Methodological obstacles in causal inference: confounding, missing data, and measurement error

Penning de Vries, B.B.L.

9

Negative controls: concepts and caveats

Bas B. L. Penning de Vries
Rolf H. H. Groenwold

[Submitted]

Abstract

Unmeasured confounding is a well-known obstacle in causal inference. In recent years, negative controls have received increasing attention as a important tool to address concerns about the problem. The literature on the topic has expanded rapidly and several authors have advocated the more routine use of negative controls in epidemiological practice. In this paper, we review concepts and methodologies based on negative controls for detection and correction of unmeasured confounding bias. We argue that negative controls may lack both specificity and sensitivity to detect unmeasured confounding and that proving the null hypothesis of a null negative control association is impossible. We focus our discussion on the control outcome calibration approach, the difference-in-difference approach, and the double-negative control approach as methods for confounding correction. For each of these methods, we highlight their assumptions and illustrate the potential impact of violations thereof. Given the potentially large impact of assumption violations, it may sometimes be desirable to replace strong conditions for exact identification with weaker, easily verifiable conditions, even when these imply at most partial identification of unmeasured confounding. Future research in this area may broaden the applicability of negative controls and in turn make them better suited for routine use in epidemiological practice. At present, however, the applicability of negative controls should be carefully judged on a case-by-case basis.

9.1 Introduction

In epidemiological research on causal effects, there are often concerns that one or more assumptions—such as exchangeability, no measurement error, or assumptions about missing data—are violated. In efforts to lessen these concerns, it has long been suggested that auxiliary variables be used that have a known (e.g., null) causal relation with the exposure or outcome of interest (Rosenbaum, 1989; Lipsitch et al., 2010; Flanders et al., 2011). Observing an association that contradicts the belief in a causal null might alert the analyst to violations of the assumptions underlying the methods used in the study. Auxiliary variables known to be causally unrelated to the variables of primary interest are called negative controls and have potential in bias detection as well as partial or complete bias correction in epidemiological research (Shi et al., 2020b).

In recent years, negative controls have received increasing attention in the epidemiological and statistical literature. The literature on how to leverage negative controls in studies on causal effects has rapidly expanded and several authors have argued that negative controls should be more commonly employed (Lipsitch et al., 2010; Arnold et al., 2016; Shi et al., 2020b). This paper aims to complement these efforts to increase the more routine implementation of negative controls with a discussion about a selection of caveats. Focusing on the use of negative controls to address possible violations of the exchangeability assumption, i.e., the assumption of no unmeasured confounding, we begin with a brief review of relevant definitions and discuss assumptions for bias detection. We then review methods for bias correction and study their sensitivity to assumption violations.

9.2 Negative controls

A negative control outcome (NCO) is a variable that is not causally affected by the exposure of interest A (Tchetgen Tchetgen, 2013; Shi et al., 2020b). Likewise, a negative control exposure (NCE) is a variable that does not causally affect the outcome of interest Y, except possibly through the exposure of interest (Shi et al., 2020b). The causal DAGs of Figure 9.1 (discussed later in this section) give examples of settings where a variable Z classifies as an NCO, an NCE or both. Given the absence of a direct causal effect of exposure A on an NCO Z or of NCE Z on outcome Y, any observed association between A and an Z, or between an Z and outcome Y given A, must be spurious. Leveraging negative controls involves translating information about such spurious associations into information about the spuriousness of associations between the primary exposure and outcome variables of interest.

9.2.1 *Negative controls for unmeasured confounding detection*

Let $Y(a)$ denote the outcome that would be realised had exposure A been set to a. Together with causal consistency (i.e., $Y(a) = Y$ if $A = a$) and positivity, epidemiologists often seek to invoke the exchangeability (or unmeasured confounding) condition $Y(a) \perp\!\!\!\perp A$ (possibly within levels of a collection of observed variables) to establish identifiability of the effect of exposure A on outcome Y (Hernán and Robins, 2020). In observational studies, however, it is seldom evident that the exchangeability condition, E, for the exposure-outcome relation of interest is achieved. A key idea of negative controls is to find a 'control' statement, C, that translates into information about E and which is more easily verified or refuted.

Control statement C may refer to the absence of bias of a measure of the association between A and Y and the NCO or NCE variable, respectively. Knowing that any control association is noncausal renders the control statement empirically verifiable. If C implies E, then a null finding for the control statement would imply conditional exchangeability for the exposure-outcome relation of interest. Conversely, if E implies C, evidence of bias of the control association corroborates the existence of unmeasured confounding.

9.2.2 *Caveats in the use of negative controls to detect unmeasured confounding*

There are a number of caveats concerning the use of negative controls for confounding detection. These caveats mainly concern the link between the control statement and exchangeability for the exposure-outcome relation of interest. Unfortunately, the extent to which one confers information about the other need

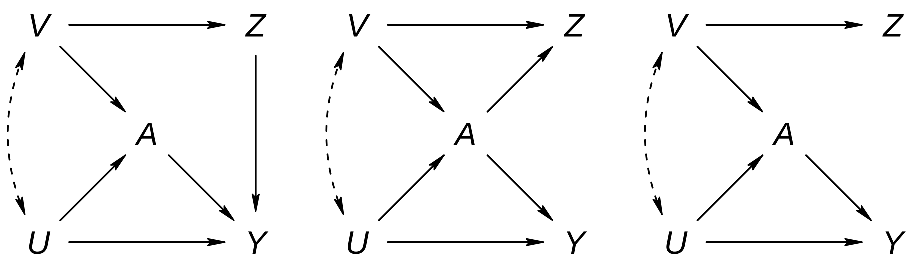

Figure 9.1: Causal directed acyclic graphs of settings where Z is a negative control outcome (left), a negative control exposure (middle) or both (right).

not be evident (Groenwold, 2013). A biased negative-control association need not imply unmeasured confounding for the exposure-outcome relation of interest and neither is the converse true generally.

First, while most applications of negative controls assume that confounding is the only source of bias, in reality it may be one of potentially many sources of bias. A spurious negative control association could have resulted, at least in part, from collider stratification, measurement error or violations of assumptions about missing data (Arnold et al., 2016). Even if unmeasured confounding for the negative control association implies unmeasured confounding for the exposure-outcome relation of interest, a biased negative control association need not be a reflection of unmeasured confounding. Conversely, a (near) null finding could be the result of opposing biases, masking the presence of unmeasured confounding. In other words, negative controls are a tool that may lack both specificity and sensitivity with respect to the type(s) of bias they are to detect.

