Single Proxy Control

Eric J Tchetgen Tchetgen
Luddy Family President's Distinguished Professor
Department of Statistics
The Wharton School
University of Pennsylvania

05/10/2023

Outline

- Preliminaries

Outline

- Preliminaries
- Review of exchangeability for causal inference

Outline

- Preliminaries
- Review of exchangeability for causal inference
- Review of Control Outcome Calibration Approach

Outline

- Preliminaries
- Review of exchangeability for causal inference
- Review of Control Outcome Calibration Approach
- Nonparametric COCA under Completeness

Outline

- Preliminaries
- Review of exchangeability for causal inference
- Review of Control Outcome Calibration Approach
- Nonparametric COCA under Completeness
- Application

Outline

- Preliminaries
- Review of exchangeability for causal inference
- Review of Control Outcome Calibration Approach
- Nonparametric COCA under Completeness
- Application
- Conclusion

Outline

- Preliminaries
- Review of exchangeability for causal inference
- Review of Control Outcome Calibration Approach
- Nonparametric COCA under Completeness
- Application
- Conclusion
- Generalized Nonparametric COCA without Completeness (if time permits)

Preliminaries

- Unmeasured confounding or endogeneity bias is a well-known threat to valid causal inference from observational data.

Preliminaries

- Unmeasured confounding or endogeneity bias is a well-known threat to valid causal inference from observational data.
- An approach that is sometimes used in practice to assess residual confounding bias, is to check whether *known null effects* can be recovered free of bias, by evaluating whether the exposure or treatment of interest is found to be associated with a so-called negative control outcome (NCO), upon adjusting for all observed confounders .

Preliminaries

- Unmeasured confounding or endogeneity bias is a well-known threat to valid causal inference from observational data.
- An approach that is sometimes used in practice to assess residual confounding bias, is to check whether *known null effects* can be recovered free of bias, by evaluating whether the exposure or treatment of interest is found to be associated with a so-called negative control outcome (NCO), upon adjusting for all observed confounders .
- An observed variable is said to be a valid *negative control outcome* or more broadly, an *outcome confounding proxy,* to the extent that it is associated with hidden factors confounding the exposure-outcome relationship in view, although not directly impacted by the exposure.

Preliminaries

- Therefore, a negative control outcome which is empirically associated with the exposure, might suggest the presence of residual confounding.

Preliminaries

- Therefore, a negative control outcome which is empirically associated with the exposure, might suggest the presence of residual confounding.
- NCOs have appeared in literature as falsification outcomes/end points, control outcomes, secondary outcomes, supplementary responses, unaffected outcomes, and donor pool in synthetic control literature, proxies.

Preliminaries

- Examples negative control outcomes abound in practice:

Preliminaries

- Examples negative control outcomes abound in practice:
  - exposure=flu vaccine;outcome=hospitalization;NCO= injury/trauma hospitalization.

Preliminaries

- Examples negative control outcomes abound in practice:
  - exposure=flu vaccine;outcome=hospitalization;NCO= injury/trauma hospitalization.
  - exposure=COVID vaccination;outcome=covid; hospitalization;NCO=flu/RSV infection.

Preliminaries

- Examples negative control outcomes abound in practice:
  - exposure=flu vaccine;outcome=hospitalization;NCO= injury/trauma hospitalization.
  - exposure=COVID vaccination;outcome=covid; hospitalization;NCO=flu/RSV infection.
  - exposure=mamographic screening;outcome=breast cancer;NCO=death from other causes.

Preliminaries

- Examples negative control outcomes abound in practice:
  - exposure=flu vaccine;outcome=hospitalization;NCO= injury/trauma hospitalization.
  - exposure=COVID vaccination;outcome=covid; hospitalization;NCO=flu/RSV infection.
  - exposure=mamographic screening;outcome=breast cancer;NCO=death from other causes.
  - exposure=air pollution; outcome=elderly hospitalization;NCO=pre-exposure elderly hospitalization

Preliminaries

- Examples negative control outcomes abound in practice:
  - exposure=flu vaccine;outcome=hospitalization;NCO= injury/trauma hospitalization.
  - exposure=COVID vaccination;outcome=covid; hospitalization;NCO=flu/RSV infection.
  - exposure=mamographic screening;outcome=breast cancer;NCO=death from other causes.
  - exposure=air pollution; outcome=elderly hospitalization;NCO=pre-exposure elderly hospitalization
  - exposure=air pollution; outcome=asthma hospitalization;NCO=appendicitis hospitalization

Preliminaries

- Examples negative control outcomes abound in practice:
  - exposure=flu vaccine;outcome=hospitalization;NCO= injury/trauma hospitalization.
  - exposure=COVID vaccination;outcome=covid; hospitalization;NCO=flu/RSV infection.
  - exposure=mamographic screening;outcome=breast cancer;NCO=death from other causes.
  - exposure=air pollution; outcome=elderly hospitalization;NCO=pre-exposure elderly hospitalization
  - exposure=air pollution; outcome=asthma hospitalization;NCO=appendicitis hospitalization
  - exposure=Zika virus; outcome=birth rate; NCO=prior year birth rate.

Preliminaries

- We will use this last example for illustration: 673 municipalities in Brazil during Zika virus outbreak or 2015,185 municipalities experienced high incidence of Zika virus in Pernambuco region reported, while 488 municipalities in Rio Grande do Sul reported no case.

Preliminaries

- We will use this last example for illustration: 673 municipalities in Brazil during Zika virus outbreak or 2015,185 municipalities experienced high incidence of Zika virus in Pernambuco region reported, while 488 municipalities in Rio Grande do Sul reported no case.
- The Zika virus, which can be transmitted from a pregnant woman to her fetus, can cause serious brain abnormalities, including microcephaly (i.e., an abnormally small head).

