Supplementary Material: A Proxy Variable View of Shared Confounding

A. Proof of Theorem 1

Proof. The proof of Theorem 1 relies on two observations. The first observation starts with the integral equation we solve:

$$P(y \mid a_{\mathcal{C}}, f(a_{\mathcal{N}})) = \int h(y, a_{\mathcal{C}}, a_{\mathcal{X}}) P(a_{\mathcal{X}} \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}))\, \mathrm{d}a_{\mathcal{X}} \tag{20}$$

$$= \int \int h(y, a_{\mathcal{C}}, a_{\mathcal{X}}) P(a_{\mathcal{X}} \mid u) P(u \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}))\, \mathrm{d}a_{\mathcal{X}}\, \mathrm{d}u. \tag{21}$$

The first equality is due to Eq. 3. The second equality is due to the conditional independence implied by Figure 1a: $A_{\mathcal{X}} \perp A_{\mathcal{C}}, f(a_{\mathcal{N}}) \mid U$.

The second observation relies on the null proxy:

$$P(y \mid a_{\mathcal{C}}, f(a_{\mathcal{N}})) = \int P(y \mid u, a_{\mathcal{C}}, f(a_{\mathcal{N}})) P(u \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}))\, \mathrm{d}u \tag{22}$$

$$= \int P(y \mid u, a_{\mathcal{C}}) P(u \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}))\, \mathrm{d}u. \tag{23}$$

The first equality is due to the definition of conditional probability. The second equality is due to the second part of Assumption 1, which implies $Y \perp f(a_{\mathcal{N}}) \mid U, A_{\mathcal{C}}$. The reason is that

$$P(y \mid u, a_{\mathcal{C}}, f(a_{\mathcal{N}})) = \int P(y \mid u, a_{\mathcal{C}}, a_{\mathcal{X}}, f(a_{\mathcal{N}})) P(a_{\mathcal{X}} \mid u, a_{\mathcal{C}}, f(a_{\mathcal{N}}))\, \mathrm{d}a_{\mathcal{X}} \tag{24}$$

$$= \int P(y \mid u, a_{\mathcal{C}}, a_{\mathcal{X}}) P(a_{\mathcal{X}} \mid u, a_{\mathcal{C}})\, \mathrm{d}a_{\mathcal{X}} \tag{25}$$

$$= P(y \mid u, a_{\mathcal{C}}). \tag{26}$$

In fact, it is sufficient to assume $Y \perp f(a_{\mathcal{N}}) \mid U, A_{\mathcal{C}}$ instead of $Y \perp f(a_{\mathcal{N}}) \mid U, A_{\mathcal{C}}, A_{\mathcal{X}}$ in Theorem 1. However, the latter is easier to check and interpret.

Comparing Eq. 21 and Eq. 23 gives

$$\int \left[P(y \mid u, a_{\mathcal{C}}) - \int h(y, a_{\mathcal{C}}, a_{\mathcal{X}}) P(a_{\mathcal{X}} \mid u)\, \mathrm{d}a_{\mathcal{X}} \right] \times P(u \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}))\, \mathrm{d}u = 0, \tag{27}$$

which, by the completeness condition in Assumption 1.2, implies

$$P(y \mid u, a_{\mathcal{C}}) = \int h(y, a_{\mathcal{C}}, a_{\mathcal{X}}) P(a_{\mathcal{X}} \mid u)\, \mathrm{d}a_{\mathcal{X}}. \tag{28}$$

Eq. 28 leads to identification:

$$P(y \mid \mathrm{do}(a_{\mathcal{C}})) = \int \int h(y, a_{\mathcal{C}}, a_{\mathcal{X}}) P(a_{\mathcal{X}} \mid u)\, \mathrm{d}a_{\mathcal{X}} P(u)\, \mathrm{d}u \tag{29}$$

$$= \int h(y, a_{\mathcal{C}}, a_{\mathcal{X}}) P(a_{\mathcal{X}})\, \mathrm{d}a_{\mathcal{X}}. \tag{30}$$

Consider the special case of a single treatment as in Figure 1b. Let $a_{\mathcal{C}} = \{A_1\}$, $a_{\mathcal{X}} = \{X\}$, $a_{\mathcal{N}} = N$, and $f(a_{\mathcal{N}}) = N$. The above proof reduces to the identification proof for proxy variables (Theorem 1 of Miao et al. (2018)). □

B. Examples of Assumption 1

As an example, if the structural equation writes

$$Y = g(A_1 + A_2, A_3, \ldots, A_m, U, \epsilon),$$

where $\epsilon \perp U, A_1, \ldots, A_m$, then Assumption 1.1 is satisfied if A_1 and A_2 are identically Gaussian: $A_{\mathcal{N}} = (A_1, A_2)$ and $f(A_{\mathcal{N}}) = A_1 - A_2$ satisfies

$$A_1 - A_2 \perp Y \,|\, U, A_3, \ldots, A_m.$$

If A_1 and A_2 are both Gaussian but not identically distributed, then $f(A_{\mathcal{N}}) = \alpha_1 A_1 - \alpha_2 A_2$ would satisfy

$$\alpha_1 A_1 - \alpha_2 A_2 \perp Y \,|\, U, A_3, \ldots, A_m,$$

for some constant α_1 and α_2.

