

Reading Note: Identifying Effects of Multiple Treatments in the Presence of Unmeasured Confounding

Miao, W., Hu, W., Ogburn, E. L., & Zhou, X. H. (2022). Identifying Effects of Multiple Treatments in the Presence of Unmeasured Confounding. *Journal of the American Statistical Association*, 118(543), 1953–1967. <https://doi.org/10.1080/01621459.2021.2023551>

Motivation & problem

- In many applications (e.g. genetics, bioinformatics, epidemiology) one may have **multiple treatments or exposures** simultaneously that all might influence an outcome.
- A major challenge is unmeasured confounding: hidden variables affect both treatments and the outcome, biasing native estimates of causal effects.
- Much of the causal inference literature focuses on a single treatment with unmeasured confounding (e.g. instrumental variables, negative controls), but less is established when there are multiple treatments and shared unmeasured confounders.
- Existing methods to deal with multiple treatments often require strong assumptions (e.g. fully parametric outcome models, many instruments) or are flawed.
- **Under what conditions can we identify causal effects of multiple treatments in the presence of unmeasured confounding, and how?**

Main contributions & methods

They propose two general strategies to achieve identification and estimation of causal effects of multiple treatments under unmeasured confounding, without relying on strong parametric models for the outcome or on estimating the confounder directly.

0. Preliminaries

Notations

- $X = (X_1, \dots, X_p)^\top$: p -vector of treatments.
- Y : outcome.
- $Y(x)$: potential outcome.
- U : q -vector of unobserved covariates. Suppose q is known a priori.
- $f(a|b) := f(A = a|B = b)$: conditional density/mass of A given B .
- $f(x, u)$: treatment-confounder distribution.

- $f(y|u, x)$: outcome model.

Frameworks

- Target: identify of $f\{Y(x)\}$:
 - $f\{Y(x)\}$ is identifiable if and only if it is uniquely determined by the joint distribution of observed variables.
- Identification assumptions:
 - *Consistency*: when $X = x$, $Y = Y(x)$.
 - *Ignorability*: $Y(x) \perp\!\!\!\perp X \mid U$.
 - *Positivity*: $0 < f(X = x|U = u) < 1$ for all (x, u) .
- When Y is not observed, joint distributions $f(y, x, u)$ cannot be uniquely determined by $f(y, x)$ because...
- Call $\tilde{f}(y, x, u)$ is admissible if it conforms to the observed data distribution $f(y, x)$, i.e. $f(y, x) = \int_u \tilde{f}(y, x, u) du$.

1. Auxiliary variables approach

- The idea is to leverage variables (auxiliary variables) that are not causally related to the outcome, but which are related to the treatments (and possibly to the unmeasured confounder).

...

2. Null treatment approach

- Notations:
 - α : a possibly infinite-dimensional parameter;
 - $f(x, u|\alpha)$: treatment-confounder distribution;
 - $f(x; \alpha)$: marginal distribution;
 - $\mathcal{C} = \{i : f(u|x) \text{ varies with } x_i\}$: indices of confounded treatments that are associated with the confounder;
 - $\mathcal{A} = \{i : f(y|x, u) \text{ varies with } x_i\}$: active treatments that affect the outcome.
- An alternative that **does not require knowing which treatments are null (i.e. having no causal effect)**.
- Assumptions:

- *Null treatments*: the cardinality of the intersection $\mathcal{C} \cap \mathcal{A}$ does not exceed $(|\mathcal{C}| - q)/2$, where $|\mathcal{C}|$ is the cardinality of the \mathcal{C} must be larger than the dimension of U ;
 - *Equivalence*: for any α , any $\tilde{f}(x, u)$ that solves $f(x; \alpha) = \int_u \tilde{f}(x, u) du$ can be written as $\tilde{f}(x, u) = f\{X = x, V(U) = u; \alpha\}$ for some invertible but not necessarily known function V .
 - *Completeness*: for any α , $f(u | x; \alpha)$ is complete in any q -dimensional subvector x_S of the confounded treatment x_C .
- The key assumption (*null treatments assumption*) is: at least half of the (possibly confounded) treatments truly have zero causal effect on the outcome, but is it not necessarily to know which treatments are active.
 - Under that assumption, the authors derive identification of the causal effects of the remaining treatments, again without requiring strong parametric assumptions:
 - **Theorem 2**: Under Assumption 1 and 3, for any joint distribution $\tilde{f}(x, u)$ that solves $f(x) = \int_u \tilde{f}(x, u) du$, there exists a unique solution $\tilde{f}(y|u, x)$ to the equation