Preliminaries

- We will use this last example for illustration: 673 municipalities in Brazil during Zika virus outbreak or 2015,185 municipalities experienced high incidence of Zika virus in Pernambuco region reported, while 488 municipalities in Rio Grande do Sul reported no case.
- The Zika virus, which can be transmitted from a pregnant woman to her fetus, can cause serious brain abnormalities, including microcephaly (i.e., an abnormally small head).
- Therefore interested in asking whether the Zika virus outbreak caused a drop in birth rates.

Preliminaries

- We will use this last example for illustration: 673 municipalities in Brazil during Zika virus outbreak or 2015,185 municipalities experienced high incidence of Zika virus in Pernambuco region reported, while 488 municipalities in Rio Grande do Sul reported no case.
- The Zika virus, which can be transmitted from a pregnant woman to her fetus, can cause serious brain abnormalities, including microcephaly (i.e., an abnormally small head).
- Therefore interested in asking whether the Zika virus outbreak caused a drop in birth rates.
- A crude comparison of exposed vs unexposed municipalities suggests that Zika virus was associated with an increase in birth rate in 2015 by 3.384 (95%CI:2.9-3.8) per 1000 persons, which is counter to expectations.

Preliminaries

- Using birth rate the prior year as NCO yields an association w Zika Virus with controls having higher prior rates, suggesting significant residual confounding bias.

Preliminaries

- Using birth rate the prior year as NCO yields an association w Zika Virus with controls having higher prior rates, suggesting significant residual confounding bias.
- A natural question is whether the NCO can be used for bias correction.

Preliminaries

- Using birth rate the prior year as NCO yields an association w Zika Virus with controls having higher prior rates, suggesting significant residual confounding bias.
- A natural question is whether the NCO can be used for bias correction.
- One such methods is Difference-in-differences approach (DiD) (Snow, 1856;, Card and Krueger, 1994) which leverages the pre-treatment outcome as NCO under parallel trends.

Preliminaries

- Using birth rate the prior year as NCO yields an association w Zika Virus with controls having higher prior rates, suggesting significant residual confounding bias.
- A natural question is whether the NCO can be used for bias correction.
- One such methods is Difference-in-differences approach (DiD) (Snow, 1856;, Card and Krueger, 1994) which leverages the pre-treatment outcome as NCO under parallel trends.
- From an NCO perspective, parallel trends can be interpreted as an additive equi-confounding assumption, that the unmeasured confounder association with the outcome of interest matches that with the pre-treatment outcome on the additive scale (Sofer et al, 2016).

Preliminaries

- In practice, equi-confounding or equivalently parallel trends may not be reasonable for a number of reasons, including if the outcome trend is also impacted by an unmeasured common cause with the treatment.

Preliminaries

- In practice, equi-confounding or equivalently parallel trends may not be reasonable for a number of reasons, including if the outcome trend is also impacted by an unmeasured common cause with the treatment.
- Furthermore, additive equiconfounding may not be realistic as a broader debiasing method in non-DiD settings where the NCO is not necessarily a pre-treatment outcome measurement, and might therefore have support on a different scale than the outcome of interest.

Preliminaries

- To address these limitations of additive equi-confounding, Tchetgen Tchetgen (2014) introduced the Control Outcome Calibration Approach (COCA) as a simple yet formal counterfactual approach to debias causal effect estimates in observational analyses.

Preliminaries

- To address these limitations of additive equi-confounding, Tchetgen Tchetgen (2014) introduced the Control Outcome Calibration Approach (COCA) as a simple yet formal counterfactual approach to debias causal effect estimates in observational analyses.
- COCA essentially treats the negative control outcome variable as a proxy measurement for the potential outcome of interest under no treatment, an assumption which is formalized in terms of a conditional independence of the NCO with the treatment mechanism, upon conditioning on the treatment-free counterfactual outcome.

Preliminaries

- This assumption formalizes the idea that, as a relevant proxy of residual confounding, the NCO would be made irrelevant to the treatment assignment mechanism if one were to hypothetically condition on the underlying potential outcome (a source of residual confounding).

Preliminaries

- This assumption formalizes the idea that, as a relevant proxy of residual confounding, the NCO would be made irrelevant to the treatment assignment mechanism if one were to hypothetically condition on the underlying potential outcome (a source of residual confounding).
- For inference, the COCA approach of Tchetgen Tchetgen (2014) involves correct specification of a regression model for the NCO, conditional on the treatment-free potential outcome together with a rank-preserving structural model of a constant individual level treatment effect.

Preliminaries

- In today's talk, I introduce nonparametric COCA identification and inference about the average causal effect for the treated

Preliminaries

- In today's talk, I introduce nonparametric COCA identification and inference about the average causal effect for the treated
- As we show, COCA identification is feasible without the need for rank preservation, therefore accommodating arbitrary degrees of effect heterogeneity across units.

Preliminaries

- In today's talk, I introduce nonparametric COCA identification and inference about the average causal effect for the treated
- As we show, COCA identification is feasible without the need for rank preservation, therefore accommodating arbitrary degrees of effect heterogeneity across units.
- Relatedly, *proximal causal inference* has recently developed in causal inference literature (Miao et al, TT et al, 2018), which leverages a pair of negative treatment and outcome control variables to nonparametrically identify treatment causal effects subject to residual confounding without invoking a rank-preservation assumption.

Preliminaries

- Importantly, while proximal causal inference relies on two proxies for causal identification, in contrast, COCA is a single-proxy control approach which therefore may present practical advantages.

Preliminaries

- Importantly, while proximal causal inference relies on two proxies for causal identification, in contrast, COCA is a single-proxy control approach which therefore may present practical advantages.
- For estimation and inference, we introduce three strategies to implement COCA which improve on prior methods: (i) an *extended propensity score approach*; (ii) a so-called *outcome calibration bridge function approach* and (iii) *a doubly robust approach* which carefully combines approaches (i) and (ii) and remains unbiased for the effect of treatment on the treated, provided that either approach is also unbiased, without necessarily knowing which method might be biased.