Similarly, if the structural equation writes

$$Y = g(A_1 \times A_2, A_3, \ldots, A_m, U, \epsilon),$$

where $\epsilon \perp U, A_1, \ldots, A_m$, then Assumption 1.1 is satisfied if A_1 and A_2 are identically log-normal: $A_{\mathcal{N}} = (A_1, A_2)$ and $f(A_{\mathcal{N}}) = A_1/A_2$ satisfies

$$A_1/A_2 \perp Y \,|\, U, A_3, \ldots, A_m.$$

As a final example, if the structural equation writes

$$Y = g(A_1 \&\& A_2, A_3, \ldots, A_m, U, \epsilon),$$

where $\epsilon \perp U, A_1, \ldots, A_m$ and A_1, A_2 are both binary, then Assumption 1.1 is satisfied: $A_{\mathcal{N}} = (A_1, A_2)$ and $f(A_{\mathcal{N}}) = A_1$ XOR A_2 satisfies

$$A_1 \text{ XOR } A_2 \perp Y \,|\, U, A_3, \ldots, A_m.$$

C. Proof of Theorem 2

Proof. Assumption 2.2 guarantees the existence of some function $\hat{h}$ such that

$$\hat{P}(y \,|\, a_{\mathcal{C}}, \hat{z}) = \int \hat{h}(y, a_{\mathcal{C}}, a_{\mathcal{X}}) \hat{P}(a_{\mathcal{X}} \,|\, \hat{z}) \, \mathrm{d}a_{\mathcal{X}} \tag{31}$$

under weak regularity conditions. (We will discuss the reason in Appendix D.)

We first claim that $\hat{h}(y, a_{\mathcal{C}}, a_{\mathcal{X}})$ solves

$$P(y \,|\, a_{\mathcal{C}}, f(a_{\mathcal{N}})) = \int \hat{h}(y, a_{\mathcal{C}}, a_{\mathcal{X}}) P(a_{\mathcal{X}} \,|\, a_{\mathcal{C}}, f(a_{\mathcal{N}})) \, \mathrm{d}a_{\mathcal{X}}. \tag{32}$$

Given this claim (Eq. 77), we have

$$\begin{aligned}
&\hat{P}(y \,|\, \mathrm{do}(a_{\mathcal{C}})) \\
=& \int \hat{P}(y \,|\, \hat{z}, a_{\mathcal{C}}) \hat{P}(\hat{z}) \, \mathrm{d}\hat{z} \\
=& \int \hat{h}(y, a_{\mathcal{C}}, a_{\mathcal{X}}) \hat{P}(a_{\mathcal{X}} \,|\, \hat{z}) \, \mathrm{d}a_{\mathcal{X}} \hat{P}(\hat{z}) \, \mathrm{d}\hat{z} \\
=& \int \hat{h}(y, a_{\mathcal{C}}, a_{\mathcal{X}}) P(a_{\mathcal{X}}) \, \mathrm{d}a_{\mathcal{X}} \\
=& P(y \,|\, \mathrm{do}(a_{\mathcal{C}})),
\end{aligned}$$

which proves the theorem. The first equality is due to Eq. 6; the second is due to Eq. 77; the third is due to the deconfounder estimate being consistent with the observed data distribution by construction; the fourth is due to the above claim (Eq. 77) and Theorem 1.

We next prove the claim (Eq. 77). Start with the right side of the equality.

$$\int \hat{h}(y, a_{\mathcal{C}}, a_{\mathcal{X}}) P(a_{\mathcal{X}} \,|\, a_{\mathcal{C}}, f(a_{\mathcal{N}})) \, \mathrm{d}a_{\mathcal{X}}$$

$$
\begin{aligned}
&= \int \int \hat{h}(y, a_{\mathcal{C}}, a_{\mathcal{X}}) \hat{P}(a_{\mathcal{X}} \,|\, \hat{z}) \hat{P}(\hat{z} \,|\, a_{\mathcal{C}}, f(a_{\mathcal{N}})) \, \mathrm{d}a_{\mathcal{X}} \, \mathrm{d}\hat{z} \\
&= \int \hat{P}(y \,|\, a_{\mathcal{C}}, \hat{z}) \hat{P}(\hat{z} \,|\, a_{\mathcal{C}}, f(a_{\mathcal{N}})) \, \mathrm{d}\hat{z} \\
&= P(y \,|\, a_{\mathcal{C}}, f(a_{\mathcal{N}})),
\end{aligned}
$$

which establishes the claim. The first equality is due to Eq. 4 and the deconfounder estimate being consistent with the observed data; the second is due to Eq. 31; the third is due to Assumption 2.1, which implies

$$\hat{P}(y \,|\, a_{\mathcal{C}}, f(a_{\mathcal{N}}), \hat{z}) = \hat{P}(y \,|\, a_{\mathcal{C}}, \hat{z}). \tag{33}$$

Similar to Assumption 1.1, it is sufficient to assume Eq. 33 directly. However, Assumption 2.1 is easier to check and more interpretable; it directly relates to the deconfounder outcome model.

□

D. Existence of solutions to the integral equations

Theorem 1 involves solving the integral equation

$$P(y \,|\, a_{\mathcal{C}}, f(a_{\mathcal{N}})) = \int h(y, a_{\mathcal{C}}, a_{\mathcal{X}}) P(a_{\mathcal{X}} \,|\, a_{\mathcal{C}}, f(a_{\mathcal{N}})) \, \mathrm{d}a_{\mathcal{X}}. \tag{34}$$

When does a solution exist for Eq. 34? We appeal to Proposition 1 of Miao et al. (2018).