Lipsitch et al. (2010) suggested a principle for establishing a link that is based on the extent to which common causes of A and Y overlap with the common causes of the exposure or outcome and the negative control variable. Clearly, for an NCO, with complete overlap (e.g., $V = U$ in Figure 9.1), the set of common causes of A and Y is empty if and only if the set of common causes of A and the NCO is empty. However, null values for certain measures of the effect of A on an NCO or of an NCE on Y need not imply that the set of unobserved common causes is empty, or, therefore, that there is conditional exchangeability for the primary exposure-outcome relation. Indeed, near null values may be the result of partially opposing confounding effects (or, more generally, opposing biases) and the relative effects may be different for the NCO versus the primary outcome Y.

With finite samples rather than complete knowledge of the theoretical or population distribution, sampling variability becomes relevant too, making it more important to acknowledge the distinction between absence of evidence and evidence of absence (Albert and Anderson, 1984). With finite samples, proving the null hypothesis of a null negative control association is impossible. Even if 'highly' powered studies cannot detect bias for the negative control relation, it may be injudicious to assume that the available data are sufficient to adequately control for confounding of the primary relation of interest, because a small degree of bias for the former relation may be associated with a substantial degree of bias for the latter. Sample size and power considerations are often ignored or left at secondary importance. While some papers have considered the power of negative control tests (Rosenbaum, 1989; Birch, 1964), it is typically ignored how the negative control association relates to the extent of bias for the exposure-outcome relation of interest, yet high power to detect 'small departures' from exposure-

NCO or NCE-outcome independence need not imply high power to detect small bias due to unmeasured confounding of the primary relation of interest. What are considered 'small departures' should therefore depend on the relation between the negative control association and the bias for the exposure-outcome relation of interest. Conversely, even if there is evidence of the contrary to the negative control null hypothesis, the bias due to uncontrolled confounding for the primary exposure-outcome relation may not be meaningful. In any case, it is important to consider the relative size of the biases in the negative control and primary exposure-outcome relations.

9.3 Negative control methods for uncontrolled confounding adjustment

The more recent literature on negative controls has considered how and under what conditions negative controls can be leveraged to partially or fully identify target causal quantities rather than merely the presence of bias. Lipsitch et al. (2010) gives conditions for valid inference about the direction of bias and thus for partial identification of the target causal quantity. These conditions are reviewed in Supplementary Appendix S9.1. In what follows, we review three methods for full identification: the control outcome calibration approach (COCA), the (generalised) difference-in-difference approach, and the double-negative control approach. Proofs of identification are given in Supplementary Appendix S9.2 for completeness. For each of the methods, we illustrate the potential impact of assumption violations on the identifiability of the targeted quantity. Throughout, departures from identification are termed bias.

9.3.1 Control outcome calibration approach

Identification

It may be tempting to regard the confounded association between the exposure of interest and an NCO as a direct measure of bias for the exposure-outcome effect of interest. However, it cannot generally be assumed that the direction or magnitude of bias are the same for the two relations. As an alternative to the restrictive and probably unrealistic "bias equivalence" assumption, i.e., the assumption of equality between between the confounded negative control association and the bias due to unmeasured confounding of the exposure-outcome effect of interest, Tchetgen Tchetgen (2013) proposed the COCA. The assumption of "bias equivalence" would especially likely be violated if the NCO and primary outcome are measured on different scales and the bias is bounded differently depending on the scale, such as would be the case if the NCO were binary and the primary outcome continuous. The COCA leverages an NCO to adjust for unmeasured confounding without requiring that the NCO and primary outcome are measured on similar scales.

The next result, due to Tchetgen Tchetgen (2013), describes a regression-based approach to implementing the COCA, which—characteristically of the COCA—relies on the assumption that a (set of) counterfactual primary outcome(s) of interest is sufficient to render the NCO conditionally independent of the exposure of interest. Some intuition behind this approach may be obtained upon noting that the counterfactual outcomes may well capture information about baseline covariates and therefore serve as a proxy for unobserved pre-exposure variables that are predictive of the NCO. The reasoning rests on the assumption that the same covariates that explain the lack of exchangeability for the outcome of interest also explain the confounding of the exposure-NCO relation. However, even then it is not evident nor guaranteed that the counterfactual outcome proxy is sufficient to render the NCO and exposure conditionally independent.

Theorem 9.1 (A regression-based approach to implementing the COCA under rank preservation)**.** *Suppose that the following conditions hold for all levels a of A:*
- Consistency: $Y(a) = Y$ *if* $a = A$.
- Rank preservation: *for some constant* θ, $Y(0) = Y(a) - \theta a$.
- Exposure-NCO independence given counterfactual outcome: $Z \perp\!\!\!\perp A | Y(0)$.
- NCO model: *for known one-to-one model link* g,

$g(\mathbb{E}[Z|A, Y]) = \beta_0 + \beta_1 A + \beta_2 Y$*, where* $\beta_0, \beta_1, \beta_2$ *are identified by a regression of* Z *on* A *and* Y*, and* $\beta_2 \neq 0$.

Then, $\mathbb{E}[Y(a) - Y(a-1)] = \theta$ *is identified by* $-\beta_1/\beta_2$.

Because counterfactual outcome $Y(0)$ may not fully account for the unmeasured confounding between the exposure and NCO, it is important that the impact of assumption violations be gauged. To this end, Tchetgen Tchetgen (2013) described a sensitivity analysis, given below in Theorem 9.2, for the special case of Theorem 9.1 where g is the identity link and A is a linear combination of $Y(0)$ and an error term Δ. When the sensitivity parameter (ρ) is set to 0, it is implicitly assumed that the NCO and exposure of interest are independent given counterfactual outcome $Y(0)$ (because χ is independent of (A, Y) and therefore of $Y(0)$) and, so, the result of Theorem 9.1 is recovered.

Theorem 9.2 (Sensitivity analysis for violations of $Z \perp\!\!\!\perp A|Y(0)$)**.** *Suppose the following conditions hold for all levels a of A:*
- Consistency: $Y(a) = Y$ *if* $a = A$.
- Rank preservation: *for some constant* θ, $Y(0) = Y(a) - \theta a$.
- Conditional exposure-NCO independence: $Z \perp\!\!\!\perp A|(Y(0), \Delta)$.
- Exposure model: $A = \alpha_0 + \alpha_1 Y(0) + \Delta$.
- NCO model: $Z = \beta_0 + \beta_1 Y(0) + \rho\Delta + \chi$, $\chi \perp\!\!\!\perp (A, Y)$.

Then, $\mathbb{E}[Z|A, Y] = \beta_0^* + \beta_1^* A + \beta_2^* Y$ *for some* $\beta_0^*, \beta_1^*, \beta_2^*$, *and if parameters* β_1^*, β_2^* *are identified (by a regression of* Z *on* A *and* Y*) and* $\beta_2^* \neq 0$, *then* $\theta = (\beta_1^* - \rho)/\beta_2^*$.