$$f(y|x) = \int_u \tilde{f}(y|u, x) \tilde{f}(u|x) du,$$

and the potential outcome distribution is identifiable by

$$f\{Y(x) = y\} = \int_u \tilde{f}(y|u, x) \tilde{f}(u) du.$$

- This theorem states that treatment effects are identifiable if fewer than half of the confounded treatments can affect the outcome.
 - The equivalence assumption allows us to use an arbitrary admissible treatment-confounder distribution for identification;
 - The null treatment assumption allows us to construct Fredholm integral equations for the first kind to solve for the outcome model;
 - The completeness assumption guarantees uniqueness of the solution.
- Intuitively, the null treatments act as a kind of “anchor” or baseline to help disentangle confounding from causal effects. That is, each X can be used as an auxiliary variable.
 - Proposition 2: Let x_S denote an arbitrary q -dimensional subvector of x_C and $x_{\bar{S}}$ the rest component of x except for x_S . Under Assumptions 1 and 3, for any choice for x_S , if there exists a function $\tilde{f}(y | u, x_{\bar{S}})$ that solves

$$f(y | x) = \int_u \tilde{f}(y | u, x_{\bar{S}}) \tilde{f}(u | x_S, x_{\bar{S}})$$

given $\tilde{f}(x, u)$ and $f(y|x)$, and that the solution depends on at most $(|C| - q)/2$ ones of the confounded treatments, then it must solve

$$f(y|x) = \int_u \tilde{f}(y | u, x) \tilde{f}(u | x) du.$$

- By distinguishing null variable sets, **hypothesis testings** can be conducted. Here provides a test of the *sharp null hypothesis* of no joint effects, which requires neither auxiliary variables nor the null treatment assumption.
 - Sharp null hypothesis: $\mathbb{H}_0 : f(y | u, x) = f(y | u)$ for all x .
 - Proposition 3: Under Assumption 1 and (ii)-(iii) of Assumption 3 and given an admissible joint distribution $\tilde{f}(x, u)$, if the null hypothesis \mathbb{H}_0 is correct, then for all q -dimensional subvector x_S of the confounded treatments x_C , the solution to the following equation exists and is unique,

$$f(y | x) = \int_u \tilde{f}(y | u, x_{\bar{S}}) \tilde{f}(u | x_S, s_{\bar{S}}) du,$$

and the solution must satisfy the following equality,

$$f(y) = \int_u \tilde{f}(y | u, x_{\bar{S}}) \tilde{f}(u) du.$$

(The null hypothesis is too restricted.)

Estimation:

- “If an issue can be addressed nonparametrically then it will often be better to tackle it parametrically; however, if it cannot be resolved nonparametrically then it is usually dangerous to resolve it parametrically.”
- See Fig.2 for algorithm under the null treatments setting.
- Step 1: requires only the equivalence assumption, which places nontrivial restrictions on the treatment-confounder distribution.
 - To estimate $\tilde{f}(x, u)$, one needs to correctly specify a treatment-confounder model that meets the equivalence assumption, such as a factor or mixture model.
 - This step estimates the distribution of U joint with X , but does not estimate U itself.

- Step 2: First, $f(y | x)$ can be estimated parametrically or nonparametrically using standard density estimation techniques. Second, solve integral equations $f(y | x) = \int_u \tilde{f}(y | x, u) \tilde{f}(u | x)$, which do not admit analytical solutions in general. However, one may apply numerical method provided in Chae et al. (2019) for a numerical method under mild conditions.
- Step 3: essential applications of the g-formula.

The null treatments algorithm:

Null-1 Obtain an arbitrary admissible joint distribution $\tilde{f}(x, u)$.