Review of exchangeability

- We aim to identify the population average causal effect for the treated (ATT), corresponding to a contrast of counterfactual averages

$$\beta = E(Y^{a=1} - Y^{a=0}|A = 1)$$

where Y^a is the potential outcome had, possibly contrary to fact a person received treatment $A = a$.

Review of exchangeability

- We aim to identify the population average causal effect for the treated (ATT), corresponding to a contrast of counterfactual averages

$$\beta = E(Y^{a=1} - Y^{a=0}|A = 1)$$

  where Y^a is the potential outcome had, possibly contrary to fact a person received treatment $A = a$.
- Standard assumptions for identification of β

Review of exchangeability

- We aim to identify the population average causal effect for the treated (ATT), corresponding to a contrast of counterfactual averages

$$\beta = E(Y^{a=1} - Y^{a=0}|A = 1)$$

where Y^a is the potential outcome had, possibly contrary to fact a person received treatment $A = a$.

- Standard assumptions for identification of β
  - Consistency

$$Y = Y_A; \tag{1}$$

Review of exchangeability

- We aim to identify the population average causal effect for the treated (ATT), corresponding to a contrast of counterfactual averages

$$\beta = E(Y^{a=1} - Y^{a=0}|A = 1)$$

where Y^a is the potential outcome had, possibly contrary to fact a person received treatment $A = a$.

- Standard assumptions for identification of β
  - Consistency

$$Y = Y_A; \tag{1}$$

  - Positivity:

$$f(A = a|L) > 0; \tag{2}$$

Review of exchangeability

- We aim to identify the population average causal effect for the treated (ATT), corresponding to a contrast of counterfactual averages

$$\beta = E(Y^{a=1} - Y^{a=0}|A = 1)$$

where Y^a is the potential outcome had, possibly contrary to fact a person received treatment $A = a$.

- Standard assumptions for identification of β
  - Consistency

$$Y = Y_A; \tag{1}$$

  - Positivity:

$$f(A = a|L) > 0; \tag{2}$$

  - Exchangeability or no unmeasured confounding (NUC) on the basis of measured covariates:

$$Y^{a=0} \amalg A|L. \tag{3}$$

Review of exchangeability

- Under assumptions (1)-(3) , it is well-known that

$$\beta = \sum_{l} \{E(Y|a=1,l) - E(Y|a=0,l)\} f(l|a=1); \qquad (4)$$

Review of exchangeability

- Under assumptions (1)-(3) , it is well-known that

$$\beta = \sum_{l} \{E(Y|a=1,l) - E(Y|a=0,l)\} f(l|a=1); \qquad (4)$$

- We are concerned that in observational settings, exchangeability may be unrealistic; for simplicity (WLOG), we will ignore measured covariates L, until stated otherwise, in which case we are concerned that observational data are generated from a graph compatible with Figure 1 below where the bow arc indicates the presence of unmeasured common causes of A and Y, i.e. endogeneity bias.

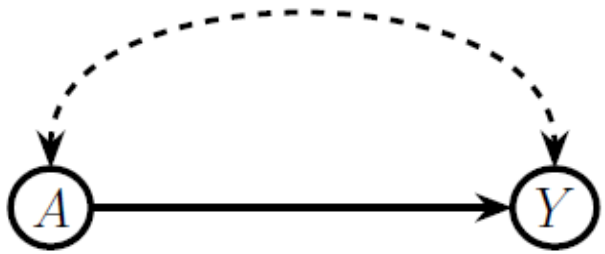

Figure 1: A Graphical Illustration of A Simple Causal Model.

Review of Control Outcome Calibration Approach (COCA) TT2014

- Instead of making an assumption of ignorabiliy/exchangeability, we will presume that one has measured a valid so-called negative control outcome, or more broadly a proxy W, which is known a priori to satisfy two conditions.

Review of Control Outcome Calibration Approach (COCA) TT2014

- Instead of making an assumption of ignorabiliy/exchangeability, we will presume that one has measured a valid so-called negative control outcome, or more broadly a proxy W, which is known a priori to satisfy two conditions.
  - First W must be associated with the mechanism leading to residual confounding in the sense that it stands for a proxy for the latter;

Review of Control Outcome Calibration Approach (COCA) TT2014

- Instead of making an assumption of ignorabiliy/exchangeability, we will presume that one has measured a valid so-called negative control outcome, or more broadly a proxy W, which is known a priori to satisfy two conditions.
  - First W must be associated with the mechanism leading to residual confounding in the sense that it stands for a proxy for the latter;
  - Secondly, W cannot be causally impacted by the treatment of interest.

Review of Control Outcome Calibration Approach (COCA) TT2014

- Instead of making an assumption of ignorabiliy/exchangeability, we will presume that one has measured a valid so-called negative control outcome, or more broadly a proxy W, which is known a priori to satisfy two conditions.
  - First W must be associated with the mechanism leading to residual confounding in the sense that it stands for a proxy for the latter;
  - Secondly, W cannot be causally impacted by the treatment of interest.
- Inspired by Tchetgen Tchetgen (2014), we formalize these conditions as followed:

$$W^a = W; \tag{5}$$

where W^a is the potential negative control outcome under an external intervention that sets $A = a$; and

$$W \not\perp\!\!\!\perp Y^{a=0}; \tag{6}$$

$$W \perp\!\!\!\perp A | Y^{a=0}. \tag{7}$$

Review of Control Outcome Calibration Approach (COCA) TT2014

- Condition (5) encodes the key assumption of a known null causal effect of the treatment on the NCO in potential outcome notation,

Review of Control Outcome Calibration Approach (COCA) TT2014

- Condition (5) encodes the key assumption of a known null causal effect of the treatment on the NCO in potential outcome notation,
- While conditions (6) and (7) formally encode the assumption that W is a valid proxy for the treatment-free potential outcome, the ultimate source of residual confounding bias.