Through the rank preservation assumption, Theorem 9.1 relies also on the strong assumption that the set of all counterfactual outcomes of an individual are deterministically linked. A prerequisite of this assumption is that the within-person ranks of counterfactuals are the same for all individuals. In the next section, we consider violations of this assumption. However, as Theorem 9.3 states, in the special case where the outcome and exposure of interest are binary, there should be no concern about violations of this assumption as it can be dropped entirely (Tchetgen Tchetgen, 2013).

Theorem 9.3 (COCA for binary primary outcome and exposure)**.** *Suppose that the following conditions hold:*
- Consistency: $Y(a) = Y$ *if* $a = A$
- Positivity: $0 < \Pr(A = a, Y = y)$ *for* $y = 0, 1$.
- Exposure-NCO independence given counterfactual outcome: $Z \perp\!\!\!\perp A|Y(a)$.
- Non-zero denominator: $\mathbb{E}[Z|A = a, Y = 1] - \mathbb{E}[Z|A = a, Y = 0] \neq 0$.

Then,

$$\begin{aligned}\mathbb{E}[Y(a)] = &\ \mathbb{E}[Y|A = a]\Pr(A = a)\\ &+ \frac{\mathbb{E}[Z|A = 1 - a] - \mathbb{E}[Z|A = a, Y = 0]}{\mathbb{E}[Z|A = a, Y = 1] - \mathbb{E}[Z|A = a, Y = 0]}\Pr(A = 1 - a).\end{aligned}$$

If the assumptions of Theorem 9.3 are met for $a = 1$, the average treatment effect among the treated (ATT) $\mathbb{E}[Y - Y(0)|A = 1]$ is identified. For identification of the average treatment effect (ATE) $\mathbb{E}[Y(1) - Y(0)]$, the result requires that the assumptions are met for $a = 0, 1$. We will consider violations of these assumptions in the next section.

Sensitivity to assumption violations

In this subsection, we consider the sensitivity of the COCA to assumption violations. In particular we illustrate the potential impact of deviating from rank preservation and of violating the assumption that counterfactual outcome $Y(0)$ renders the exposure and NCO conditionally independent. While classical measurement error in the outcome does not hamper inference in terms of bias in the classical linear regression setting, we also illustrate that this for of measurement error does result in bias of the COCA.

First, to illustrate the potential impact of deviating from rank preservation, consider the setting where A is binary and where the following models hold:

$$\left.\begin{aligned} \theta|A &\sim \text{Normal}(\mathbb{E}[\theta], \sigma_\theta^2), \\ Y(0)|A, \theta &\sim \text{Normal}(\alpha_0 + \alpha_1 A, \sigma_Y^2), \\ Y = Y(A) &= Y(0) + \theta A, \\ Z|(A, \theta, Y(0)) &\sim \text{Normal}(\gamma_0 + \gamma_1 Y(0), \sigma_Z^2). \end{aligned}\right\} \tag{9.1}$$

A standard implementation of the COCA as per Theorem 9.1 yields $\hat{\theta} = -\hat{\beta}_1/\hat{\beta}_2$, where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the coefficients for A and Y of an ordinary least squares regression of Z on A and Y.

Given a value of the ATE (i.e, $\mathbb{E}[\theta]$), the parameter values are fully determined under models (9.1) by the joint distribution of the observed variables A, Y, Z (Supplementary Appendix S9.3). In particular, given a fixed distribution of (A, Y, Z), the variance of the individual effects $Y(1) - Y(0)$ (i.e., $\text{Var}(\theta) = \sigma_\theta^2$) and the ATE are linearly related via

$$\text{Var}(\theta) = \frac{\text{Var}(A)\text{Var}(Y) - \text{Cov}(A, Y)^2}{(\text{Var}(A) + \mathbb{E}[A]^2)\text{Cov}(A, Z)}(\hat{\beta}_1 - \hat{\beta}_2\mathbb{E}[\theta])$$

(Supplementary Appendix S9.3). For values of the ATE between -4 and 2, we chose parameter values such that the distribution of (A, Y, Z) has marginal means $\mathbb{E}[A] = 0.25$, $\mathbb{E}[Y] = 0$ and $\mathbb{E}[Z] = 0$, and covariance matrix

$$\begin{bmatrix} 3/16 & 1/2 & 1/2 \\ 1/2 & 3 & 2 \\ 1/2 & 2 & 4 \end{bmatrix}. \tag{9.2}$$

Figure 9.2 shows the bias of the COCA for the ATE. As shown, the magnitude of bias is zero under rank preservation but increases linearly with increasing variance of individual exposure-outcome effects.

In illustrating the sensitivity of the COCA against violations of rank preservation, it was assumed that the other assumptions were maintained. We now turn to the assumption of Exposure-NCO independence given counterfactual outcome $Y(0)$ and likewise assume that all other assumptions, including rank preservation, are met. In particular, we consider the setting where $Y(0)$ is the sum of two independent variables U_1, U_2. By assuming the following models, we also stipulate that some (albeit not necessarily the same) linear combination $\alpha_0' + \alpha_1' U_1 + \alpha_2' U_2$ is sufficient to render the exposure of interest and NCO conditionally independent:

$$\left.\begin{aligned} U_1 &\perp\!\!\!\perp U2, \\ A|(U_1, U_2) &\sim \text{Normal}(\alpha_0 + \alpha_1 U_1 + \alpha_2 U_2, \sigma_A^2), \\ Y = Y(A) &= U_1 + U_2 + \theta A,\ \theta \text{ constant}, \\ Z|(U_1, U_2, A, Y) &\sim \text{Normal}(\alpha_0' + \alpha_1' U_1 + \alpha_2' U_2, \sigma_Z^2) \end{aligned}\right\} \tag{9.3}$$

Variables U_1 and U_2 can be viewed as common causes of the NCO and the exposure and outcome of interest. Again, the COCA identifies the quantity $\hat{\theta} = -\hat{\beta}_1/\hat{\beta}_2$ based on an ordinary least squares regression of NCO Z on A and Y, but this quantity is not generally equal to θ. Figure 9.3 shows the asymptotic bias (departure from identification of the ATE) of the COCA plotted against α_2 over the interval $(-5, 5)$ for the special case where U_1 and U_2 take the standard

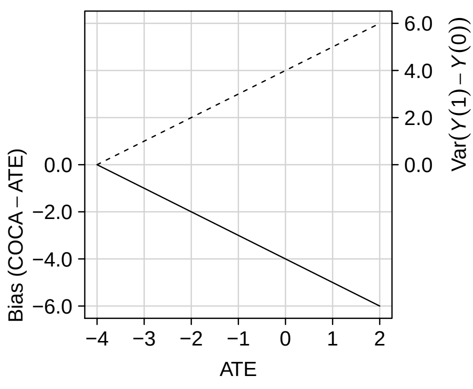

Figure 9.2: Illustration of the effect of violating the rank preservation assumption on the difference between the quantity identified by the COCA and the ATE (bias). The dashed line depicts the relation between the variance of individual exposure outcome effects $Y(1) - Y(0)$ and the mean $\mathbb{E}[Y(1) - Y(0)]$ (the ATE) under a fixed observed data distribution; the solid line describes the relation between the ATE and the bias of the implementation of the COCA.

normal distribution and where $\alpha_0, \alpha_0', \alpha_2' = 0$, $\alpha_1, \sigma_A^2, \sigma_Z^2 = 1$ and $\alpha_1' = 2$. The bias is zero only when counterfactual outcome $Y(0)$ is proportional to the linear combination of common causes U_1 and U_2 that renders the NCO and exposure of interest conditionally independent.

