0.95	0.94	0.95	0.20	0.97	0.00	1.06	0.21	1.08	0.01
	2000	5.69	0.28	0.03	0.31	0.31	0.95	0.95	0.93	0.15	0.94	0.00	1.03	0.15	1.04	0.00
	4000	5.95	0.15	0.02	0.22	0.22	0.94	0.94	0.92	0.10	0.92	0.00	1.02	0.11	1.03	0.00

Table A1: Operating Characteristics of Estimators under Different Networks.

Note: We consider three different networks; the small world network model with the average degree of four (SW-4) and eight (SW-8), and the Add Health network. Absolute mean bias, standard error, and RMSE for both estimators are standardized by the true ACPE.

As the required assumption is violated, the difference-in-differences style estimator is also biased, and the RMSE is as high as the OLS estimator, and it is much higher than that of the proposed DNC estimator.

Compared to the low-density small-world network (SW-4), bias, standard errors, and RMSE of the DNC estimator were larger in the high-density small-world network (SW-8). Results for the Add Health network fell somewhere in between. The coverage of 95% confidence intervals was close to the nominal level when the analytical bandwidth was chosen. While coverage with default bandwidth tended to under-cover slightly at smaller sample sizes in the Add Health network structure, it improved as sample size increased. They indicated that our proposed standard error estimation provided valid inference. These results confirmed our theoretical results in finite sample and demonstrated the advantages of the proposed DNC estimator.

E.2 Violation of Negative Control Assumptions

Setup. In this section, we consider violations of the negative control assumption (Assumption 2.2). In particular, we modify the data generating mechanism of Appendix E as follows. For units $i = 1, \ldots, n$,

(1) Unobserved confounder with network dependence: $U_i = \sum_{s \geq 0} \zeta^s \sum_{j \in \mathcal{N}(i;s)} \widetilde{U}_j / |\mathcal{N}(i;s)|$ where $\zeta = 0.8$ and $\widetilde{U}_j \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0,1)$. This part is the same as the one used in Appendix E.

(2) Observed covariates with network dependence: $X_i = (X_{i1}, X_{i2}, X_{i3})$ where, for $k \in \{1,2,3\}$, $X_{ik} = \sum_{s \geq 0} \zeta^s \sum_{j \in \mathcal{N}(i;s)} \widetilde{X}_{jk} / |\mathcal{N}(i;s)|$, $\zeta = 0.8$, and $\widetilde{X}_{jk} \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0,1)$. This part is the same as the one used in Appendix E.

(3) Observed auxiliary variable: $C_i = \sum_{s \geq 0} (\zeta_C)^s \sum_{j \in \mathcal{N}(i;s)} \widetilde{C}_j / |\mathcal{N}(i;s)|$ where $\widetilde{C}_i = U_i + \beta_c^\top X_i + \epsilon_{i0}$ where $\epsilon_{i0} \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0,1)$ and $\beta_c = (0.05, 0.05, 0.05)$. This part is the difference from the one used in Appendix E.

(4) Focal behavior at the baseline: $Y_{i1} = U_i + 0.05 C_i + \beta_1^\top X_i + \epsilon_{i1}$ where $\epsilon_{i1} \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0,1)$ and $\beta_1 = (-1,-1,-1)$. This part is the same as the one used in Appendix E.

(5) Focal behavior at the follow-up: $Y_{i2} = \tau A_i + 0.2 Y_{i1} + 3 U_i + 0.05 C_i + \beta_2^\top X_i + \epsilon_{i2}$ where $\epsilon_{i2} \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0,1)$, and $\beta_2 = (-1,-1,-1)$. The treatment variable A_i is defined as $A_i = \sum_{j \in \mathcal{N}(i;1)} Y_{j1} / |\mathcal{N}(i;1)|$. This part is the same as the one used in Appendix E.

The main and only difference is in (3) where we allow for network association between auxiliary variable C across units. Because we use $W_i = C_i$ as NCO, and $Z_i = \sum_{j \in \mathcal{N}(i;1)} C_j / |\mathcal{N}(i;1)|$ as NCE, this network association violates assumptions for NCO and NCE (Assumption 2.2).

We consider three different levels of the violation using parameter $\zeta_C \in \{0.02, 0.10, 0.50\}$. We call them "Small", "Moderate", and "Large" violations in Table A2. We fix sample size to be 1000, and we generate 2000 simulations to evaluate estimators in terms of the absolute mean bias, the standard error (computed as the standard deviation of point estimates across simulations), the root mean squared error (RMSE), and coverage of 95% confidence intervals based on the network HAC variance estimator. We standardize the first three quantities by the true ACPE to ease interpretation.

Results. Table A2 summarizes the results of the simulation study. Our proposed DNC estimator has small bias and has reasonable coverage when the violation is "small." However, as we expect, the larger is the violation, the bias is larger and coverage performance becomes poorer.