Proposition 7. *(Proposition 1 of Miao et al. (2018)) Denote $L^2\{F(t)\}$ as the space of all square-integrable function of t with respect to a c.d.f. $F(t)$. A solution to integral equation*

$$P(y \,|\, z, x) = \int h(w, x, y) P(w \,|\, z, x) \, \mathrm{d}w \tag{35}$$

exists if

1. *the conditional distribution $P(z \,|\, w, x)$ is complete in w for all x,*
2. $\int \int P(w \,|\, z, x) P(z \,|\, w, x) \, \mathrm{d}w \, \mathrm{d}z < +\infty$,
3. $\int [P(y \,|\, z, x)]^2 P(z \,|\, x) \, \mathrm{d}z < +\infty$,
4. $\sum_{n=1}^{+\infty} | < P(y \,|\, z, x), \psi_{x,n} > |^2 < +\infty$,

where the inner product is $< g, h >= \int g(t)h(t) \, \mathrm{d}F(t)$, and $(\lambda_{x,n}, \phi_{x,n}, \psi_{x,n})_{n=1}^{\infty}$ is a singular value decomposition of the conditional expectation operator $K_x : L^2\{F(w \,|\, x)\} \to L^2\{F(z \,|\, x)\}$, $K_x(h) = \mathbb{E}\left[h(w) \,|\, z, x\right]$ for $h \in L^2\{F(w \,|\, x)\}$.

Leveraging Proposition 7, we can establish sufficient conditions for existence of a solution to Eq. 34.

Corollary 8. *A solution exist for the integral equation Eq. 34 if*

1. *the conditional distribution $P(f(a_{\mathcal{N}}) \,|\, a_{\mathcal{X}}, a_{\mathcal{C}})$ is complete in $a_{\mathcal{X}}$ for all $a_{\mathcal{C}}$,*
2. $\int \int P(a_{\mathcal{X}} \,|\, f(a_{\mathcal{N}}), a_{\mathcal{C}}) P(f(a_{\mathcal{N}}) \,|\, a_{\mathcal{X}}, a_{\mathcal{C}}) \, \mathrm{d}a_{\mathcal{X}} \, \mathrm{d}f(a_{\mathcal{N}}) < +\infty$,
3. $\int [P(y \,|\, f(a_{\mathcal{N}}), a_{\mathcal{C}})]^2 P(f(a_{\mathcal{N}}) \,|\, a_{\mathcal{C}}) \, \mathrm{d}f(a_{\mathcal{N}}) < +\infty$,
4. $\sum_{n=1}^{+\infty} | < P(y \,|\, f(a_{\mathcal{N}}), a_{\mathcal{C}}), \psi_{a_{\mathcal{C}},n} > |^2 < +\infty$,

where $\psi_{a_{\mathcal{C}},n}$ is similarly defined as a component of the singular value decomposition.

We remark that the first condition is precisely Theorem 1.3; others are weak regularity conditions.

By the same token, we can establish sufficient conditions for solution existence of Eq. 8, Eq. 14. The same argument also applies to the integral equation involved in Theorem 6:

$$\hat{P}(y \mid a_{\mathcal{C}}, \hat{z}, u_{\mathcal{C}}^{\mathrm{sng}}, s=1) = \int \hat{h}(y, a_{\mathcal{C}}, a_{\mathcal{X}}, u_{\mathcal{C}}^{\mathrm{sng}}) \hat{P}(a_{\mathcal{X}} \mid \hat{z}, u_{\mathcal{C}}^{\mathrm{sng}}, s=1) \, \mathrm{d}a_{\mathcal{X}}. \tag{36}$$

It is easy to show that the conditions described in the main text are sufficient to guarantee the existence of solutions under weak regularity conditions. We omit the details here.

E. Proof of Lemma 3

The idea of the proof is to start with the structural equations of the expanded class of causal graphs Figure 2b. Then posit the existence of a latent variable Z that renders all the treatments conditionally independent; Figure 2c features this conditional independence structure. We will quantify the information (i.e. the σ-algebra) of this latent variable Z; Z contains the information of the union of multi-treatment confounders U^{mlt}, multi-treatment null confounders W^{mlt}, and some independent error. This result lets us establish

$$P(y \mid u^{\mathrm{sng}}, u^{\mathrm{mlt}}, w^{\mathrm{mlt}}, a_1, \ldots, a_m, s=1) = P(y \mid u^{\mathrm{sng}}, z, a_1, \ldots, a_m, s=1). \tag{37}$$

We start with a generic structural equation model for multiple treatments.

$$W_k = f_{W_k}(\epsilon_{W_k}), \qquad k = 1, \ldots, K, K \geq 0, \tag{38}$$
$$U_j = f_{U_j}(\epsilon_{U_j}), \qquad j = 1, \ldots, J, J \geq 0, \tag{39}$$
$$V_l = f_{V_l}(\epsilon_{V_l}), \qquad l = 1, \ldots, L, L \geq 0, \tag{40}$$
$$A_i = f_{A_i}(W_{S_{A_i}^W}, U_{S_{A_i}^U}, \epsilon_{A_i}), \qquad i = 1, \ldots, m, m \geq 2, \tag{41}$$
$$Y = f_y(A_1, \ldots, A_m, U_1, \ldots, U_K, V_1, \ldots V_L, \epsilon_Y), \tag{42}$$

where all the errors $\epsilon_{W_k}, \epsilon_{U_j}, \epsilon_{V_l}, \epsilon_{A_i}, \epsilon_Y$ are independent. Notation wise, we note that $S_{A_i}^W \subset \{1, \ldots, K\}$ is an index set; if $S_{A_1}^W = \{1, 3, 4\}$, then $W_{S_{A_i}^W} = (W_1, W_3, W_4)$. The same notion applies to $S_{A_i}^U \subset \{1, \ldots, J\}$.