With $\alpha_2, \alpha_2' = 0$, models (9.3) imply the same joint distribution of observed variables A, Y, Z as models (9.4):

$$\left.\begin{array}{c} U_1 \perp\!\!\!\perp U2, \\ A|(U_1, U_2) \sim \text{Normal}(\alpha_0 + \alpha_1 U_1, \sigma_A^2), \\ Y(A) = U_1 + \theta A,\ \theta \text{ constant}, \\ Y = Y(A) + U_2, \\ Z|(U_1, U_2, A, Y) \sim \text{Normal}(\alpha_0' + \alpha_1' U_1, \sigma_Z^2) \end{array}\right\} \tag{9.4}$$

An important difference between (9.3) and (9.4) is that the consistency assumption is violated (provided that $\text{Var}(U_2) > 0$). The observed outcome Y is now the sum of the outcome of interest $Y(A)$ and an independent mean-zero error term. Figure 9.3 therefore also illustrates that the validity of the COCA also critically rests on the absence of classical measurement error in the outcome. At $\alpha_2 = 0$, Figure 9.3 gives the bias of the COCA under (9.4) with the values for the parameters given above. Although ATE θ may not be identified in the presence of classical measurement error, in Supplementary Appendix S9.3, partial identification bounds are derived for θ.

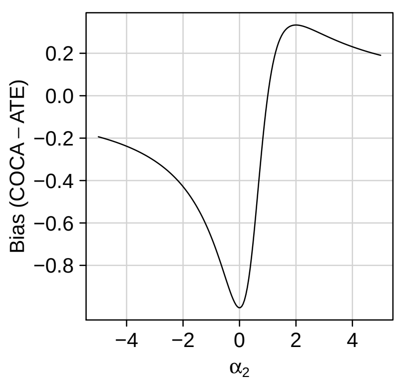

Figure 9.3: Illustration of the potential impact of violating the the assumption that the NCO and exposure of interest are independent given counterfactual outcome $Y(0)$.

9.3.2 *Difference-in-difference approach*

Identification

The difference-in-difference approach (DiD) proposed by Sofer et al. (2016) is an alternative approach to the COCA and does not assume rank preservation, nor does it require that the counterfactual outcome $Y(0)$ renders the NCO and exposure of interest conditionally independent. Instead, the approach relies on bias equivalence for the primary exposure-outcome relation and the exposure-NCO relation. The simplest version of the DiD approach identifies the ATT under additive equi-confounding, as stated in Theorem 9.4, via the difference between the crude difference in primary outcome means and the bias of the exposure-NCO relation.

Theorem 9.4 (Difference-in-difference approach for the ATT under additive equi-confounding)**.** *Suppose that the following conditions hold for all levels* $a = 0, 1$*:*
- Consistency: $Y(a) = Y$ *if* $a = A$*.*
- Additive equi-confounding:
$\mathbb{E}[Y(0)|A = 1] - \mathbb{E}[Y(0)|A = 0] = \mathbb{E}[N|A = 1] - \mathbb{E}[N|A = 0]$*.*

Then, $\mathbb{E}[Y(1) - Y(0)|A = 1] = (\mathbb{E}[Y|A = 1] - \mathbb{E}[Y|A = 0]) - (\mathbb{E}[N|A = 1] - \mathbb{E}[N|A = 0])$*.*

Additive equi-confounding is relatively easy to interpret. However, the assumption may be particularly likely to be violated when primary outcome Y and NCO Z are measured on different scales. A generalized DiD approach still identifies the ATT under a different constraint on the dependence between $Y(0)$ and A in relation to the dependence between N and A. In particular, Theorem 9.5, based on Sofer et al. (2016), relies on quantile-quantile equi-confounding, an example of which is depicted in Figure 9.4.

Theorem 9.5 (Generalized difference-in-difference approach for the ATT under quantile-qualine equi-confounding)**.** *Suppose that the following conditions hold for all levels* $a = 0, 1$*:*
- Consistency: $Y(a) = Y$ *if* $a = A$*.*
- Quantile-quantile equi-confounding: $F_0(F_1^{-1}(p)) = G_0(G_1^{-1}(p))$ *for all* $p \in [0, 1]$*, where* $F_a(y) = \Pr(Y(0) \leq y|A = a)$*,* $F_a^{-1}(p) = \min\{y : p \leq F_a(y)\}$*,* $G_a(z) = \Pr(Z \leq z|A = a)$*,* $G_a^{-1}(p) = \min\{z : p \leq G_a(z)\}$*.*
- F_1 *is strictly increasing.*

Then, $\mathbb{E}[Y(1) - Y(0)|A = 1] = \mathbb{E}[Y|A = 1] - \mathbb{E}[F_0^{-1}(G_0(G_1^{-1}(V)))]$*, where* $V \sim \text{Uniform}[0, 1]$*.*

Sensitivity to assumption violations

We now give a simple setting where neither additive nor quantile-quantile equi-confounding is guaranteed to hold. The setting is characterised by two common causes U_1, U_2 of the primary exposure and outcome and of the NCO. As before, we let the relative effects of these common causes to differ between exposure, primary outcome and NCO, and we suppose that the following models hold:

$$\left.\begin{aligned} A &\sim \text{Bernoulli}(p_A), \\ U_1|A &\sim \text{Normal}(\alpha_0 + \alpha_1 A, \sigma_1^2), \\ U_2|(U_1, A) &\sim \text{Normal}(\alpha_0' + \alpha_1' A, \sigma_2^2), \\ Y(0)|(U_1, U_2, A) &\sim \text{Normal}(U_1 + U_2, \sigma_Y^2), \\ Y = Y(A) &= Y(0) + \theta A,\ \theta \text{ constant}, \\ Z|(U_1, U_2, A, Y(0)) &\sim \text{Normal}(\beta_0 + \beta_1 U_1 + \beta_2 U_2, \sigma_Z^2). \end{aligned}\right\} \tag{9.5}$$