Simulation Design		**DNC**				
Network	Violation	Bias	Standard Error	RMSE	Coverage (Analytical)	Coverage (Default)
SW-4	Small	0.04	0.38	0.39	0.94	0.93
	Moderate	0.32	0.30	0.45	0.78	0.78
	Large	0.74	0.25	0.78	0.12	0.12
SW-8	Small	0.02	0.53	0.53	0.95	0.94
	Moderate	0.26	0.43	0.50	0.87	0.86
	Large	0.66	0.35	0.74	0.46	0.46
Add Health	Small	0.03	0.41	0.41	0.94	0.93
	Moderate	0.34	0.34	0.48	0.80	0.78
	Large	0.81	0.28	0.86	0.13	0.13

Table A2: Operating Characteristics when the Negative Control Assumptions are Violated.

Note: We consider three different levels of violation: "Small" ($\zeta_C = 0.02$), "Moderate" ($\zeta_C = 0.10$), and "Large" ($\zeta_C = 0.50$). We examine the same three different networks; the small world network model with the average degree of four (SW-4) and eight (SW-8), and the Add Health network. For the DNC estimator, we report the absolute mean bias, the standard error, the RMSE, and coverage of the 95% confidence intervals based on the analytical bandwidth and the default bandwidth. The absolute mean bias, the standard error, and the RMSE for both estimators are standardized by the true ACPE.

E.3 Violation of Confounding Bridge Assumption due to Completeness

Setup. In this section, we consider violations of the outcome confounding bridge assumption (Assumption 2.3). In particular, we consider violation of the completeness condition (Assumption 5) we use to prove the existence of an outcome confounding bridge function.

In particular, we modify the data generating mechanism of Appendix E as follows. For units $i = 1, \ldots, n$,

(1) Two unobserved confounders with network dependence: For $k \in \{1, 2\}$, $U_{ik} = \sum_{s \geq 0} \zeta^s \sum_{j \in \mathcal{N}(i;s)} \widetilde{U}_{jk} / |\mathcal{N}(i; s)|$ where $\zeta = 0.8$ and $\widetilde{U}_{jk} \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0, 1)$. This part is the difference from the one used in Appendix E.

(2) Observed covariates with network dependence: $X_i = (X_{i1}, X_{i2}, X_{i3})$ where, for $k \in \{1, 2, 3\}$, $X_{ik} = \sum_{s \geq 0} \zeta^s \sum_{j \in \mathcal{N}(i;s)} \widetilde{X}_{jk} / |\mathcal{N}(i; s)|$, $\zeta = 0.8$, and $\widetilde{X}_{jk} \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0, 1)$. This part is the same as the one used in Appendix E.

(3) Observed auxiliary variable: $C_i = U_{i1} + \beta_{UC} U_{i2} + \beta_c^\top X_i + \epsilon_{i0}$ where $\epsilon_{i0} \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0, 1)$ and $\beta_c = (0.05, 0.05, 0.05)$. The part of U_{i2} is the difference from the one used in Appendix E.

(4) Focal behavior at the baseline: $Y_{i1} = U_{i1} + \beta_{UY1} U_{i2} + 0.05 C_i + \beta_1^\top X_i + \epsilon_{i1}$ where $\epsilon_{i1} \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0, 1)$ and $\beta_1 = (-1, -1, -1)$. The part of U_{i2} is the difference from the one used in Appendix E.

(5) Focal behavior at the follow-up: $Y_{i2} = \tau A_i + 0.2 Y_{i1} + 3 U_{i1} + \beta_{UY2} U_{i2} + 0.05 C_i + \beta_2^\top X_i + \epsilon_{i2}$ where $\epsilon_{i2} \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0, 1)$, and $\beta_2 = (-1, -1, -1)$. The treatment variable A_i is defined as $A_i = \sum_{j \in \mathcal{N}(i;1)} Y_{j1} / |\mathcal{N}(i; 1)|$. The part of U_{i2} is the difference from the one used in Appendix E.

The main difference is in (1) where we allow for two separate unmeasured confounders U_{i1} and U_{i2}. Yet, we use $W_i = C_i$ as NCO, and $Z_i = \sum_{j \in \mathcal{N}(i;1)} C_j / |\mathcal{N}(i; 1)|$ as NCE. Therefore, the number of unmeasured confounders is larger than the number of NCO, and this violates the completeness condition (Assumption 5). In this case, an outcome confounding bridge does not exist and Assumption 2.3 is violated.

We consider three different levels of violation using parameters $(\beta_{UC}, \beta_{UY1}, \beta_{UY2})$. We define "Small", "Moderate", and "Large" violations as follows.

- "Small": $\beta_{UC} = 0.1, \beta_{UY1} = \beta_{UY2} = 0.005$
- "Moderate": $\beta_{UC} = 0.25, \beta_{UY1} = \beta_{UY2} = 0.0125$
- "Large": $\beta_{UC} = 0.5, \beta_{UY1} = \beta_{UY2} = 0.025$

We fix sample size to be 1000, and we generate 2000 simulations to evaluate estimators in terms of the absolute mean bias, the standard error (computed as the standard deviation of point estimates across simulations), the root mean squared error (RMSE), and coverage of 95% confidence intervals based on the network HAC variance estimator. We standardize the first three quantities by the true ACPE to ease interpretation.