Review of Control Outcome Calibration Approach (COCA) TT2014

- Condition (5) encodes the key assumption of a known null causal effect of the treatment on the NCO in potential outcome notation,
- While conditions (6) and (7) formally encode the assumption that W is a valid proxy for the treatment-free potential outcome, the ultimate source of residual confounding bias.
- We illustrate these assumptions graphically in Figure 2 below.

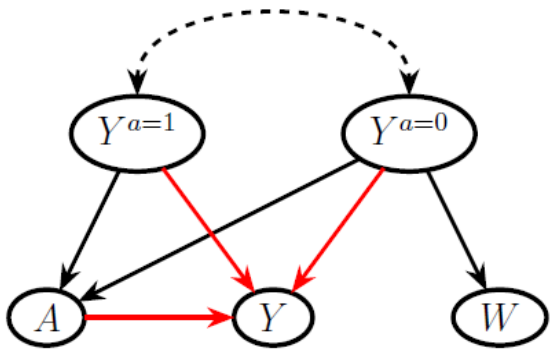

Figure 2: A Graphical Illustration of the Assumptions for Control Outcome Calibration Approach.

Review of Control Outcome Calibration Approach (COCA) TT2014

- Examples of negative controls/outcome confounding proxies abound in observational studies of health and social science: e.g. Pre-treatment outcome measurements typically used for DiD are excellent candidates and thus COCA can be viewed as scale-free approach for DiD, without making a parallel trend/equiconfounding condition, but rather based on a "structural assumption" we believe is easier for practitioners to reason about.

Review of Control Outcome Calibration Approach (COCA) TT2014

- Examples of negative controls/outcome confounding proxies abound in observational studies of health and social science: e.g. Pre-treatment outcome measurements typically used for DiD are excellent candidates and thus COCA can be viewed as scale-free approach for DiD, without making a parallel trend/equiconfounding condition, but rather based on a "structural assumption" we believe is easier for practitioners to reason about.
- Specifically, COCA conceptualizes NCO as an error-prone proxy of the treatment free potential outcome, that is, it views NCO from perspective of nonparametric measurement model.

Review of Control Outcome Calibration Approach (COCA) TT2014

- For identification and estimation in the case of a continuous outcome, Tchetgen Tchetgen (2014) further assumed the rank preserving structural model:

$$Y = Y^{a=0} + \psi_0 A, \tag{8}$$

which by consistency, implies constant individual level causal effect $\psi_0 = Y^{a=1} - Y^{a=0}$.

Review of Control Outcome Calibration Approach (COCA) TT2014

- For identification and estimation in the case of a continuous outcome, Tchetgen Tchetgen (2014) further assumed the rank preserving structural model:

$$Y = Y^{a=0} + \psi_0 A, \tag{8}$$

which by consistency, implies constant individual level causal effect $\psi_0 = Y^{a=1} - Y^{a=0}$.

- Under this model, he noted that upon defining $Y(\psi) = Y - \psi A$, then one can deduce from conditions (5)-(7) that

$$W \amalg A | Y(\psi) \tag{9}$$

if and only if $\psi = \psi_0$.

Review of Control Outcome Calibration Approach (COCA) TT2014

- For identification and estimation in the case of a continuous outcome, Tchetgen Tchetgen (2014) further assumed the rank preserving structural model:

$$Y = Y^{a=0} + \psi_0 A, \tag{8}$$

which by consistency, implies constant individual level causal effect $\psi_0 = Y^{a=1} - Y^{a=0}$.

- Under this model, he noted that upon defining $Y(\psi) = Y - \psi A$, then one can deduce from conditions (5)-(7) that

$$W \amalg A | Y(\psi) \tag{9}$$

if and only if $\psi = \psi_0$.

- COCA inverts a test of the above null hypothesis (9) to construct 95%CI for ψ_0.

Nonparametric COCA via IPW

- Though practically convenient, validity of either of above approach relies on both correct specification of the linear model for W given $Y^{a=0}$, and on the causal effect being constant across unit, an assumption that is rarely appropriate. In the following, we describe alternative methods aimed at addressing these limitations.

Nonparametric COCA via IPW

- Though practically convenient, validity of either of above approach relies on both correct specification of the linear model for W given $Y^{a=0}$, and on the causal effect being constant across unit, an assumption that is rarely appropriate. In the following, we describe alternative methods aimed at addressing these limitations.
- In order to introduce the approach, consider the extended propensity score function:

$$\pi(y) = \Pr\left(A = 1 | Y^{a=0} = y\right)$$

which makes explicit the fact that in presence of unmeasured confounding, the treatment mechanism will generally depend on the treatment-free potential outcome.

Nonparametric COCA via IPW

- Next, we note that were $\pi(y)$ known, identification of the average treatment-free potential outcome in the treated would then be empirically identified by the expression

$$E\left(Y^{a=0}|A=1\right) = \frac{E\left((1-A)\,Y\frac{\pi(Y)}{1-\pi(Y)}\right)}{E\left((1-A)\frac{\pi(Y)}{1-\pi(Y)}\right)} \tag{10}$$

and therefore, the average causal effect of treament on the treated would be identified by

$$E\left(Y^{a=1} - Y^{a=0}|A=1\right) = E\left(Y|A=1\right) - \frac{E\left((1-A)\,Y\frac{\pi(Y)}{1-\pi(Y)}\right)}{E\left((1-A)\frac{\pi(Y)}{1-\pi(Y)}\right)} \tag{11}$$

Nonparametric COCA via IPW

- As $\pi(y)$ is unknown, we next demonstrate how the NCO assumption can be leveraged to identify the latter quantity. Let $p(w) = \Pr(A = 1|W = w)$ .The proposed approach to identify π is based on the following equality, which we prove in the appendix:

Nonparametric COCA via IPW

- As $\pi(y)$ is unknown, we next demonstrate how the NCO assumption can be leveraged to identify the latter quantity. Let $p(w) = \Pr(A = 1|W = w)$ .The proposed approach to identify π is based on the following equality, which we prove in the appendix:
- *Result 1: Under consistency and Assumptions* $(5) - (7)$, *we have that*

$$\frac{p(w)}{1-p(w)} = E\left\{\frac{\pi(Y)}{1-\pi(Y)}|W = w, A = 0\right\} \tag{12}$$
$$= \sum_y \frac{\pi(y)}{1-\pi(y)} g(y|w, A = 0)$$

where $g(y|w, A = 0)$ *is the law of* Y *given* Z *and* A, *evaluated at* $(y, w, a = 0)$.