Parameters $\alpha_1, \alpha_1', \beta_1, \beta_2$ control the dependence (confounding), through U_1 and U_2, between A and $Y(0)$ and between A and NCO Z; in the special case where these parameters take the value 0, there is no confounding. The models of (9.5) imply

$$Y(0)|A \sim \text{Normal}((\alpha_0 + \alpha_0') + (\alpha_1 + \alpha_1')A, \sigma_1^2 + \sigma_2^2 + \sigma_Y^2),$$

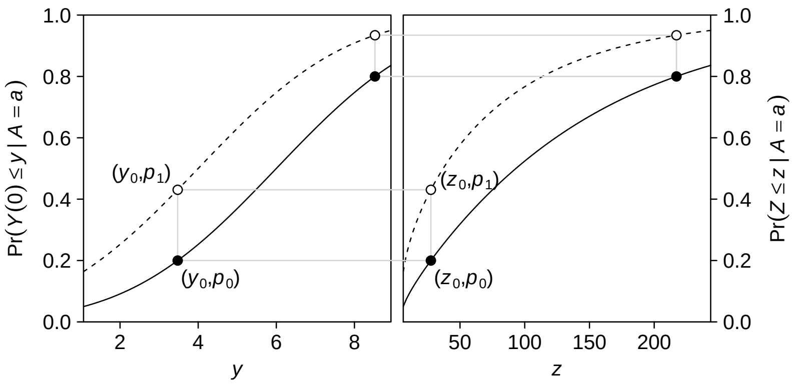

Figure 9.4: Example of quantile-quantile equi-confounding. Dashed curves represents $a = 1$, solid curves $a = 0$. There is quantile-quantile equi-confounding because for every two points (y_0, p_0) and (y_0, p_1) on the solid and dashed curves, respectively, of the left panel, there exists z_0 such that (z_0, p_0) and (z_0, p_1) lie on the solid and dashed curves, respectively, of the right panel; quantiles y_0 and z_0 need not be the same.

$$Y|A \sim \text{Normal}((\alpha_0 + \alpha_0') + (\alpha_1 + \alpha_1' + \theta)A, \sigma_1^2 + \sigma_2^2 + \sigma_Y^2),$$
$$Z|A \sim \text{Normal}((\beta_0 + \beta_1\alpha_0 + \beta_2\alpha_0') + (\beta_1\alpha_1 + \beta_2\alpha_1')A), \beta_1^2\sigma_1^2 + \beta_2^2\sigma_2^2 + \sigma_Z^2).$$

Implementing the DiD for the ATT θ would therefore identify, under (9.5), the quantity

$$\begin{aligned}&(\mathbb{E}[Y|A=1] - \mathbb{E}[Y|A=0]) - (\mathbb{E}[N|A=1] - \mathbb{E}[N|A=0])\\ &\quad= (1-\beta_1)\alpha_1 + (1-\beta_2)\alpha_1' + \theta,\end{aligned}$$

with a bias of $(1-\beta_1)\alpha_1 + (1-\beta_2)\alpha_1'$. The generalised DiD would instead identify

$$\begin{aligned}&\mathbb{E}[Y|A=1] - \mathbb{E}[F_0^{-1}(G_0(G_1^{-1}(V)))]\\ &\quad= (\alpha_0 + \alpha_0') + (\alpha_1 + \alpha_1' + \theta) - \int_{-\infty}^{+\infty} F_0^{-1}(G_0(G_1^{-1}(p)))\ \mathrm{d}p,\end{aligned}$$

where G_1^{-1} is the quantile function of the associated with the distribution of $Z|A=1$, G_0 is the cumulative distribution function for $Z|A=1$ and F_0^{-1} the quantile function of $Y|A=0$.

Figure 9.5 shows, for various parameter specifications, the bias of the (generalised) DiD for the ATT θ. Specifically, β_1 was varied over $(-2, 2)$ and α_1' over $\{0, 1\}$, while β_2 was set to $2 - \beta_1$, and $p_A = 0.5$, $\alpha_0, \alpha_0', \beta_0, \theta = 0$ and $\alpha_1 = 1, \sigma_1^2, \sigma_2^2, , \sigma_Y^2, \sigma_Z^2 = 1$. The figure illustrates that under additive and quantile-quantile equi-confounding the DiD and generalised DiD, respectively, identify the ATT. It also shows that both approaches are sensitive—albeit differently—to violations of their respective assumptions. Interestingly, even in the absence of additive equi-confounding the generalised DiD could be subject to considerable bias (Figure 9.5, right panel, where the bias for the DiD is $(1-\beta_1)\alpha_1 + (1-\beta_2)\alpha_2' = 2 - (\beta_1 + \beta_2) = 0$). Beside the interpretability of its assumptions, an appealing property of the standard DiD approach is that the effects of common causes need not be the same for the NCO and primary outcome of interest; if the net additive confounding is (close to) the same for the NCO and primary outcome, then the ATT may be (nearly) identified.

9.3.3 Double-negative control approach

Identification

Recent developments on the use of negative controls to adjust for unmeasured confounding leverage multiple negative control variables or proxies of unmeasured common causes (Miao et al., 2018a,b; Shi et al., 2020a,b; Tchetgen et al., 2020).

For example, the next result, due to Miao et al. (2018b), gives a set of conditions sufficient to identify the expected marginal counterfactual outcome $\mathbb{E}[Y(a)]$ by leveraging a pair of proxy variables B, Z of an unobserved variable U that renders the counterfactual outcomes independent of the exposure of interest (i.e., conditional exchangeability given U).

Theorem 9.6 (The confounding bridge approach)**.** *Suppose that for all levels a of A, the following conditions hold:*
- Consistency: $Y(a) = Y$ *if* $a = A$.
- Positivity: $0 < \Pr(A = a|B) < 1$ *with probability 1.*
- Latent ignorability: $Y(a) \perp\!\!\!\perp (A, B)|U$ *and* $Z \perp\!\!\!\perp (A, B)|U$.
- Confounding bridge assumption: $\mathbb{E}[Y|A = a, U] = \mathbb{E}[h(Z)|A = a, U]$ *with probability 1 for some h.*
- Completeness: *for all g, if* $\mathbb{E}[g(Z)|A = a, B] = 0$ *with probability 1, then* $\Pr(g(Z) = 0|A = a) = 1$.