Results. Table A3 summarizes the results of the simulation study. Our proposed DNC estimator has small bias and has reasonable coverage when the violation is "small." However, as we expect, the larger is the violation, the bias is larger and coverage performance becomes poorer.

Simulation Design		**DNC**				
Network	Violation	Bias	Standard Error	RMSE	Coverage (Analytical)	Coverage (Default)
SW-4	Small	0.02	0.41	0.41	0.94	0.94
	Moderate	0.20	0.38	0.43	0.89	0.89
	Large	0.75	0.38	0.84	0.39	0.39
SW-8	Small	0.07	0.56	0.56	0.95	0.95
	Moderate	0.11	0.54	0.56	0.91	0.91
	Large	0.58	0.48	0.75	0.68	0.68
Add Health	Small	0.04	0.44	0.45	0.95	0.94
	Moderate	0.17	0.44	0.47	0.89	0.89
	Large	0.70	0.41	0.81	0.49	0.49

Table A3: Operating Characteristics when the Confounding Bridge Assumption and the Completeness Condition are Violated.

Note: We consider three different levels of violation: "Small", "Moderate", and "Large" (see above for their definitions). We examine the same three different networks; the small world network model with the average degree of four (SW-4) and eight (SW-8), and the Add Health network. For the DNC estimator, we report the absolute mean bias, the standard error, the RMSE, and coverage of the 95% confidence intervals based on the analytical bandwidth and the default bandwidth. The absolute mean bias, the standard error, and the RMSE for both estimators are standardized by the true ACPE.

E.4 Association of W between U

In this section, we consider how the performance of the DNC estimator changes as we change the association between NCO W and unobserved confounder U.

For the data generating process, we change only one aspect of the data generating process in Appendix E.1. In particular, we vary parameter $\beta_{UC} \in \{0.0, 0.5, 1.0, 1.5\}$ in the following data generating process. In Appendix E.1, we used $\beta_{UC} = 1.0$.

(3) Observed auxiliary variable: $C_i = \beta_{UC} U_i + \beta_c^\top X_i + \epsilon_{i0}$ where $\epsilon_{i0} \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0, 1)$.

Importantly, when $\beta_{UC} = 0$, NCO $W_i = C_i$ is unrelated to unobserved confounder U, and thus, the outcome confounding bridge assumption (Assumption 2.3) is violated. When the association between W and U is small, even if the outcome confounding bridge is identified, the performance of the estimator might be poor.

Results. We generate 2000 simulations and evaluate estimators in terms of absolute mean bias, standard error (computed as standard deviation of point estimates across simulations), root mean squared error (RMSE), and coverage of 95% confidence intervals based on the network HAC variance estimator. We set the sample size to be 1000, and we standardize the first three quantities by the true ACPE to ease interpretation.

Table A4 summarizes the results of the simulation study. Several points are worth noting. First, when $\beta_{UC} = 0$, because the required causal assumption is violated, the DNC estimator is not consistent and does not have correct coverage. Second, as long as $\beta_{UC} > 0$, the ACPE is identified, and thus, the simulation results show that the DNC estimator is consistent and has correct coverage. Importantly, even when $\beta_{UC} = 0.5$, we find that the DNC estimator maintains correct coverage. However, it is important to see that standard errors are large and the RMSEs are larger than those of the OLS estimator. This is similar to the phenomena of weak instrumental variable (i.e., identified but have poor finite sample performance), and this is what we can call the weak negative control problem. Finally, we see that when the association between U and W (i.e., β_{UC}) becomes stronger, the performance of the DNC estimator improves dramatically. While maintaining correct coverage, the DNC estimator now has much smaller standard errors and RMSEs. This simulation study highlights the importance of choosing NCO W that is strongly associated with unobserved confounder U.

Simulation Design				DNC					OLS			
β_{UC}	Network	Average degree	Density	Bias	Standard Error	RMSE	Coverage (Analytical)	Coverage (Default)	Bias	Standard Error	RMSE	Coverage
0.0	SW-4	4.00	0.40	3.26	2.38	4.04	0.24	0.24	1.48	0.22	1.50	0.00
0.0	SW-8	8.00	0.80	2.72	4.86	5.57	0.30	0.30	1.30	0.29	1.33	0.00
0.0	Add Health	4.82	0.48	3.24	3.87	5.04	0.25	0.24	1.40	0.26	1.42	0.00
0.5	SW-4	4.00	0.40	0.81	13.74	13.77	0.94	0.94	1.32	0.20	1.33	0.00
0.5	SW-8	8.00	0.80	1.61	21.53	21.59	0.94	0.94	1.16	0.28	1.19	0.01
0.5	Add Health	4.80	0.48	0.39	10.82	10.83	0.94	0.94	1.26	0.24	1.28	0.00
1.0	SW-4	4.00	0.40	0.06	0.41	0.42	0.95	0.95	0.99	0.18	1.01	0.00
1.0	SW-8	8.00	0.80	0.09	0.56	0.57	0.95	0.95	0.87	0.24	0.90	0.04
1.0	Add Health	4.87	0.49	0.06	0.45	0.45	0.95	0.94	0.94	0.20	0.96	0.00
1.5	SW-4	4.00	0.40	0.02	0.26	0.26	0.95	0.95	0.71	0.15	0.73	0.00
1.5	SW-8	8.00	0.80	0.04	0.34	0.35	0.95	0.95	0.62	0.21	0.66	0.14
1.5	Add Health	4.82	0.48	0.03	0.29	0.29	0.94	0.93	0.67	0.18	0.69	0.03