Nonparametric COCA via IPW

- As $\pi(y)$ is unknown, we next demonstrate how the NCO assumption can be leveraged to identify the latter quantity. Let $p(w) = \Pr(A = 1|W = w)$ .The proposed approach to identify π is based on the following equality, which we prove in the appendix:
- *Result 1: Under consistency and Assumptions* $(5) - (7)$, *we have that*

$$\frac{p(w)}{1-p(w)} = E\left\{\frac{\pi(Y)}{1-\pi(Y)}|W = w, A = 0\right\} \tag{12}$$
$$= \sum_y \frac{\pi(y)}{1-\pi(y)} g(y|w, A = 0)$$

  where $g(y|w, A = 0)$ *is the law of* Y *given* Z *and* A, *evaluated at* $(y, w, a = 0)$.
- the result relates p, g to π; the first two are uniquely determined by the observed data, which are then related to the unknown function of interest π in the Result.

Nonparametric COCA via IPW

- *Result 2:If Assumptions* $(5)-(7)$ *and equation* (12) *admits a unique solution, then* $\pi(Y)$ *is nonparametrically identified from the observed data; and* $E\left(Y^{a=0}|A=1\right)$ *is nonparametrically identified by*

$$E\left(Y^{a=1}-Y^{a=0}|A=1\right)=E\left(Y|A=1\right)-\frac{E\left((1-A)\,Y\frac{\pi(Y)}{1-\pi(Y)}\right)}{E\left((1-A)\frac{\pi(Y)}{1-\pi(Y)}\right)}$$

Nonparametric COCA via IPW

- *Result 2:If Assumptions* $(5)-(7)$ *and equation* (12) *admits a unique solution, then* $\pi(Y)$ *is nonparametrically identified from the observed data; and* $E\left(Y^{a=0}|A=1\right)$ *is nonparametrically identified by*

$$E\left(Y^{a=1}-Y^{a=0}|A=1\right)=E\left(Y|A=1\right)-\frac{E\left((1-A)\,Y\frac{\pi(Y)}{1-\pi(Y)}\right)}{E\left((1-A)\frac{\pi(Y)}{1-\pi(Y)}\right)}$$

- Equation (12) for $\pi(Y)$ is known in mathematics as an integral equation, more precisely a Fredholm integral equation of the first kind. Sufficient conditions for the existence and uniqueness of a solution to such an equation are well studied by mathematicians. Such conditions are also familiar to econometricians as they play an important role in so-called nonparametric instrumental variable model (see papers by Newey, Blundell, Chen etc....) and were also recently discussed by statisticians in the context of proximal causal inference (Miao et al, 2016, Tchetgen Tchetgen et al, 2020).

Nonparametric COCA via Outcome Bridge Function

- We introduce an alternative approach which does not rely upon identifying the EPS π^*, but instead relies on the existence of a so-called *outcome bridge function*, which we introduce in the following assumption.

Nonparametric COCA via Outcome Bridge Function

- We introduce an alternative approach which does not rely upon identifying the EPS π^*, but instead relies on the existence of a so-called *outcome bridge function*, which we introduce in the following assumption.
- *Condition 2: for any y, there exist a solution $b^*(W)$ to the following equation*

$$y = E\left\{b^*(W) \,|\, Y = y, A = 0\right\}. \tag{13}$$

Nonparametric COCA via Outcome Bridge Function

- We introduce an alternative approach which does not rely upon identifying the EPS π^*, but instead relies on the existence of a so-called *outcome bridge function*, which we introduce in the following assumption.
- *Condition 2: for any y, there exist a solution $b^*(W)$ to the following equation*

$$y = E\left\{b^*(W) \,|\, Y = y, A = 0\right\}. \tag{13}$$

- Intuitively, upon noting that under assumptions $(5) - (7)$, Condition 2 can equivalently be stated in terms of potential outcomes

$$Y^{a=0} = E\left\{b^*(W) \,|\, Y^{a=0}\right\} \tag{14}$$

which essentially formalizes the idea that W is a sufficiently relevant proxy for the potential outcome $Y^{a=0}$ if there exist a (possibly unknown) transformation of W whose conditional expectation given $Y^{a=0}$, recovers $Y^{a=0}$.

Nonparametric COCA via Outcome Bridge Function

$$Y^{a=0} = E\left\{b^*(W) \,|\, Y^{a=0}\right\} \tag{15}$$

- Note that classical measurement error is a special case of the equation in the display above in which case b is the identity map and

$$W = Y^{a=0} + \varepsilon \tag{16}$$

where ε is a mean zero independent error.

Nonparametric COCA via Outcome Bridge Function

$$Y^{a=0} = E\left\{b^*(W) \,|\, Y^{a=0}\right\} \tag{15}$$

- Note that classical measurement error is a special case of the equation in the display above in which case b is the identity map and

$$W = Y^{a=0} + \varepsilon \tag{16}$$

  where ε is a mean zero independent error.
- The condition can therefore be viewed as a nonparametric generalization of classical measurement error which allows W and Y to be of arbitrary nature and does not assume the error to be unbiased on the additive scale.

Nonparametric COCA via Outcome Bridge Function

- Similar to Condition 1, Condition 2 formally defines a Fredholm integral equation of the first kind, for which sufficient conditions for existence and uniqueness of a solution are well characterized; we again refer the reader to Miao et al (2016).