The notation in this structural equation model is consistent with the set up in Figure 2b. W_k's are null confounders; U_j's are confounders; V_l's are covariates. Moreover, $U_{S_{A_i}^U}$ indicates the set of confounders that have an arrow to both A_i and Y. $W_{S_{A_i}^W}$ indicates the set of null confounders that have an arrow to A_i; they do not have arrows to Y.

Relating to the single-treatment and multi-treatment notion, we have single-treatment null confounders as

$$W^{\mathrm{sng}} \triangleq \{W_1, \ldots, W_K\} / \bigcup_{i,j \in \{1,\ldots,m\}: i \neq j} (W_{S_{A_i}^W} \cap W_{S_{A_j}^W}). \tag{43}$$

To parse the notation above, recall that $W_{S_{A_i}^W}$ is the set of null confounders that affects A_i. $\bigcup_{i,j\in\{1,\ldots,m\}:i\neq j}(W_{S_{A_i}^W} \cap W_{S_{A_j}^W})$ describes the set of null confounders that affect at least two of the A_i's. Hence, W^{sng} denotes the set of null confounders that affect only one of the A_i's, a.k.a. single-treatment null confounders.

Before proving Lemma 3, we first prove the following lemma that quantifies the information in Z (in Figure 2c).

Lemma 9. *The random variable Z in Figure 2c "captures" all multi-treatment confounders, all multi-treatment null confounders and some independent error:*

$$\sigma(Z) = \sigma\left(\{\epsilon_Z\} \bigcup (\cup_{i,j\in\{1,\ldots,m\}:i\neq j} (W_{S_{A_i}^W} \cap W_{S_{A_j}^W}) \cup (U_{S_{A_i}^U} \cap U_{S_{A_j}^U}))\right), \tag{44}$$
$$= \sigma\left(\{\epsilon_Z\} \bigcup W^{\mathrm{mlt}} \bigcup U^{\mathrm{mlt}}\right). \tag{45}$$

where $\epsilon_Z \perp (\epsilon_Y, V_1, \ldots, V_L, \cup_{i,j\in\{1,\ldots,m\}:i\neq j}(W_{S_{A_i}^W} \cap W_{S_{A_j}^W}) \cup (U_{S_{A_i}^U} \cap U_{S_{A_j}^U}), S)$.

We can parse the notation in Lemma 9 in the same way as in Eq. 43: $\cup_{i,j\in\{1,\ldots,m\}:i\neq j}(W_{S^W_{A_i}} \cap W_{S^W_{A_j}})$ denotes the set of all multi-treatment confounders; $\cup_{i,j\in\{1,\ldots,m\}:i\neq j}(U_{S^U_{A_i}} \cap U_{S^U_{A_j}})$ denotes the set of all multi-treatment null confounders.

Proof. Without the loss of generality, we assume the compactness of representation in Eqs. 41 and 42. For any subset $\mathcal{S}$ of the random variables $\mathcal{S} \subset \{A_1, \ldots, A_m, Y\}$, we assume the σ-algebra $\sigma(\bigcap_\tau(S^W_{S_\tau}, S^U_{S_\tau}, S^V_{S_\tau}))$ is the *smallest* σ-algebra that makes all the random variables in $\mathcal{S}$ jointly independent. The assumption is made for technical convenience. We simply ensure the arrows from the W, U, V's to the A_i's do exist. In other words, all the W, U, V's "whole-heartedly" contribute to the A_i's when they appear in Eq. 41. This assumption does not limit the class of causal graphs we study.

First we show that all multi-treatment confounders and all multi-treatment null confounders are measurable with respect to the substitute confounder Z:

$$\sigma\left(\bigcup_{i,j\in\{1,\ldots,m\}:i\neq j}(W_{S^W_{A_i}} \cap W_{S^W_{A_j}}) \cup (U_{S^U_{A_i}} \cap U_{S^U_{A_j}})\right) \subset \sigma(Z). \tag{46}$$

Consider any pair of A_i and A_j. Figure 2c implies that

$$A_i \perp A_j \,|\, Z, \tag{47}$$

for $i \neq j$ and $i, j \in \{1, \ldots, M\}$. On the other hand, we have

$$A_i \perp A_j \,|\, \sigma\left((W_{S^W_{A_i}} \cap W_{S^W_{A_j}}), (U_{S^U_{A_i}} \cap U_{S^U_{A_j}})\right), \tag{48}$$

by the independence of errors assumption. Therefore, by the compactness of representation assumption, $\sigma((W_{S^W_{A_i}} \cap W_{S^W_{A_j}}), (U_{S^U_{A_i}} \cap U_{S^U_{A_j}}))$ is the smallest σ-algebra that renders A_i independent of A_j. This implies

$$\sigma\left((W_{S^W_{A_i}} \cap W_{S^W_{A_j}}), (U_{S^U_{A_i}} \cap U_{S^U_{A_j}})\right) \subset \sigma(Z). \tag{49}$$