Null-2 Use the estimate of $\tilde{f}(u | x)$ from Step 1, along with an estimate of $f(y | x)$, to solve Equation (6) for an estimate of $\tilde{f}(y | u, x)$. The constructive method described in Proposition 2 can be implemented to solve (6).

Null-3 Plug the estimate of $\tilde{f}(u)$ from Step 1 and $\tilde{f}(y | u, x)$ from Step 2 into Equation (7) to estimate $f\{Y(x)\}$.

Figure 1: Fig.2: the null treatments algorithm

An Example:

Consider the linear models

$$X = \alpha U + \epsilon, \quad \mathbb{E}(Y | X, U) = \beta^\top X + \delta^\top U,$$

$$\Sigma_U = I_q, \quad \mathbb{E}(U) = 0, \quad U \perp\!\!\!\perp \epsilon, \quad \Sigma_\epsilon \text{ diagonal}.$$

Here, U and ϵ are not necessarily normally distributed. - The coefficient β encoding the average treatment effects is of interest. - Let $\gamma = \Sigma_X^{-1} \alpha$ denotes the coefficients of regressing U on X . - Let $\mathcal{C} = \{i : \alpha_i \text{ is not a zero vector}\}$ denote the indices of cno-founded treatments, and - γ_C the submatrix consisting of the corresponding rows of γ . - Let A_i denote the i -th row of a matrix or a column vector A . **Theorem 3** can be expressed accordingly: - Theorem 3: The parameter β is identified under the model above and the following assumptions: - (i) at most $(|\mathcal{C}| - q)/2$ entries of β_C are nonzero; - (ii) after deleting any row, there remain two disjoint submatrices of α of full rank; - (iii) any submatrix of γ_C consisting of q rows has full rank.

Estimation of β . - Denote ξ the coefficients by regressing Y on X , i.e., $\mathbb{E}[Y | X] = X\xi$. - Given n iid samples, we first obtain $\hat{\gamma}$ by factor analysis of X (Step 1): - The matrix factorization model:

$X \approx U\alpha^\top + \epsilon$, where $U \in \mathbb{R}^{n \times q}$ contains the latent confounder scores per individual, $\alpha \in \mathbb{R}^{p \times q}$ contains the factor loading for each treatment. - Thus, in the population model, $X_j = \alpha_j U + \epsilon_j$. - Then, $\hat{\gamma} = \Sigma_X^{-1} \hat{\alpha}$. - (?) Matrix dimensions? - Then we obtain $\hat{\xi}$ by regression of Y on X , which can be viewed as a crude estimator of β with asymptotic bias $\gamma\delta$, i.e. $\xi = \beta + \gamma\delta$. - Note that $\xi_C = \beta_C + \gamma_C\delta$ for confounded treatments, then estimation of $\hat{\delta}$ can be cast as a standard robust regression given $\hat{\xi}$ and $\hat{\gamma}$: - Suppose $\hat{\gamma}_C$ is the design matrix and $\hat{\xi}_C$ is observations with outliers corresponding to nonzero entries of β_C . - Given a $n^{1/2}$ -consistent estimator $(\hat{\xi}, \hat{\gamma})$, we solve

$$\hat{\delta}^{lms} = \arg \min_{\delta} \text{median}\{(\hat{\xi}_i - \hat{\gamma}_i \delta)^2, i \in \hat{\mathcal{C}}\}, \quad \hat{\mathcal{C}} = \{i : \|\hat{\gamma}_i\|_2^2 > \log(n)/n\},$$

which is a least median of squares estimator minimizing median of the squared errors $(\hat{\xi}_i - \hat{\gamma}_i \delta)^2$ among the confounded treatments consistently selected by $\hat{\mathcal{C}}$. - The corresponding estimate of β is $\hat{\beta}^{lms} = \hat{\xi} - \hat{\gamma} \hat{\delta}^{lms}$.

- $(\hat{\delta}^{lms}, \hat{\beta}^{lms})$ are consistent under the assumptions of Theorem 3 and an additional regularity condition given $n^{1/2}$ -consistency of $(\hat{\xi}, \hat{\gamma})$, but they are not necessarily asymptotically normal.

Simulation studies.