Table A4: Operating Characteristics when the Association between U and W Changes.

Note: We consider four different levels of association: $\beta_{UC} \in \{0.0, 0.5, 1.0, 1.5\}$. We examine the same three different networks; the small world network model with the average degree of four (SW-4) and eight (SW-8), and the Add Health network. For the DNC estimator, we report the absolute mean bias, the standard error, the RMSE, and coverage of the 95% confidence intervals based on the analytical bandwidth and the default bandwidth. The absolute mean bias, the standard error, and the RMSE for both estimators are standardized by the true ACPE.

E.5 Binary Outcomes

In this section, we consider a binary outcome to show that our nonparametric identification results can accommodate different types of outcomes. Given three different networks as we introduced in Appendix E.1, we simulate data with the following data-generating mechanism: For units $i = 1, \ldots, n$,

(1) Unobserved confounder with network dependence: $U_i = 0.5 \sum_{s \geq 0} \zeta^s \sum_{j \in \mathcal{N}(i;s)} \widetilde{U}_j / |\mathcal{N}(i;s)|$ where $\zeta = 0.8$ and $\widetilde{U}_j \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0, 0.5)$. This data generating process for a network-dependent variable follows a simulation setup of Kojevnikov *et al.* (2020).

(2) Observed auxiliary variable: $C_i = 2U_i + \epsilon_{i0}$ where $\epsilon_{i0} \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0, 0.1)$.

(3) Focal behavior at the baseline: Y_{i1} follows a Bernoulli distribution with $\Pr(Y_{i1} = 1 \mid U_i) = \Phi(0.2 + U_i)$ where $\Phi(\cdot)$ is the CDF of the standard normal distribution.

(4) Negative Control Outcome: $W_i = Z_i + A_i + \epsilon_{i1}$ where $\epsilon_{i1} \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(0, 0.5)$, $Z_i = \sum_{j \in \mathcal{N}(i;1)} C_j / |\mathcal{N}(i;1)|$, and $A_i = \sum_{j \in \mathcal{N}(i;1)} Y_{j1} / |\mathcal{N}(i;1)|$. This data generating process follows a result for the binary outcome in Tchetgen Tchetgen *et al.* (2020a).

(5) Focal behavior at the follow-up: Y_{i2} follows a Bernoulli distribution with $\Pr(Y_{i2} = 1 \mid \widetilde{W}_i, A_i) = \Phi(\beta_1 \widetilde{W}_i + \beta_2 A_i)$ where $\widetilde{W}_i = Z_i - 1.2A_i$ and $\Phi(\cdot)$ is the CDF of the standard normal distribution. As before, $Z_i = \sum_{j \in \mathcal{N}(i;1)} C_j / |\mathcal{N}(i;1)|$, and $A_i = \sum_{j \in \mathcal{N}(i;1)} Y_{j1} / |\mathcal{N}(i;1)|$. We follow a result for the binary outcome in Tchetgen Tchetgen *et al.* (2020a), and thus, we set $(\beta_1, \beta_2) = (\eta_1, \eta_2) \times \phi$ where $\phi = (1 + (0.5\eta_1)^2)^{-1/2}$.

The above models imply that the oucome confounding bridge is $h(W_i, A_i) = \Phi(\eta_1 W_i + \eta_2 A_i)$. In this simulation, we set $(\eta_1, \eta_2) = (0.3, 0.6)$ and, the true ACPE is about 0.23, while the exact value of the true ACPE depends on simulation settings. The key difference from other simulations is that the outcome variables Y_{i1} and Y_{i2} are both binary variables: in this simulation, we use the probit link. To have a clear simulation, we follow Tchetgen Tchetgen *et al.* (2020a) and use the closed form solution for the outcome confounding bridge in the case of binary outcomes. Please see Appendix of Tchetgen Tchetgen *et al.* (2020a) for more details about how to generate simulations for binary outcomes in the double negative control setting.