Nonparametric COCA via Outcome Bridge Function

- Similar to Condition 1, Condition 2 formally defines a Fredholm integral equation of the first kind, for which sufficient conditions for existence and uniqueness of a solution are well characterized; we again refer the reader to Miao et al (2016).
- *Result 3:If Assumptions* $(5)-(7)$ *and Condition 2 hold, then*

$$E\left(Y^{a=0}|A=1\right)=E\left(b^{*}\left(W\right)|A=1\right) \tag{17}$$

therefore

$$E\left(Y^{a=1}-Y^{a=0}|A=1\right)=E\left(Y|A=1\right)-E\left(b^{*}\left(W\right)|A=1\right)$$

Nonparametric COCA via Outcome Bridge Function

- Similar to Condition 1, Condition 2 formally defines a Fredholm integral equation of the first kind, for which sufficient conditions for existence and uniqueness of a solution are well characterized; we again refer the reader to Miao et al (2016).
- *Result 3:If Assumptions* $(5)-(7)$ *and Condition 2 hold, then*

$$E\left(Y^{a=0}|A=1\right)=E\left(b^{*}\left(W\right)|A=1\right) \tag{17}$$

therefore

$$E\left(Y^{a=1}-Y^{a=0}|A=1\right)=E\left(Y|A=1\right)-E\left(b^{*}\left(W\right)|A=1\right)$$

- Important to note that unique identification of b^{*} is not necessary for identifying the ATT, any solution to equation in Condition 2 yields the same ATT.

Nonparametric COCA under Completeness

- Estimation of b and π can be formulated as a minimax optimization program of the form :

$$\begin{aligned}\widehat{b} &= \arg \min_{b \in \mathcal{H}} \max_{t \in \mathcal{T}} \mathbb{P}_n \left\{ t(Y)(1-A)(Y - b(W)) - t^2(Y) \right\} \\ &\quad -\text{Pen}_1 \times ||t|| + \text{Pen}_2 \times ||b||\end{aligned}$$

Nonparametric COCA under Completeness

- Estimation of b and π can be formulated as a minimax optimization program of the form :

$$\begin{aligned}\widehat{b} &= \arg \min_{b \in \mathcal{H}} \max_{t \in \mathcal{T}} \mathbb{P}_n \left\{ t(Y)(1-A)(Y - b(W)) - t^2(Y) \right\} \\ &\quad -\text{Pen}_1 \times ||t|| + \text{Pen}_2 \times ||b||\end{aligned}$$

- One can view $\mathcal{H}$, $\mathcal{T}$ as functional spaces that are sufficiently small in order to regularize the inverse pb defining b^*.

Nonparametric COCA under Completeness

- Estimation of b and π can be formulated as a minimax optimization program of the form :

$$\begin{aligned}\widehat{b} &= \arg \min_{b \in \mathcal{H}} \max_{t \in \mathcal{T}} \mathbb{P}_n \left\{ t(Y)(1-A)(Y - b(W)) - t^2(Y) \right\} \\ &\quad -\text{Pen}_1 \times ||t|| + \text{Pen}_2 \times ||b||\end{aligned}$$

- One can view $\mathcal{H}$, $\mathcal{T}$ as functional spaces that are sufficiently small in order to regularize the inverse pb defining b^*.
- π^* can likewise be estimated via adversarial learning.

Nonparametric COCA under Completeness

- Functional spaces $\mathcal{H}$ and $\mathcal{T}$ can either be finite dimensional (parametric models) or infinite dimensional (semiparametric/nonparametric models).

Nonparametric COCA under Completeness

- Functional spaces $\mathcal{H}$ and $\mathcal{T}$ can either be finite dimensional (parametric models) or infinite dimensional (semiparametric/nonparametric models).
- Have formal asymptotic results when these spaces are chosen to be (i) finite dimensional or (ii) nonparametric methods using Reproducing Kernel Hilbert Spaces (RKHS). (Ghassami et al. 2021).

Nonparametric COCA under Completeness

- Functional spaces $\mathcal{H}$ and $\mathcal{T}$ can either be finite dimensional (parametric models) or infinite dimensional (semiparametric/nonparametric models).
- Have formal asymptotic results when these spaces are chosen to be (i) finite dimensional or (ii) nonparametric methods using Reproducing Kernel Hilbert Spaces (RKHS). (Ghassami et al. 2021).
- In fact, can show that combining such estimators w cross-fittiing (Schick,1986; Chernozukov et al,2018), the estimator of :

$$\mathbb{P}_n \left\{ A\widehat{b}(W) + \frac{(1-A)\,\widehat{\pi}(Y)}{[1-\widehat{\pi}(Y)]} \left\{ Y - \widehat{b}(W) \right\} \right\} / \mathbb{P}_n (A=1)$$

is root-n consistent and asymptotically normal provided that $\|\widehat{\pi} - \pi\| \times \left\|\widehat{b} - b\right\| = o_p\left(n^{-1/2}\right)$, so called mixed bias criterion (Rotnitzky, Smucler, Robins, 2019), i.e. it is doubly robust.

Zika in Brazil

- Recall that we are considerig 673 municipalities in Brazil during Zika virus outbreak or 2015,with 185 municipalities experienced high incidence of Zika virus in Pernambuco region reported, while 488 municipalities in Rio Grande do Sul reported no case.

Zika in Brazil

- Recall that we are considerig 673 municipalities in Brazil during Zika virus outbreak or 2015,with 185 municipalities experienced high incidence of Zika virus in Pernambuco region reported, while 488 municipalities in Rio Grande do Sul reported no case.
- Recall that a crude estimate suggested that Zika virus increased birth rate by 2015 by **3.384 (95%CI:(2.9,3.8) per 1000 persons in 2015**, which is counter to expectations and suggests bias due to hidden confounding factors.