The argument can be applied to any pair of $i \neq j, i, j \in \{1, \ldots, M\}$, so we have

$$\sigma\left(\bigcup_{i,j\in\{1,\ldots,m\}:i\neq j}(W_{S^W_{A_i}} \cap W_{S^W_{A_j}}) \cup (U_{S^U_{A_i}} \cap U_{S^U_{A_j}})\right) \subset \sigma(Z). \tag{50}$$

Next Figure 2c implies

$$\sigma(A_1, \ldots, A_M) \not\subset \sigma(Z), \tag{51}$$

and

$$\sigma(Y) \not\subset \sigma(Z). \tag{52}$$

Therefore, we have

$$\sigma(Z) \subset \sigma\left(\{\epsilon_Z\} \bigcup (\cup_{i,j\in\{1,\ldots,m\}:i\neq j}(W_{S^W_{A_i}} \cap W_{S^W_{A_j}}) \cup (U_{S^U_{A_i}} \cap U_{S^U_{A_j}}))\right), \tag{53}$$

where ϵ_Z is independent of all the other errors in the structural model, including those of A and Y.

The error ϵ_Z can have an empty σ-algebra: for example, ϵ_Z is a constant. Therefore, the left side of Eq. 50 can be made equal to the right side of Eq. 53. We have

$$\sigma(Z) = \sigma\left(\{\epsilon_Z\} \bigcup (\cup_{i,j\in\{1,\ldots,m\}:i\neq j}(W_{S^W_{A_i}} \cap W_{S^W_{A_j}}) \cup (U_{S^U_{A_i}} \cap U_{S^U_{A_j}}))\right) \tag{54}$$

$$= \sigma\left(\{\epsilon_Z\} \bigcup W^{\mathrm{mlt}} \bigcup U^{\mathrm{mlt}}\right). \tag{55}$$

for some random variable ϵ_Z that is independent of all other random errors ϵ's. $\square$

As a direct consequence of Lemma 9, we have

$$P(y \mid u^{\mathrm{sng}}, u^{\mathrm{mlt}}, w^{\mathrm{mlt}}, a_1, \ldots, a_m, s=1) = P(y \mid u^{\mathrm{sng}}, z, a_1, \ldots, a_m, s=1), \tag{56}$$

due to the definition of conditional probabilities and $\epsilon_Z \perp Y \mid S, U^{\mathrm{sng}}, U^{\mathrm{mlt}}, W^{\mathrm{mlt}}, A_1, \ldots, A_m$. The latter is because ϵ_Z is independent of all other errors.

F. Proof of Lemma 4

Proof. Denote $U^{\mathrm{sng}}_{\mathcal{C}}$ as the set of single-treatment confounders that affects $A_{\mathcal{C}}$.

The proof of Lemma 4 relies on two observations.

The first observation starts with the integral equation we solve:

$$\begin{aligned}
&P(y \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}), u^{\mathrm{sng}}_{\mathcal{C}}, s=1) && (57)\\
=&\int h(y, a_{\mathcal{C}}, a_{\mathcal{X}}, u^{\mathrm{sng}}_{\mathcal{C}}) P(a_{\mathcal{X}} \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}), u^{\mathrm{sng}}_{\mathcal{C}}, s=1)\,\mathrm{d}a_{\mathcal{X}} && (58)\\
=&\int\int h(y, a_{\mathcal{C}}, a_{\mathcal{X}}, u^{\mathrm{sng}}_{\mathcal{C}}) P(a_{\mathcal{X}} \mid z) P(z \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}), u^{\mathrm{sng}}_{\mathcal{C}}, s=1)\,\mathrm{d}a_{\mathcal{X}}\,\mathrm{d}z && (59)
\end{aligned}$$

The first equality is due to Eq. 14. The second equality is due to Assumption 3.2.

The second observation relies on the null proxy:

$$\begin{aligned}
&P(y \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}), u^{\mathrm{sng}}_{\mathcal{C}}, s=1) && (60)\\
=&\int P(y \mid z, a_{\mathcal{C}}, f(a_{\mathcal{N}}), u^{\mathrm{sng}}_{\mathcal{C}}, s=1) P(z \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}), u^{\mathrm{sng}}_{\mathcal{C}}, s=1)\,\mathrm{d}z && (61)\\
=&\int P(y \mid z, a_{\mathcal{C}}, u^{\mathrm{sng}}_{\mathcal{C}}, s=1) P(z \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}), u^{\mathrm{sng}}_{\mathcal{C}}, s=1)\,\mathrm{d}z && (62)
\end{aligned}$$