We evaluate the performance of the proposed DNC estimator and the network HAC variance estimator. For reference, we also report the usual regression estimator where we regress Y_{i2} on the treatment variable A_i and a set of observed variables (Y_{i1}, C_i) in the probit regression. This estimator is consistent only under conditional ignorability, which is violated due to unmeasured network confounder under this simulation setup. Thus, this probit regression estimator quantifies the amount of network confounding that the DNC estimator has to correct for.

Results. Table A5 summarizes the results of the simulation study. Our proposed DNC estimator remains stable with relatively small bias across all scenarios and has correct coverage across simulation settings. In contrast, the probit regression has a large bias and incorrect coverage across simulation settings.

Simulation Design				DNC				Probit			
	Network	Average degree	Density	Bias	Standard Error	RMSE	Coverage	Bias	Standard Error	RMSE	Coverage
SW-4	500	4.00	0.80	0.00	0.42	0.42	0.95	0.49	0.36	0.61	0.74
	1000	4.00	0.40	0.00	0.30	0.30	0.95	0.49	0.26	0.55	0.54
	2000	4.00	0.20	0.00	0.21	0.21	0.94	0.48	0.19	0.52	0.27
	4000	4.00	0.10	0.00	0.15	0.15	0.94	0.49	0.13	0.50	0.04
SW-8	500	8.00	1.60	0.02	0.61	0.61	0.95	0.46	0.53	0.70	0.85
	1000	8.00	0.80	0.02	0.43	0.43	0.95	0.48	0.37	0.60	0.76
	2000	8.00	0.40	0.01	0.30	0.30	0.95	0.46	0.27	0.54	0.59
	4000	8.00	0.20	0.02	0.22	0.22	0.95	0.48	0.19	0.51	0.29
Add Health	500	3.82	0.77	0.01	0.45	0.45	0.93	0.52	0.36	0.63	0.68
	1000	4.80	0.48	0.01	0.33	0.33	0.94	0.51	0.27	0.58	0.51
	2000	5.69	0.28	0.01	0.24	0.24	0.94	0.50	0.20	0.54	0.29
	4000	5.95	0.15	0.01	0.16	0.16	0.95	0.49	0.14	0.51	0.06

Table A5: Operating Characteristics with Binary Outcome Variables.

Note: We examine the same three different networks; the small world network model with the average degree of four (SW-4) and eight (SW-8), and the Add Health network. For the DNC estimator, we report the absolute mean bias, the standard error, the RMSE, and coverage of the 95% confidence intervals. The absolute mean bias, the standard error, and the RMSE for both estimators are standardized by the true ACPE.

F Extensions

F.1 Higher-order Peer Effects

Following standard causal peer effect literature, we have focused on the causal effect from peers as the causal estimand of primary interest (the ACPE defined in equation (8)). It is important to emphasize that all results in Section 3 do not rule out causal effects from higher-order peers (e.g., peers-of-peers). If they exist, one can simply adjust for focal behaviors of higher-order peers as observed pre-treatment covariates X_i. We have considered such higher-order peer effects as nuisance when studying identification and estimation of the ACPE. In this section, we clarify that the proposed double negative control approach can also be used for identification and estimation of higher-order causal peer effects as well.

The study of such higher-order peer effects can be important for several reasons. First, in some applications, focal behaviors might be directly affected by higher-order peers even if peers might not change their behaviors. For example, information can diffuse from higher-order peers even if there is no behavioral change among peers. Second, estimation of higher-order peer effects can account for some forms of misspecification of underlying networks. It is possible that observed network and time might not perfectly match the underlying process through which units causally affect peers. For example, it is possible that units affect peers faster, and units can affect their peers-of-peers within one observed time interval. Additionally, the observed network might miss some ties between units, and thus, two units with the observed shortest distance of two might in fact be connected directly in the underlying true network. In such cases, we want to estimate causal effects from peers and peers-of-peers jointly.

One can explicitly include focal behaviors of higher-order peers into the potential outcome. Suppose we are interested in causal effects from all units within network distance $s^\dagger$. We define a vector of the treatment variable $\tilde{A}_i = (A_{i1}, \dots, A_{is^\dagger})$ where $A_{is} = \phi(\{Y_{j1} : j \in \mathcal{N}(i;s)\}) \in \mathbb{R}$, $s \in \{1, \dots, s^\dagger\}$, and function ϕ is specified by a researcher based on subject matter knowledge. When $s^\dagger = 1$, this setup reduces to the one in Section 3. The potential outcome $Y_{i2}(\tilde{a})$ is defined as the outcome that would realize when the treatment vector is set to $\tilde{A}_i = \tilde{a}$. We can then define the higher-order ACPE as

$$\tau(\tilde{a}, \tilde{a}') := \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left\{Y_{i2}(\tilde{a}) - Y_{i2}(\tilde{a}')\right\} \tag{A.40}$$

where $\tilde{a}, \tilde{a}' \in \widetilde{\mathcal{A}}$ where $\widetilde{\mathcal{A}}$ is the support of $\tilde{A}$. For example, $\tau((a_1, a_2), (a_1, a_2'))$ captures the second-order peer effect by fixing the treatment value of peers and changing the treatment value of peers-of-peers. Importantly, while this setup considers up to the $s^\dagger$-th order peer effects as the causal estimand, this does not assume the absence of causal effects from peers at distance more than $s^\dagger$. We only view them as nuisance.