Zika in Brazil

- Recall that we are considerig 673 municipalities in Brazil during Zika virus outbreak or 2015,with 185 municipalities experienced high incidence of Zika virus in Pernambuco region reported, while 488 municipalities in Rio Grande do Sul reported no case.
- Recall that a crude estimate suggested that Zika virus increased birth rate by 2015 by **3.384 (95%CI:(2.9,3.8) per 1000 persons in 2015**, which is counter to expectations and suggests bias due to hidden confounding factors.
- COCA appears to correct for such residual bias to yield **-1.817 (95%CI:(-2.9,-0.6) fewer birth per 1000 persons in 2015**. In contrast DiD gave somewhat smaller effect estimate, however consistent w COCA estimate **-1.191 95% CI (-1.507, -0.876)**

Conclusions

- COCA offers alternative approach to address residual confounding with a single proxy, therefore adding to bag of tools using proxies (IVs,DiD, PCI)

Conclusions

- COCA offers alternative approach to address residual confounding with a single proxy, therefore adding to bag of tools using proxies (IVs,DiD, PCI)
  - Relative to IV, no monotonicity or homogeneity assumptions needed for identification

Conclusions

- COCA offers alternative approach to address residual confounding with a single proxy, therefore adding to bag of tools using proxies (IVs,DiD, PCI)
  - Relative to IV, no monotonicity or homogeneity assumptions needed for identification
  - Relative to DiD, no equi-confounding assumption needed on any scale

Conclusions

- COCA offers alternative approach to address residual confounding with a single proxy, therefore adding to bag of tools using proxies (IVs,DiD, PCI)
  - Relative to IV, no monotonicity or homogeneity assumptions needed for identification
  - Relative to DiD, no equi-confounding assumption needed on any scale
  - Relative to PCI, Single proxy control instead of double proxy control

Conclusions

- COCA offers alternative approach to address residual confounding with a single proxy, therefore adding to bag of tools using proxies (IVs,DiD, PCI)
  - Relative to IV, no monotonicity or homogeneity assumptions needed for identification
  - Relative to DiD, no equi-confounding assumption needed on any scale
  - Relative to PCI, Single proxy control instead of double proxy control
- COCA is universal in the sense that it can be used to identify any causal effect on the treated that can in principle be identified under unconfoundedness.

Conclusions

- COCA offers alternative approach to address residual confounding with a single proxy, therefore adding to bag of tools using proxies (IVs,DiD, PCI)
  - Relative to IV, no monotonicity or homogeneity assumptions needed for identification
  - Relative to DiD, no equi-confounding assumption needed on any scale
  - Relative to PCI, Single proxy control instead of double proxy control
- COCA is universal in the sense that it can be used to identify any causal effect on the treated that can in principle be identified under unconfoundedness.
- Sensitivity analysis can be conducted to evaluate potential impact of violation of key assumption $W \amalg A | Y^{a=0}$ (see ETT, Park and Richardson, 2023)

Conclusions

- COCA offers alternative approach to address residual confounding with a single proxy, therefore adding to bag of tools using proxies (IVs,DiD, PCI)
  - Relative to IV, no monotonicity or homogeneity assumptions needed for identification
  - Relative to DiD, no equi-confounding assumption needed on any scale
  - Relative to PCI, Single proxy control instead of double proxy control
- COCA is universal in the sense that it can be used to identify any causal effect on the treated that can in principle be identified under unconfoundedness.
- Sensitivity analysis can be conducted to evaluate potential impact of violation of key assumption $W \amalg A | Y^{a=0}$ (see ETT, Park and Richardson, 2023)
- Much work remains to be done:e.g. COCA for front door, Mediation, interference, repairing invalid IV, complex longitudinal studies, SC.

Acknowledgments

- This work is funded by NIH grants R01AG065276, R01CA222147 and R01AI27271. (PI:Tchetgen Tchetgen)
- Co-authors of Single Proxy Control Paper : Chan Park and David Richardson; and JR, TR, AG and IS for helpful discussions.

  - TT, EJ., 2014. The control outcome calibration approach for causal inference with unobserved confounding. American journal of epidemiology, 179(5), pp.633-640.
  - TT EJ, Park Chan and David Richardson (2023). Single Proxy Control. https://arxiv.org/abs/2302.06054.
  - Li W, Miao W, TT EJ. (2021)Nonparametric inference about mean functionals of nonignorable nonresponse data without identifying the joint distribution. JRSSB. In Press
  - Ghassami A, Ying A, Shpitser I, TT EJ. (2022) Minimax Kernel Machine Learning for a Class of Doubly Robust Functionals with Application to Proximal Causal Inference. International Conference on Artificial Intelligence and Statistics (pp. 7210-7239). PMLR.
  - Miao, W., Geng, Z. and TT, E.J., 2018. Identifying causal effects with proxy variables of an unmeasured confounder. Biometrika, 105(4),

Generalized Nonparametric COCA without Completeness

- Suppose we wish to identify a functional μ

$$\mu = E\left(\tau\left(Y^{a=0}\right)|A=1\right)$$

for any user specified function τ under COCA conditions

$$W^a = W; W \not\perp\!\!\!\perp Y^{a=0}; W \perp\!\!\!\perp A | Y^{a=0}$$

without invoking any completeness condition.

Generalized Nonparametric COCA without Completeness

- Suppose we wish to identify a functional μ

$$\mu = E\left(\tau\left(Y^{a=0}\right)|A=1\right)$$

for any user specified function τ under COCA conditions

$$W^a = W; W \not\perp\!\!\!\perp Y^{a=0}; W \perp\!\!\!\perp A | Y^{a=0}$$

without invoking any completeness condition.

- *Result 5: (*Li, Miao and TT, 2021), *We establish that μ is identified if and only if*

$$\tau(Y) \in \mathcal{N}(T)^{\perp}$$

where $\mathcal{N}(T)$ is the null space of the operator $T: L_2(Y) \longrightarrow L_2(W)$ given by $T(n) = E(n(Y)|A=0, W)$.

Generalized Nonparametric COCA without Completeness

- Suppose we wish to identify a functional μ

$$\mu = E\left(\tau\left(Y^{a=0}\right)|A=1\right)$$

for any user specified function τ under COCA conditions

$$W^a = W; W \not\perp\!\!\!\perp Y^{a=0}; W \perp\!\!\!\perp A | Y^{a=0}$$

without invoking any completeness condition.