We generate 2 confounders (U), 8 treatments (X), and an outcome (Y) as follows,

$$X = \alpha U + \varepsilon_X, \quad Y = \beta^\top X + \delta_Y U + \varepsilon_Y, \quad U \sim N(0, I_2), \quad \varepsilon_X \sim N(0, I_6), \quad \varepsilon_Y \sim N(0, 1);$$

$$\alpha^\top = \begin{pmatrix} 0 & 0.4 & 0.8 & 1.2 & 1.5 & -0.4 & -0.8 & -1.2 \\ 0 & 0.2 & 0.4 & 0.6 & 0.8 & -0.5 & -1.0 & -1.2 \end{pmatrix}, \quad \delta_Y = (1, 1).$$

We consider two choices for β ,

Case 1: $\beta_1 = \beta_2 = \beta_3 = 1, \beta_4 = \dots = \beta_8 = 0$;

Case 2: $\beta_1 = \beta_2 = \beta_3 = 1, \beta_4 = \beta_5 = 0.2, \beta_6 = \beta_7 = \beta_8 = 0$.

Under these settings, (X_2, \dots, X_8) are confounded but X_1 is not; the null treatments assumption is satisfied in Case 1 as only two of the confounded treatments are active, but it is violated in Case 2 where β_4 and β_5 also have a small effect on Y .

Figure 2: Fig.3: Simulations on the null treatments settings

- In case 1, the null treatments assumption is satisfied, as only two of the confounded treatments are active.

- In case 2, this assumption is violated.
- Three methods are used:
 - the null treatments estimation with correct dimension of the confounder (Null1),
 - the null treatments estimation with one confounder (Null2), i.e., the number of “specified” confounder is smaller than the truth,
 - the OLS.

Table 2: Coverage probability of the 95% bootstrap confidence interval for Case 1

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	Note: The first row is for
	0.967	0.970	0.973	0.975	0.987	0.974	0.990	0.983	sample size 2000 and the
Null ₁	0.943	0.958	0.959	0.967	0.969	0.961	0.963	0.975	second for 5000.

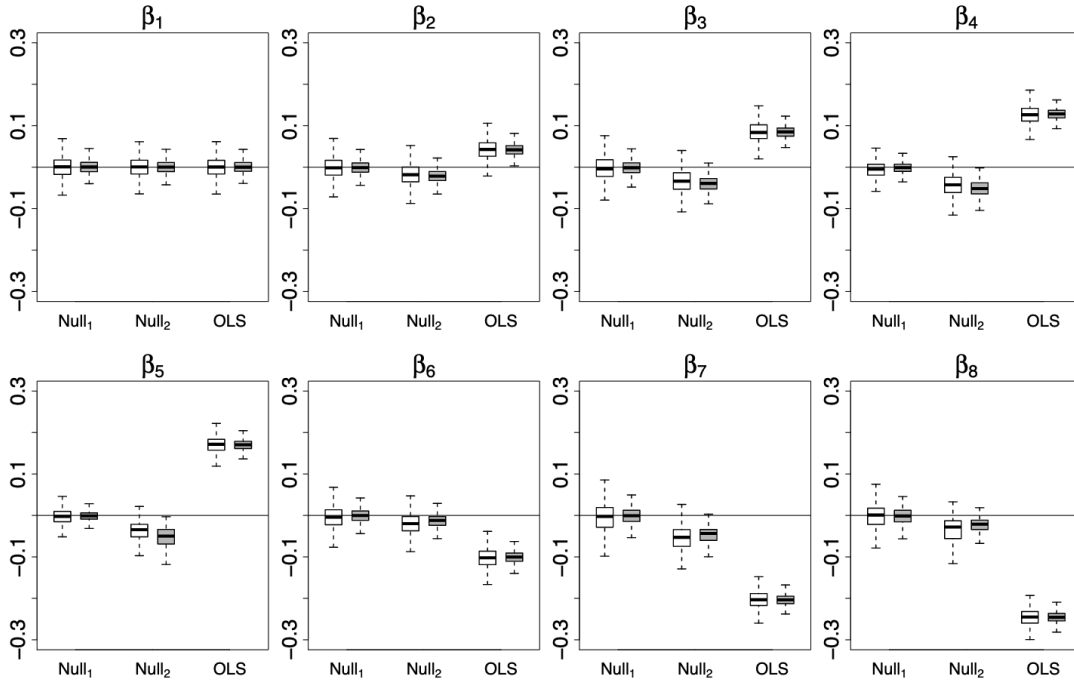


Figure 3: Bias of estimators in Case 1. White boxes are for sample size 2000 and gray ones for 5000.