We can straightforwardly generalize Assumption 2 and Theorem 2 to this setting of higher-order peer effects by replacing A_i with $\tilde{A}_i$. The selection of negative controls can also proceed in similar fashion. A plausible candidate is again an auxiliary variable C_i that (a) does not affect network relationships and (b) does not affect variables of other units. For example, even if we add the second-order peer effects to Figure 3.(i) (i.e., a causal arrow from Y_{41} to Y_{22}), the original choice of negative controls — C_2 as the NCO and $\{C_1, C_3, C_4\}$ as the NCEs — remains valid.

Another candidate for negative controls is the focal behavior itself. For example, if one were to add the second-order peer effects to Figure 3.(ii) (i.e., an causal arrow from Y_{41} to Y_{22}), the original choice of NCO Y_{21} would remain valid, while the original choice of NCEs $\{Y_{41}, Y_{42}\}$ would no longer be valid. If all third-order peer effects are absent, focal behaviors of third-order peers would be a plausible candidate for the NCEs. In summary, while the specific choice of negative controls need to be adjusted when examining higher-order ACPE, the two primary ways of sfxrelecting negative controls we discussed in Section 3 continue to be useful.

Finally, estimation and inference can proceed as in Theorems 3 and 4 can be extended by replacing A_i with $\tilde{A}_i$ in the definition of $L_{n,i}$.

F.2 Misspecification

For the sake of clarity in the exposition, we restricted presentation of all main results to the causal effect from peers. While it is impossible to account for all forms of misspecification of exposure mapping ϕ, we consider one general form. Specifically, we now suppose that higher-order peers (e.g., peers-of-peers) can have causal effects and yet researchers do not know exactly how far peer effects exist. Formally, suppose causal peer effects are nonzero from all units within network distance s^*. We call it the *causal* network distance as causal peer effects exist for all units within this network distance s^*. We assume in this section that analysts do not know the exact value of the causal network distance s^*, which is a common scenario in practice.

This potential misspecification is important because it is possible that the observed network does not perfectly match the underlying process through which units causally affect network peers. For example, units can affect higher-order peers possibly because the observed network might miss some ties between units, and thus, two units with the observed shortest distance of two might in fact be connected directly in the underlying true network.

When researchers know the exact value of the causal network distance s^*, they can use results in Appendix F.1 to identify higher-order peer effects directly. Here, we consider how to test the potential misspecification using double negative controls.

F.2.1 Specification Test

We assume that, while researchers do not know the exact value of the causal network distance s^*, they know its upper bound s^*_0 where $s^* \leq s^*_0$. This is usually a plausible assumption in practice, and more importantly, by picking a larger value of s^*_0 (e.g., 5), researchers can almost always satisfy this requirement, while the proposed test below might be more conservative.

Importantly, as we discussed in Appendix F.1, when researchers use an auxiliary variable C_i that (a) does not affect network relationships and (b) does not affect variables of other units for double negative controls, researchers do not need to change the choice of negative controls. By including peers up to s^*_0 as discussed in Appendix F.1, researchers can directly test whether causal peer effects from high-order peer effects are nonzero. In practice, when researchers find that causal peer effects are nonzero even for peers at network distance s_0^*, it is likely that the chosen value of s_0^* is small, so we recommend checking a larger value of s_0^*.

It is also interesting to consider cases when researchers use the focal behavior as candidates for negative controls. In this case, the focal behaviors of the second-order peers $\{Y_{jt} : j \in \mathcal{N}(i;2), t \in \{1,2\}\}$ (those we discussed in Section 3.3.2) are no longer valid negative control exposures as they might affect ego's outcomes Y_{i2} or be affected by the negative control outcome Y_{i1} (ego's focal behavior at the baseline).

Under this setup, we can use the same negative control outcome Y_{i1}, but we change the choice of the negative control exposures.

$$Z_i^{s_0^*} \equiv \{Y_{jt} : j \in \mathcal{N}_n(i;\ell),\ t \in \{1,2\} \ \text{ where } \ \ell > s_0^*,\ \ell \in \mathbb{Z}\}, \tag{A.41}$$

which are the focal behaviors of neighbors that are at the distance more than s_0^* from unit i.

We can then construct a specification test. Importantly, $Z_i^s \subseteq Z_i^{s'}$ where $s > s'$. This nested structure implies that, under the assumption that the causal network distance is at most s_0^*, we can test whether the true causal network distance is equal to or smaller than $\tilde{s}$, which is smaller than s_0^*. For example, if we assume that the true causal network distance is at most 2, we can test whether there is any misspecification or not, i.e., $s^* = 1$. This can be done within the GMM framework as the J-test (Hansen, 1982).