Let $\mathcal{H}(a)$ be the collection of all h that satisfy $\mathbb{E}[Y - h(Z)|A = a, B] = 0$ with probability 1. Then, $\mathcal{H}(a)$ is non-empty, and for all $h \in \mathcal{H}(a)$, $\mathbb{E}[Y(a)] = \mathbb{E}[h(Z)]$.

Figure 9.6 shows a directed acyclic graph that is consistent with the assumptions of Theorem 9.6. The proxy variables can be seen to be negative control variables in the sense of Shi et al. (2020b), thus making the confounding bridge approach a (double-)negative control approach. Like the primary

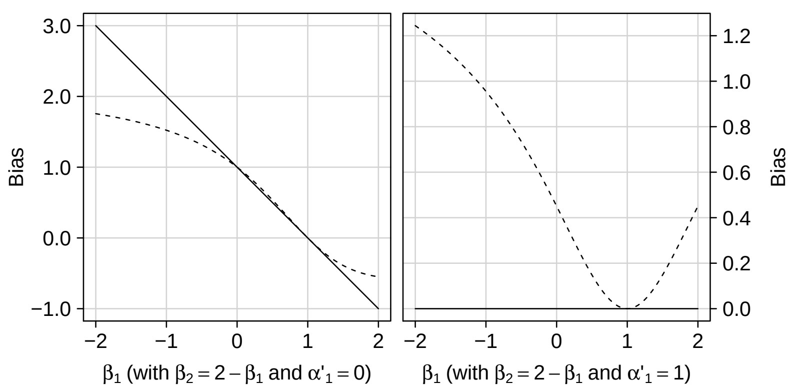

Figure 9.5: Illustrating of the potential impact of violating additive or quantile-quantile equi-confounding on the bias of the (generalised) difference-in-difference approach. Solid lines represent the difference-in-difference approach; dashed lines the generalised difference-in-difference.

exposure-outcome association, the exposure-NCO association is confounded by U. The function h is referred to as a confounding bridge because the confounding bridge assumption indicates that it links the Y-U association with the NCO-U association. The NCE is not part of this link but is meant to help identify it.

The confounding bridge and completeness assumptions can be difficult to grasp. For categorical variables, however, the assumptions are subsumed under the conditions of the next result, due to Miao et al. (2018a) and Shi et al. (2020a).

Theorem 9.7 (The proximal g-formula for categorical variables)**.** *Let U, B, Z be discrete random variables with finite support such that U has no more categories than B or Z. Suppose that for all levels a of A, the following conditions hold:*

- Consistency: $Y(a) = Y$ *if* $a = A$.
- Positivity: $0 < \Pr(A = a, B = b)$ *for all categories* b *of* B.
- Latent ignorability: $Y(a) \perp\!\!\!\perp (A, B)|U$ *and* $Z \perp\!\!\!\perp (A, B)|U$.
- Full rank: $\Pr(\boldsymbol{Z}|\boldsymbol{U})$ *and* $\Pr(\boldsymbol{U}|A = a, \boldsymbol{B})$ *have rank equal to the number of levels of* U.

Then, $\mathbb{E}[Y(a)] = h(\boldsymbol{Z})\Pr(\boldsymbol{Z})$*, where* $h(\boldsymbol{Z}) = \mathbb{E}[Y|A = a, \boldsymbol{B}]\Pr(\boldsymbol{Z}|A = a, \boldsymbol{B})^{-1}$.

Here, following Miao et al. (2018a), for any categorical variables X, Y, Z, $\Pr(\boldsymbol{X}|Y, \boldsymbol{Z})$ denotes the matrix of probabilities $\Pr(X = x|Y, \boldsymbol{Z})$ with a one-to-one correspondence between rows and categories x of X and a one-to-one correspondence between columns and categories z of Z. Interestingly, the proximal g-formula can also be written as a weighted version of the standard

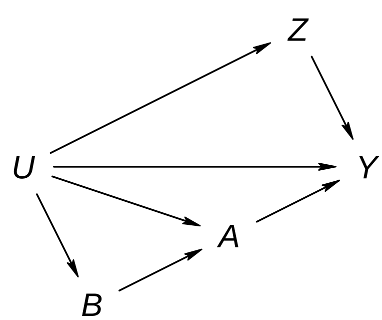

Figure 9.6: Causal directed acyclic graph with negative control pair satisfying the latent ignorability condition of Theorem 9.6.

g-formula:

$$\mathbb{E}[Y|A=a,\boldsymbol{B}]\mathrm{diag}(\boldsymbol{W}(a))\Pr(\boldsymbol{B})$$

with weights $\boldsymbol{W}(a) = (\mathrm{diag}\Pr(\boldsymbol{B}))^{-1}\Pr(\boldsymbol{Z}|A=a,\boldsymbol{B})^{-1}\Pr(\boldsymbol{Z})$ and $\mathrm{diag}(\boldsymbol{W}(a))$ and $\mathrm{diag}(\boldsymbol{B})$ denoting the diagonal matrices with main diagonals $\boldsymbol{W}(a)$ and $\boldsymbol{B}$, respectively. In the case that proxy variables B and Z are binary, the expression simplifies to

$$\mathbb{E}\{\mathbb{E}[WY|A=a,B]\}$$

with weights

$$\begin{aligned} W = &\frac{(1-B)}{\Pr(B=0)}\frac{\Pr(Z=1|A,B=1)-\Pr(Z=1)}{\Pr(Z=1|A,B=1)-\Pr(Z=1|A,B=0)} \\ &+\frac{-B}{\Pr(B=1)}\frac{\Pr(Z=1|A,B=0)-\Pr(Z=1)}{\Pr(Z=1|A,B=1)-\Pr(Z=1|A,B=0)}. \end{aligned}$$

Sensitivity to assumption violations

Theorem 9.7 can accommodate any number of categories of U by taking proxy variables with sufficiently many categories, e.g., by combining sufficiently many proxies. However, upon increasing the number of proxy variables, the latent ignorability assumption becomes more difficult to satisfy in the sense that $Y(a)$ must be independent of increasingly many proxies given A and U. In this subsection, we consider the sensitivity of the proximal g-formula for violations of latent ignorability as well as of the assumption that U has no more categories than the proxy variables.