The first equality is due to the definition of conditional probability. The second equality is due to the second part of Assumption 4; it implies $Y \perp f(a_{\mathcal{N}}) \mid Z, U^{\mathrm{sng}}_{\mathcal{C}}, A_{\mathcal{C}}, S=1$. The reason is that

$$\begin{aligned}
&P(y \mid z, a_{\mathcal{C}}, f(a_{\mathcal{N}}), u^{\mathrm{sng}}_{\mathcal{C}}, s=1) && (63)\\
=&\int P(y \mid z, a_{\mathcal{C}}, a_{\mathcal{X}}, f(a_{\mathcal{N}}), u^{\mathrm{sng}}_{\mathcal{C}}, s=1) P(a_{\mathcal{X}} \mid z, a_{\mathcal{C}}, f(a_{\mathcal{N}}), u^{\mathrm{sng}}_{\mathcal{C}}, s=1)\,\mathrm{d}a_{\mathcal{X}} && (64)\\
=&\int P(y \mid z, a_{\mathcal{C}}, a_{\mathcal{X}}, u^{\mathrm{sng}}_{\mathcal{C}}, s=1) P(a_{\mathcal{X}} \mid z, a_{\mathcal{C}}, u^{\mathrm{sng}}_{\mathcal{C}}, s=1)\,\mathrm{d}a_{\mathcal{X}} && (65)\\
=&P(y \mid z, a_{\mathcal{C}}, u^{\mathrm{sng}}_{\mathcal{C}}, s=1). && (66)
\end{aligned}$$

The second equality is again due to Assumption 3.2.

Comparing Eq. 59 and Eq. 62 gives

$$\int \left[P(y \mid z, a_{\mathcal{C}}, u^{\mathrm{sng}}_{\mathcal{C}}, s=1) - \int h(y, a_{\mathcal{C}}, a_{\mathcal{X}}, u^{\mathrm{sng}}_{\mathcal{C}}) P(a_{\mathcal{X}} \mid z)\,\mathrm{d}a_{\mathcal{X}}\right] \times P(z \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}), u^{\mathrm{sng}}_{\mathcal{C}}, s=1)\,\mathrm{d}z = 0, \tag{67}$$

which implies

$$P(y \mid z, a_{\mathcal{C}}, u^{\mathrm{sng}}_{\mathcal{C}}, s=1) = \int h(y, a_{\mathcal{C}}, a_{\mathcal{X}}, u^{\mathrm{sng}}_{\mathcal{C}}) P(a_{\mathcal{X}} \mid z)\,\mathrm{d}a_{\mathcal{X}}. \tag{68}$$

This step is due to the completeness condition in Assumption 4.2.

Eq. 68 leads to identification:

$$P(y \mid \mathrm{do}(a_{\mathcal{C}})) \tag{69}$$

$$=P(y \mid z, a_{\mathcal{C}}, u_{\mathcal{C}}^{\text{sng}})P(z)P(u_{\mathcal{C}}^{\text{sng}})\,\mathrm{d}z\,\mathrm{d}u_{\mathcal{C}}^{\text{sng}} \tag{70}$$
$$=P(y \mid z, a_{\mathcal{C}}, u_{\mathcal{C}}^{\text{sng}}, s=1)P(z)P(u_{\mathcal{C}}^{\text{sng}})\,\mathrm{d}z\,\mathrm{d}u_{\mathcal{C}}^{\text{sng}} \tag{71}$$
$$=\int\int\int h(y, a_{\mathcal{C}}, a_{\mathcal{X}}, u_{\mathcal{C}}^{\text{sng}})P(a_{\mathcal{X}} \mid z)\,\mathrm{d}a_{\mathcal{X}}P(z)P(u_{\mathcal{C}}^{\text{sng}})\,\mathrm{d}z\,\mathrm{d}u_{\mathcal{C}}^{\text{sng}} \tag{72}$$
$$=\int\int h(y, a_{\mathcal{C}}, a_{\mathcal{X}}, u_{\mathcal{C}}^{\text{sng}})P(a_{\mathcal{X}})P(u_{\mathcal{C}}^{\text{sng}})\,\mathrm{d}a_{\mathcal{X}}\,\mathrm{d}u_{\mathcal{C}}^{\text{sng}}. \tag{73}$$

In particular, the second equality is due to Assumption 3.2.

□

G. Proof of Theorem 5

We first state the variant of Assumption 3 and Assumption 4 required by Theorem 5. We essentially replace Z with $(U^{\text{mlt}}, W^{\text{mlt}})$ in these assumptions.

Assumption 6. *(Assumption 3') The causal graph Figure 2b satisfies the following conditions:*

1. *All single-treatment confounders U_i^{sng}'s are observed.*
2. *The selection operator S satisfies*
$$S \perp (A, Y) \mid U^{\text{mlt}}, W^{\text{mlt}}, U^{\text{sng}}. \tag{74}$$
3. *We observe the non-selection-biased distribution*
$$P(a_1, \ldots, a_m, u^{\text{sng}})$$
and the selection-biased distribution
$$P(y, u^{\text{sng}}, a_1, \ldots, a_m \mid s=1).$$

Assumption 7. *(Assumption 4') There exists some function f and a set $\emptyset \neq \mathcal{N} \subset \{1, \ldots, m\} \backslash \mathcal{C}$ such that*