Under the assumption that the true causal network distance is at most s_0^*, the following J statistic follows an asymptotic chi-square distribution when the true network causal distance is also equal to or smaller than $\tilde{s}$.

$$J_{\tilde{s},s_0^*} \equiv n \left\{ \frac{1}{n} \sum_{i=1}^{n} m(L_{n,i}; \widehat{\theta}_n^{\tilde{s},s_0^*}) \right\}^{\top} \Omega \left\{ \frac{1}{n} \sum_{i=1}^{n} m(L_{n,i}; \widehat{\theta}_n^{\tilde{s},s_0^*}) \right\} \xrightarrow{d} \chi^2_{k-p}$$

where k is the number of moments, p is the number of parameters. $m(\cdot)$ is the moment function and $\widehat{\theta}_n^{\tilde{s},s_0^*}$ is the GMM estimator with the following negative control exposures:

$$Z_i^{\tilde{s},s_0^*} \equiv \{Y_{jt} : j \in \mathcal{N}_n(i;\ell),\ t \in \{1,2\} \ \text{ where } \ \tilde{s}+1 \leq \ell \leq s_0^*+1,\ \ell \in \mathbb{Z}\}. \tag{A.42}$$

F.3 Identification of Spillover Effects in Observational Studies

In this paper, we have focused on the causal peer effect, which is defined as the causal effect of peers' focal behaviors on an ego's behavior. In our application, we studied the causal peer effect on the GPA in a friendship network. peers' GPA served as a treatment variable. The spillover effect is a related but different causal quantity of interest. In the literature of spillover effects, we have a treatment variable defined separately from the focal behavior of interest. For example, one might be interested in the causal effect of a scholarship, which is separately defined from the GPA, the outcome of interest. In this example, the spillover effect of a scholarship is the causal effect of whether peers receive scholarships on the GPA of an ego. See Ogburn and VanderWeele (2014) for further discussion on the difference between causal peer effects and spillover effects.

The vast majority of the spillover effect literature has focused on randomized experiments where identification of the spillover effect stems from randomized treatments (e.g., Sobel, 2006; Hudgens and Halloran, 2008; Tchetgen Tchetgen and VanderWeele, 2010; Aronow and Samii, 2017). See Halloran and Hudgens (2016) for a review. Recent papers examine identification and estimation of spillover effects in observational studies under conditional ignorability (e.g., Ogburn *et al.*, 2017; Tchetgen Tchetgen *et al.*, 2020b; Forastiere *et al.*, 2020). In the absence of randomization in observational studies, conditional ignorability assumption might be violated due to unmeasured network confounding. Researchers can use our proposed double negative control approach for identification and estimation of spillover effects as well. The key difference from the main results of Section 3 is in selection of negative controls as the definition of the treatment differs. Regardless, we can naturally generalize Assumptions 2 and 3 to prove identification and asymptotic properties for the spillover effect estimation analogous to Theorems 2 and 3.

References

Ai, C. and Chen, X. (2003). Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions. *Econometrica*, **71**(6), 1795–1843.

Andrews, D. W. (2017). Examples of L2-complete and Boundedly-complete Distributions. *Journal of Econometrics*, **199**(2), 213–220.

Aronow, P. M. and Samii, C. (2017). Estimating Average Causal Effects Under General Interference, With Application to A Social Network Experiment. *Annals of Applied Statistics*.

Carrasco, M., Florens, J.-P., and Renault, E. (2007). Linear Inverse Problems in Structural Econometrics Estimation based on Spectral Decomposition and Regularization. *Handbook of Econometrics*, **6**, 5633–5751.

Chernozhukov, V., Imbens, G. W., and Newey, W. K. (2007). Instrumental Variable Estimation of Nonseparable Models. *Journal of Econometrics*, **139**(1), 4–14.

Cobzaru, R., Welsch, R., Finkelstein, S., Ng, K., and Shahn, Z. (2022). Bias Formulas for Violations of Proximal Identification Assumptions. *arXiv preprint arXiv:2208.00105*.

Darolles, S., Fan, Y., Florens, J.-P., and Renault, E. (2011). Nonparametric Instrumental Regression. *Econometrica*, **79**(5), 1541–1565.

Deaner, B. (2018). Proxy Controls and Panel Data. *arXiv preprint arXiv:1810.00283*.

Egami, N. (2018). Identification of Causal Diffusion Effects Under Structural Stationarity. arXiv preprint available at `https://arxiv.org/abs/1810.07858`.

Forastiere, L., Airoldi, E. M., and Mealli, F. (2020). Identification and Estimation of Treatment and Interference Effects in Observational Studies On Networks. *Journal of the American Statistical Association*, pages 1–18.

Freyberger, J. (2018). Non-Parametric Panel Data Models With Interactive Fixed Effects. *The Review of Economic Studies*, **85**(3), 1824–1851.

Granger, C. W. (1988). Some Recent Development in A Concept of Causality. *Journal of Econometrics*, **39**(1-2), 199–211.