In particular, we consider the case where the variables A, Y of interest and the proxy variables B, Z are binary, where U is a pair (U_1, U_2) of independent binary variables, and where the following models hold:

$$\begin{aligned} U_1 &\sim \mathrm{Bernoulli}(1/2), \\ U_2|U_1 &\sim \mathrm{Bernoulli}(\rho), \\ B|U_1,U_2 &\sim \mathrm{Bernoulli}(\mathrm{expit}\{\alpha_0+U_1+U_2\}), \\ A|U_1,U_2,B &\sim \mathrm{Bernoulli}(\mathrm{expit}\{\beta_0+U_1+\beta_1U_2+B\}), \\ Z|U_1,U_2,B,A &\sim \mathrm{Bernoulli}(\mathrm{expit}\{\gamma_0+U_1-1/2U_2+\gamma_1A\}), \\ Y|U_1,U_2,B,A,Z &\sim \mathrm{Bernoulli}(\mathrm{expit}\{\theta_0+U_1+U_2+Z+\theta_1B\}), \end{aligned}$$

where $\mathrm{expit}(x) = 1/(1+\exp[-x])$ for all x. Intercepts $\alpha_0, \beta_0, \gamma_0, \theta_0$ were chosen to ensure that $\Pr(B=1) = \Pr(A=1) = 1/2$ and $\Pr(Z=1) = \Pr(Y=1) = 1/5$.

We let $\rho = 0, \beta_1 = 1, \gamma_1 = 0, \theta_1 = 0$ by default. In scenario A, instead of taking $\beta_1 = 1, \rho = 0$, we vary β_1 over $(-4, 4)$ under $\rho = 1/2$ to violate the full rank assumption, which implies that U has no more categories than B or Z. In scenario B, instead of taking $\gamma_1 = 0$, we violate the latent ignorability assumption by varying γ_1 over $(-4, 4)$ (i.e., Z is not a negative control outcome). In scenario C, we violate the same assumption, now by varying θ_1 over $(-4, 4)$ (i.e., B is not a negative control exposure) instead of taking $\theta_1 = 0$.

Figure 9.7 gives the bias of the proximal g-formula for the ATE $\mathbb{E}[Y(1)-Y(0)]$ for all scenarios. Also shown are the differences between the crude risk differences $\mathbb{E}[Y|A = 1] - \mathbb{E}[Y|A = 0]$ and the ATE. The bias is zero under the default parameters, which are consistent with the assumptions of Theorem 9.7. The figure also illustrates that violations of these unverifiable assumptions can have a large impact on the validity of the double-negative control approach.

In an other study, Vlassis et al. (2020) found bias of the crude risk difference to be consistently smaller than that of the proximal g-formula. Our results demonstrate that in some settings, the proximal g-formula results in considerably more bias than what would result from ignoring unmeasured confounding.

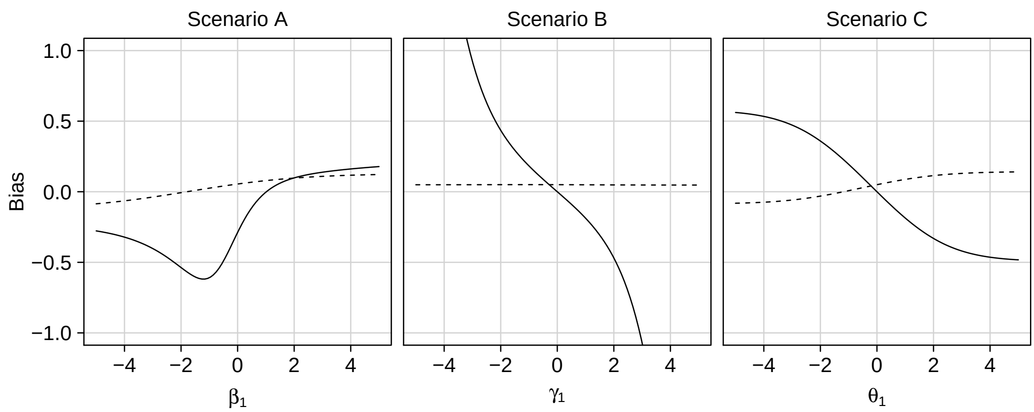

Figure 9.7: Bias of crude approach (dashed) and proximal g-formula (solid) under violations of the cardinality assumption (Scenario A), negative control outcome condition (Scenario B), or negative control exposure condition (Scenario C).

9.4 Conclusion

Negative controls have gained increasing interest in addressing concerns about the validity of a study. The literature on the topic has tended to consider increasingly ambitious tasks, from confounding detection to full identification of causal effects, typically at the cost of stronger and more complex assumptions. Efforts have been made to introduce negative controls to a broader audience and ensure they are adopted in epidemiological practice (Shi et al., 2020b). However, little attention has yet been given to the methods' assumptions and potential impact of assumptions violations. While the assumptions may be tenable enough in some specific cases to justify an application, in others substantial violations are possible. We have illustrated that assumption violations, some of which are likely even in very simple settings, may have a considerable impact on the validity of the negative control approach, thereby limiting its utility. Despite the possible abundance of negative controls, their routine use in epidemiological practice may fail to strengthen evidence about exposure-outcome effects unless it can be safely assumed that assumption violations are absent or else if the robustness against these violations is well understood. Given the potential impact of assumption violations, it may sometimes be desirable to replace strong conditions for identification with weaker conditions that are easier to verify, even when these weaker conditions imply at most partial identification. Future research in this area may broaden the applicability of negative controls and in turn make them more suited for routine use in epidemiological practice. When they are used, we advise that researches consider the results of their applications carefully and explicitly in light of the methods' limitations and assumptions.

References

Albert, A. and J. Anderson (1984): "On the existence of maximum likelihood estimates in logistic regression models," *Biometrika*, 71, 1–10.

Arnold, B. F., A. Ercumen, J. Benjamin-Chung, and J. M. Colford Jr (2016): "Brief report: negative controls to detect selection bias and measurement bias in epidemiologic studies," *Epidemiology (Cambridge, Mass.)*, 27, 637.

Birch, M. (1964): "The detection of partial association, I: the 2× 2 case," *Journal of the Royal Statistical Society. Series B (Methodological)*, 313–324.

Flanders, W. D., M. Klein, L. A. Darrow, M. J. Strickland, S. E. Sarnat, J. A. Sarnat, L. A. Waller, A. Winquist, and P. E. Tolbert (2011): "A method for detection of residual confounding in time-series and other observational studies," *Epidemiology (Cambridge, Mass.)*, 22, 59.

Groenwold, R. H. (2013): "Falsification end points for observational studies," *JAMA*, 309, 1769–1771.

Hernán, M. and J. Robins (2020): *Causal Inference: What If*, Boca Raton: Chapman & Hall/CRC.

Lipsitch, M., E. Tchetgen Tchetgen, and T. Cohen (2010): "Negative controls: a tool for detecting confounding and bias in observational studies," *Epidemiology*, 21, 383–388.

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

Miao, W., X. Shi, and E. Tchetgen Tchetgen (2018b): "A confounding bridge approach for double negative control inference on causal effects," *arXiv e-prints*, arXiv–1808.

Rosenbaum, P. (1989): "The role of known effects in observational studies," *Biometrics*, 45, 557–569.