1. *The outcome Y does not causally depend on $f(a_{\mathcal{N}})$:*
$$f(a_{\mathcal{N}}) \perp Y \mid A_{\mathcal{C}}, A_{\mathcal{X}}, U^{\text{mlt}}, W^{\text{mlt}}, U^{\text{sng}}, S=1 \tag{75}$$
where $\mathcal{X} = \{1, \ldots, m\} \backslash (\mathcal{C} \cup \mathcal{N}) \neq \emptyset$.
2. *The conditional $P(u^{\text{mlt}}, w^{\text{mlt}} \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}), u_{\mathcal{C}}^{\text{sng}}, s=1)$ is complete in $f(a_{\mathcal{N}})$ for almost all $a_{\mathcal{C}}$ and $u_{\mathcal{C}}^{\text{sng}}$, where $U_{\mathcal{C}}^{\text{sng}}$ is the single-treatment confounders affecting $A_{\mathcal{C}}$.*
3. *The conditional $P(f(a_{\mathcal{N}}) \mid a_{\mathcal{C}}, a_{\mathcal{X}}, u_{\mathcal{C}}^{\text{sng}}, s=1)$ is complete in $a_{\mathcal{X}}$ for almost all $a_{\mathcal{C}}$ and $u_{\mathcal{C}}^{\text{sng}}$.*

Under these assumptions, Theorem 5 is a direct consequence of Lemma 3 and Lemma 4. The reason is that $U^{\text{mlt}}, W^{\text{mlt}}, U^{\text{sng}}$ constitutes an admissible set to identify the intervention distributions $P(y \mid \mathrm{do}(a_{\mathcal{C}}))$.

H. Proof of Theorem 6

We assume Assumption 6 and Assumption 7 as described in Appendix G.

Proof. Assumption 5.2 guarantees the existence of some function $\hat{h}$ such that

$$\hat{P}(y \mid a_{\mathcal{C}}, \hat{z}, u_{\mathcal{C}}^{\text{sng}}, s=1) = \int \hat{h}(y, a_{\mathcal{C}}, a_{\mathcal{X}}, u_{\mathcal{C}}^{\text{sng}})\hat{P}(a_{\mathcal{X}} \mid \hat{z}, u_{\mathcal{C}}^{\text{sng}}, s=1)\,\mathrm{d}a_{\mathcal{X}} \tag{76}$$

under weak regularity conditions. (We discuss the reason in Appendix D.)

We first claim that $\hat{h}(y, a_{\mathcal{C}}, a_{\mathcal{X}}, u_{\mathcal{C}}^{\text{sng}})$ solves

$$P(y \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}), u_{\mathcal{C}}^{\text{sng}}, s=1) = \int \hat{h}(y, a_{\mathcal{C}}, a_{\mathcal{X}}, u_{\mathcal{C}}^{\text{sng}}) P(a_{\mathcal{X}} \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}), u_{\mathcal{C}}^{\text{sng}}, s=1) \, \mathrm{d}a_{\mathcal{X}}. \tag{77}$$

Given this claim (Eq. 77), we have

$$\begin{aligned}
&\hat{P}(y \mid \mathrm{do}(a_{\mathcal{C}})) \\
=&\int\int \hat{P}(y \mid \hat{z}, u_{\mathcal{C}}^{\text{sng}}, a_{\mathcal{C}}, s=1) \hat{P}(\hat{z}) P(u_{\mathcal{C}}^{\text{sng}}) \, \mathrm{d}\hat{z} \, \mathrm{d}u_{\mathcal{C}}^{\text{sng}} \\
=&\int\int\int \hat{h}(y, a_{\mathcal{C}}, a_{\mathcal{X}}, u_{\mathcal{C}}^{\text{sng}}) \hat{P}(a_{\mathcal{X}} \mid \hat{z}, u_{\mathcal{C}}^{\text{sng}}, s=1) \, \mathrm{d}a_{\mathcal{X}} \hat{P}(\hat{z}) P(u_{\mathcal{C}}^{\text{sng}}) \, \mathrm{d}\hat{z} \, \mathrm{d}u_{\mathcal{C}}^{\text{sng}} \\
=&\int\int\int \hat{h}(y, a_{\mathcal{C}}, a_{\mathcal{X}}, u_{\mathcal{C}}^{\text{sng}}) \hat{P}(a_{\mathcal{X}} \mid \hat{z}) \, \mathrm{d}a_{\mathcal{X}} \hat{P}(\hat{z}) P(u_{\mathcal{C}}^{\text{sng}}) \, \mathrm{d}\hat{z} \, \mathrm{d}u_{\mathcal{C}}^{\text{sng}} \\
=&\int\int \hat{h}(y, a_{\mathcal{C}}, a_{\mathcal{X}}, u_{\mathcal{C}}^{\text{sng}}) P(a_{\mathcal{X}}) \, \mathrm{d}a_{\mathcal{X}} P(u_{\mathcal{C}}^{\text{sng}}) \, \mathrm{d}u_{\mathcal{C}}^{\text{sng}} \\
=&P(y \mid \mathrm{do}(a_{\mathcal{C}})),
\end{aligned}$$

which proves the theorem. The first equality is due to Eq. 15; the second is due to Eq. 76; the third is due to Assumption 5 and $U_{\mathcal{C}}^{\text{sng}}$ being the single-treatment confounders for $A_{\mathcal{C}}$; the fourth is due to marginalizing out $\hat{Z}$; the fifth is due to the above claim (Eq. 77) and Theorem 5.