Halloran, M. E. and Hudgens, M. G. (2016). Dependent Happenings: a Recent Methodological Review. *Current Epidemiology Reports*, **3**(4), 297–305.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, pages 1029–1054.

Hansen, L. P., Heaton, J., and Yaron, A. (1996). Finite-Sample Properties of Some Alternative GMM Estimators. *Journal of Business & Economic Statistics*, **14**(3), 262–280.

Hu, Y. and Schennach, S. M. (2008). Instrumental Variable Treatment of Nonclassical Measurement Error Models. *Econometrica*, **76**(1), 195–216.

Hu, Y. and Shum, M. (2012). Nonparametric Identification of Dynamic Models with Unobserved State Variables. *Journal of Econometrics*, **171**(1), 32–44.

Hudgens, M. G. and Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, **103**(482), 832–842.

Kallus, N., Mao, X., and Uehara, M. (2021). Causal Inference Under Unmeasured Confounding With Negative Controls: A Minimax Learning Approach. *arXiv preprint arXiv:2103.14029*.

Kojevnikov, D., Marmer, V., and Song, K. (2020). Limit Theorems for Network Dependent Random Variables. *Journal of Econometrics*.

Kress, R. (1989). *Linear Integral Equations*, volume 82. Springer.

Lehmann, E. L. and Scheffé, H. (2012a). Completeness, Similar Regions, and Unbiased Estimation-Part I. In *Selected Works of EL Lehmann*, pages 233–268. Springer.

Lehmann, E. L. and Scheffé, H. (2012b). Completeness, Similar Regions, and Unbiased Estimation—Part II. In *Selected Works of EL Lehmann*, pages 269–286. Springer.

Liu, L. and Tchetgen Tchetgen, E. (2020). Regression-based Negative Control of Homophily in Dyadic Peer Effect Analysis. *arXiv preprint arXiv:2002.06521*.

McFowland III, E. and Shalizi, C. R. (2021). Estimating Causal Peer Influence in Homophilous Social Networks by Inferring Latent Locations. *Journal of the American Statistical Association*, pages 1–27.

Miao, W., Shi, X., and Tchetgen Tchetgen, E. (2018a). A Confounding Bridge Approach for Double Negative Control Inference on Causal Effects. *arXiv preprint arXiv:1808.04945*.

Miao, W., Geng, Z., and Tchetgen Tchetgen, E. J. (2018b). Identifying Causal Effects With Proxy Variables of An Unmeasured Confounder. *Biometrika*, **105**(4), 987–993.

Newey, W. K. and McFadden, D. (1994). Large Sample Estimation and Hypothesis. In R. Engle and D. McFadden, editors, *Handbook of Econometrics*, pages 2112–2245. North Holland.

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, **71**(5), 1565–1578.

Ogburn, E. L. and VanderWeele, T. J. (2014). Causal Diagrams for Interference. *Statistical Science*, **29**(4), 559–578.

Ogburn, E. L., Sofrygin, O., Diaz, I., and van der Laan, M. J. (2017). Causal Inference for Social Network Data. *arXiv preprint arXiv:1705.08527*.

Severini, T. A. and Tripathi, G. (2006). Some Identification Issues in Nonparametric Linear Models with Endogenous Regressors. *Econometric Theory*, **22**(2), 258–278.

Shalizi, C. R. and Thomas, A. C. (2011). Homophily and Contagion are Generically Confounded in Observational Social Network Studies. *Sociological Methods & Research*, **40**(2), 211–239.

Shao, J. (2003). *Mathematical Statistics*. Springer Science & Business Media.

Shi, X., Miao, W., Nelson, J. C., and Tchetgen Tchetgen, E. J. (2020). Multiply Robust Causal Inference With Double-Negative Control Adjustment for Categorical Unmeasured Confounding. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **82**(2), 521–540.

Sobel, M. E. (2006). What Do Randomized Studies Of Housing Mobility Demonstrate? Causal Inference In The Face Of Interference. *Journal of the American Statistical Association*, **101**(476), 1398–1407.

Sofer, T., Richardson, D. B., Colicino, E., Schwartz, J., and Tchetgen Tchetgen, E. J. (2016). On Negative Outcome Control of Unobserved Confounding as a Generalization of Difference-in-Differences. *Statistical Science*, **31**(3), 348.

Tchetgen Tchetgen, E. J. and VanderWeele, T. J. (2010). On Causal Inference in The Presence of Interference. *Statistical Methods in Medical Research*, **21**(1), 55–75.

Tchetgen Tchetgen, E. J., Ying, A., Cui, Y., Shi, X., and Miao, W. (2020a). An Introduction to Proximal Causal Learning. *arXiv preprint arXiv:2009.10982*.

Tchetgen Tchetgen, E. J., Fulcher, I., and Shpitser, I. (2020b). Auto-G-Computation of Causal Effects on a Network. *Journal of the American Statistical Association*.