- *Result 5: (*Li, Miao and TT, 2021), *We establish that μ is identified if and only if*

$$\tau(Y) \in \mathcal{N}(T)^{\perp}$$

where $\mathcal{N}(T)$ is the null space of the operator $T: L_2(Y) \longrightarrow L_2(W)$ given by $T(n) = E(n(Y)|A=0, W)$.

- This result establishes that identification of μ is possible if and only if $\tau(Y) \in \overline{\mathcal{R}(T')}$. This result is unfortunately not sufficient for root-n estimation of μ as we show next.

Nonparametric COCA without Completeness

- *Result 6:* (Li, Miao and TT, 2021), *We establish that* μ *is root-n-estimable if and only if*

$$\tau(Y) \in \mathcal{R}\left(T'\right)$$

where $\mathcal{R}\left(T'\right)$ is range of the adjoint operator $T' : L_2(W) \longrightarrow L_2(Y)$ given by $T(m) = E(m(W)|A = 0, Y)$, the adjoint of T.

Nonparametric COCA without Completeness

- *Result 6:* (Li, Miao and TT, 2021), *We establish that* μ *is root-n-estimable if and only if*

$$\tau(Y) \in \mathcal{R}(T')$$

where $\mathcal{R}(T')$ is range of the adjoint operator $T' : L_2(W) \longrightarrow L_2(Y)$ given by $T(m) = E(m(W)|A = 0, Y)$, the adjoint of T.

- This result can be stated equivalently as: μ *is root-n estimable, if and only if there exist an* L_2 *integrable function* $b(W)$*, such that*

$$E(b(W)|A = 0, Y = y) = \tau(Y)$$

This is our outcome bridge function condition 2 in case $\tau(Y) = Y$.

Nonparametric COCA without Completeness

- This result provides an only if condition for root-n estimation of any functional that can be defined as a solution to a moment condition;e.g. θ =median$(Y^{a=0}|A=1)$ identified by setting $\tau(Y;\theta) = (I(Y<\theta) - 1/2)$.

Nonparametric COCA without Completeness

- This result provides an only if condition for root-n estimation of any functional that can be defined as a solution to a moment condition;e.g. θ =median$(Y^{a=0}|A=1)$ identified by setting $\tau(Y;\theta) = (I(Y<\theta) - 1/2)$.
- Importantly, this result implies that completeness is strictly not necessary for root-n estimation of μ, nor is identification of the EPS π necessary.

Nonparametric COCA without Completeness

- This result provides an only if condition for root-n estimation of any functional that can be defined as a solution to a moment condition;e.g. θ =median$(Y^{a=0}|A=1)$ identified by setting $\tau(Y;\theta) = (I(Y<\theta) - 1/2)$.
- Importantly, this result implies that completeness is strictly not necessary for root-n estimation of μ, nor is identification of the EPS π necessary.
- This however complicates inference considerably. We have proposed such inference methods in following step (Li, Miao, TT, 2021).

Nonparametric COCA without Completeness

- To construct a root-n-consistent estimator of μ, we follow an empirical version of the following steps (Li, Miao, TT, 2021):

Nonparametric COCA without Completeness

- To construct a root-n-consistent estimator of μ, we follow an empirical version of the following steps (Li, Miao, TT, 2021):
  1. We characterize an approximate solution set

$$\begin{aligned}\mathcal{B}_K &= \{b : E(b(W)|A=0, Y=y) \approx \tau(Y)\} \\ &\quad \cap \{b : span(\psi_k : k \leq K)\}\end{aligned}$$

  for a complete bases $\{\psi_k : k\}$

Nonparametric COCA without Completeness

- To construct a root-n-consistent estimator of μ, we follow an empirical version of the following steps (Li, Miao, TT, 2021):
  1. We characterize an approximate solution set

$$\begin{aligned}\mathcal{B}_K &= \{b : E(b(W)|A=0, Y=y) \approx \tau(Y)\} \\ &\quad \cap \{b : span(\psi_k : k \leq K)\}\end{aligned}$$

     for a complete bases $\{\psi_k : k\}$
  2. We select a uniquely defined minimum norm element of $\mathcal{B}_K$, say

$$b_K = \arg\min_{b \in \mathcal{B}_K} E\left(\{b(W)(1-A)\}^2\right)$$

Nonparametric COCA without Completeness

- To construct a root-n-consistent estimator of μ continued...:

Nonparametric COCA without Completeness

- To construct a root-n-consistent estimator of μ continued...:
  3. As the EPS π may not be identified, rather than leveraging the latter, we obtain a "Riesz representer" $h(W)$ of the continuous linear functional

$$E(b(W)|A=1) =< b, h > \text{ for all } b \in \mathcal{H}$$

by solving the optimization

$$h^* \approx \arg\min_{h \in \mathcal{H}} E\left((1-A)E(h(W)|A=0,Y)^2\right) - 2E((1-A)h(W))$$

Nonparametric COCA without Completeness

- To construct a root-n-consistent estimator of μ continued...:
  3. As the EPS π may not be identified, rather than leveraging the latter, we obtain a "Riesz representer" $h(W)$ of the continuous linear functional

$$E(b(W)|A=1) =< b, h > \text{ for all } b \in \mathcal{H}$$

by solving the optimization

$$h^* \approx \arg\min_{h \in \mathcal{H}} E\left((1-A)E(h(W)|A=0,Y)^2\right) - 2E((1-A)h(W))$$

  4. Our debiased root-n estimator is based on an empirical version of the doubly robust representation $\mu =$

$$E\left\{\begin{array}{c} \frac{A}{\Pr(A=1)} b_k(W) \\ + \frac{(1-A)}{\Pr(A=1)} \underbrace{E(h^*(W)|A=1,Y)}_{\text{replaces } \pi/(1-\pi)} \{\tau(Y) - b_k(W)\} \end{array}\right\}$$