We next prove the claim (Eq. 77). Start with the right side of the equality.

$$\begin{aligned}
&\int \hat{h}(y, a_{\mathcal{C}}, a_{\mathcal{X}}, u_{\mathcal{C}}^{\text{sng}}) P(a_{\mathcal{X}} \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}), u_{\mathcal{C}}^{\text{sng}}, s=1) \, \mathrm{d}a_{\mathcal{X}} \\
=&\int\int \hat{h}(y, a_{\mathcal{C}}, a_{\mathcal{X}}, u_{\mathcal{C}}^{\text{sng}}) \hat{P}(a_{\mathcal{X}} \mid \hat{z}, u_{\mathcal{C}}^{\text{sng}}, a_{\mathcal{C}}, s=1) \hat{P}(\hat{z} \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}), u_{\mathcal{C}}^{\text{sng}}, s=1) \, \mathrm{d}a_{\mathcal{X}} \, \mathrm{d}\hat{z} \\
=&\int \hat{P}(y \mid a_{\mathcal{C}}, \hat{z}, u_{\mathcal{C}}^{\text{sng}}, s=1) \hat{P}(\hat{z} \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}), u_{\mathcal{C}}^{\text{sng}}, s=1) \, \mathrm{d}\hat{z} \\
=&\int \hat{P}(y \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}), \hat{z}, u_{\mathcal{C}}^{\text{sng}}, s=1) \hat{P}(\hat{z} \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}), u_{\mathcal{C}}^{\text{sng}}, s=1) \, \mathrm{d}\hat{z} \\
=&P(y \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}), u_{\mathcal{C}}^{\text{sng}}, s=1),
\end{aligned}$$

which establishes the claim. The first equality is due to Eq. 15; the second is due to Eq. 76; the third equality is due to Assumption 5.2, which implies

$$\hat{P}(y \mid a_{\mathcal{C}}, f(a_{\mathcal{N}}), \hat{z}, u_{\mathcal{C}}^{\text{sng}}, s=1) = \hat{P}(y \mid a_{\mathcal{C}}, \hat{z}, u_{\mathcal{C}}^{\text{sng}}, s=1). \tag{78}$$

The fourth equality is due to marginalizing out $\hat{z}$. $\square$

I. Constructing candidate $f(a_{\mathcal{N}})$'s from the deconfounder outcome model

We illustrate how to construct candidate $f(a_{\mathcal{N}})$'s in the deconfounder outcome model.

Consider a fitted linear outcome model

$$Y = \sum_{i=1}^{10} \alpha_{YA_i} A_i + \alpha_{YZ} \hat{Z} + \alpha_{YU'} U^{\text{sng}} + \epsilon_Y. \tag{79}$$

where all the random variables are Gaussian.

It implies that there exists $f_1(A_9, A_{10}) = A_9 + \alpha_{9,10} A_{10}$ that satisfies

$$f_1(A_9, A_{10}) \perp Y \mid \hat{Z}, U^{\text{sng}}, A_1, \ldots, A_8,$$

where

$$\alpha_{9,10} = -\frac{\alpha_9 \mathrm{Var}(A_9) + \alpha_{10}\mathrm{Cov}(A_9, A_{10})}{\alpha_9 \mathrm{Cov}(A_9, A_{10}) + \alpha_{10}\mathrm{Var}(A_{10})}.$$

The reason is that $f(A_9, A_{10}) \perp (\alpha_9 A_9 + \alpha_{10} A_{10})$. Hence $f(a_{\mathcal{N}}) = A_9 + \alpha_{9,10} A_{10}$ satisfies Assumption 5.2.

J. Details of the simulation study

Figure 3a. We simulate $n = 10,000$ data points from a linear Gaussian model and apply the deconfounder. For $\gamma_U = 0, 1, 2, 3, 4, 5,$

$$U_{n\times 1} \sim \mathcal{N}(0, I), \tag{80}$$
$$\theta_{1\times 3} \sim \mathrm{Unif}(0, I), \tag{81}$$
$$A_{n\times 3} \sim \mathcal{N}(U\theta, I), \tag{82}$$
$$\beta_{1\times 3} \sim \mathrm{Unif}(0, I), \tag{83}$$
$$\beta_0 \sim \mathrm{Unif}(0, 1), \tag{84}$$
$$Y \sim \mathcal{N}(\beta_0 + A\beta^\top + \gamma_U \cdot U, I). \tag{85}$$

To apply the deconfounder, we perform maximum likelihood estimation of PPCA on A and then fit a linear model of Y against both A and the PPCA factor.

As (1) the distributions of U, A, Y are all Gaussian, and (2) the Gaussianity of A leads to the existence of null proxy (as is discussed in Appendix I, the completeness conditions in Assumption 1 are satisfied.

Figure 3b. We perform the same simulation as above except that $U_{n\times 1} \sim \mathrm{Unif}(0, I)$. In this case, the distributions of A and Y no longer belong to the exponential family and violate the completeness conditions in Assumption 1.

Figures 3c and 3d. We perform the same pair of simulation as above except that we add an additional selection step to U. After generating U from $U_{n\times 1} \sim \mathcal{N}(0, I)$, we select U w.p. proportional to $\mathcal{N}(U\,;\,0, 0.5^2)/\mathcal{N}(U\,;\,0, I)$ and $\mathrm{Unif}(U\,;\,0, 0.5)/\mathrm{Unif}(U\,;\,0, I)$ respectively. The resulting U distribution is $\mathcal{N}(U\,;\,0, 0.5^2)$ and $\mathrm{Unif}(U\,;\,0, 0.5)$ respectively.

References

Miao, W., Geng, Z., & Tchetgen Tchetgen, E. J. (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4), 987–993.