Figure 3: Fig.4: Simulation results of case 1 under the null treatments settings

- In case 1:

- Null1, as the groundtruth method, provides unbiased estimates of all coefficients of interests.
 - Null2 with a wrong number of specified confounder, provides unbiased estimate for active but unconfounded treatment X_1 , but the estimates for the other treatments, either active or inactive, are biased, but the biases are less than those of the OLS.
 - OLS is always biased.
- In case 2:
 - All biased. See Fig.5.

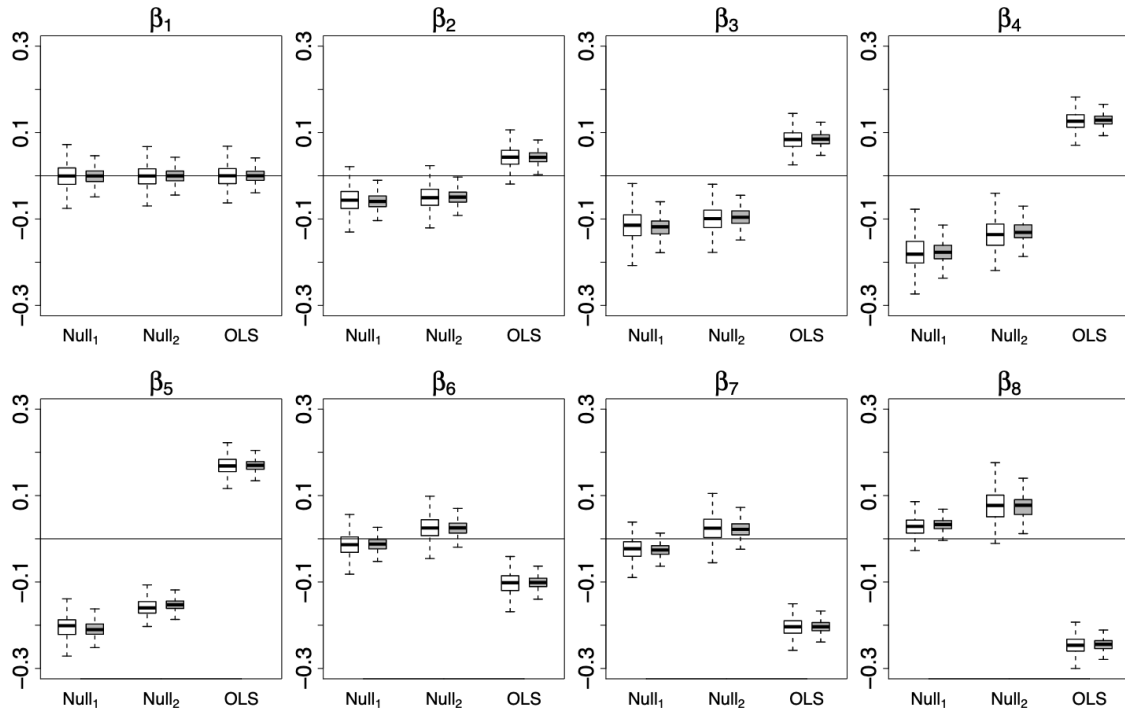


Figure S.6: Bias of estimators in Case 2 of the null treatments setting. White boxes are for sample size 2000 and gray ones for 5000.

Figure 4: Fig.5: Simulation results of case 2 under the null treatments settings

Discussion

- The proposed estimation strategies can also be used to test whether unmeasured confounding is present, by assessing how far the proposed estimates are from the crude

ones.

- The proposed estimation methods, comprised of standard factor analysis, linear and robust linear regression, inherit properties from the classical theory of statistical inference.
- However, statistical inference for nonparametric and semiparametric models remains to be studied.
- So far the study considers the low-dimensional settings, $n > p$.