Shi, X., W. Miao, J. C. Nelson, and E. J. Tchetgen Tchetgen (2020a): "Multiply robust causal inference with double-negative control adjustment for categorical unmeasured confounding," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82, 521–540.

Shi, X., W. Miao, and E. Tchetgen Tchetgen (2020b): "A selective review of negative control methods in epidemiology," *Current Epidemiology Reports*, 1–13.

Sofer, T., D. B. Richardson, E. Colicino, J. Schwartz, and E. J. Tchetgen Tchetgen (2016): "On negative outcome control of unobserved confounding as a generalization of difference-in-differences," *Statistical science: a review journal of the Institute of Mathematical Statistics*, 31, 348.

Tchetgen, E. J. T., A. Ying, Y. Cui, X. Shi, and W. Miao (2020): "An introduction to proximal causal learning," *arXiv preprint arXiv:2009.10982.*

Tchetgen Tchetgen, E. (2013): "The control outcome calibration approach for causal inference with unobserved confounding," *American Journal of Epidemiology*, 179, 633–640.

Vlassis, N., P. Hebda, S. McBride, and A. Noulas (2020): "On proximal causal learning with many hidden confounders," *arXiv preprint arXiv:2012.06725*.

Supplementary Material

S9.1 Identifiability of the direction of bias using an NCO/NCE

Theorem (Identification of the direction of bias using an NCO/NCE)**.** *Suppose the following conditions hold:*

- Latent ignorability for some scalar U: $Z \perp\!\!\!\perp (A, Y)|U$ *and* $Y \perp\!\!\!\perp A|U$.
- Primary exposure model: $A = \alpha_0 + \alpha_1 U + \epsilon$, $\epsilon \perp\!\!\!\perp U$, $\mathbb{E}[\epsilon] = 0$.
- Primary outcome model: $Y = \gamma_0 + \gamma_1 U + \theta A + \varepsilon$, $\varepsilon \perp\!\!\!\perp (A, U)$, $\mathbb{E}[\varepsilon] = 0$.
- NCO/NCE model: $Z = \beta_0^* + \beta_1^* U + \delta$, $\delta \perp\!\!\!\perp (A, Y, U), \mathbb{E}[\delta] = 0$.

Then, $\hat{\theta} - \theta$ *has the same sign as*

$$\frac{\mathrm{Cov}(Y, Z)}{\mathrm{Cov}(A, Z)} - \hat{\theta}.$$

Proof. For the ordinary least squares coefficient $\hat{\theta} = \mathrm{Cov}(Y, A)/\mathrm{Var}(A)$ in the regression of Y on A, we have

$$\begin{aligned} \hat{\theta} - \theta &= \gamma_1 \frac{\mathrm{Cov}(U, A)}{\mathrm{Var}(A)} && \text{(by the primary outcome model)} \\ &= \gamma_1 \alpha_1 \frac{\mathrm{Var}(U)}{\mathrm{Var}(A)}. && \text{(by the primary exposure model)} \end{aligned}$$

Note that $\mathrm{Var}(\mathbb{E}[A|U]) = \alpha_1^2 \mathrm{Var}(U)$ and, by the law of total variance, $\mathrm{Var}(A) = \mathrm{Var}(\mathbb{E}[A|U]) + \mathbb{E}[\mathrm{Var}(A|U)]$. Thus, $\mathrm{Var}(A) - \mathbb{E}[\mathrm{Var}(A|U)] = \alpha_1^2 \mathrm{Var}(U)$

$$\hat{\theta} - \theta = \frac{\gamma_1}{\alpha_1} \frac{\mathrm{Var}(A) - \mathbb{E}[\mathrm{Var}(A|U)]}{\mathrm{Var}(A)}.$$

The fraction $(\mathrm{Var}(A) - \mathbb{E}[\mathrm{Var}(A|U)])/\mathrm{Var}(A)$ can be interpreted as the proportion of variance of A that is explained by U. By the law of total variance, the fraction is bounded by 0 and 1.

Next, note that

$$\begin{aligned} \frac{\mathrm{Cov}(Y, Z)}{\mathrm{Cov}(A, Z)} &= \frac{\gamma_1 \mathrm{Cov}(Z, U) + \theta \mathrm{Cov}(A, Z)}{\mathrm{Cov}(A, Z)} && \text{(by the primary outcome model)} \\ &= \gamma_1 \frac{\mathrm{Cov}(Z, U)}{\mathrm{Cov}(A, Z)} + \theta \\ &= \gamma_1 \frac{\mathrm{Cov}(Z, U)}{\alpha_1 \mathrm{Cov}(U, Z)} + \theta && \text{(by the primary exposure model)} \end{aligned}$$

$$= \frac{\gamma_1}{\alpha_1} + \theta.$$

Hence,

$$\hat{\theta} - \theta = \left(\frac{\mathrm{Cov}(Y,Z)}{\mathrm{Cov}(A,Z)} - \theta\right)\frac{\mathrm{Var}(A) - \mathbb{E}[\mathrm{Var}(A|U)]}{\mathrm{Var}(A)},$$

$$\theta = \frac{\hat{\theta} - \lambda\dfrac{\mathrm{Cov}(Y,Z)}{\mathrm{Cov}(A,Z)}}{1-\lambda},$$

$$\hat{\theta} - \theta = \left(\frac{\mathrm{Cov}(Y,Z)}{\mathrm{Cov}(A,Z)} - \hat{\theta}\right)\frac{\lambda}{1-\lambda},$$

where $\lambda = (\mathrm{Var}(A) - \mathbb{E}[\mathrm{Var}(A|U)])/\mathrm{Var}(A)$. Clearly, since $\lambda \in [0,1]$, the sign of the bias $\hat{\theta} - \theta$ is identified by $\mathrm{Cov}(Y,Z)/\mathrm{Cov}(A,Z) - \hat{\theta}$. □

No identifiability of the direction of bias when Z is not an NCE. *Consider the following models*

$$\begin{aligned}
&U \sim \mathrm{Normal}(\mathbb{E}[U], \mathrm{Var}(U)),\\
&A = \alpha_0 + \alpha_1 U + \epsilon,\ \ \epsilon|U \sim \mathrm{Normal}(0, \mathrm{Var}(\epsilon)),\\
&Z = \beta_0^* + \beta_1^* U + \delta,\ \ \delta|(U,A) \sim \mathrm{Normal}(0, \mathrm{Var}(\delta)),\\
&Y = \gamma_0 + \gamma_1 U + \gamma_2 Z + \theta A + \varepsilon,\ \ \varepsilon|(Z,A,U) \sim \mathrm{Normal}(0, \mathrm{Var}(\varepsilon)),
\end{aligned}$$

which are compatible with those of the above theorem if $\gamma_2 = 0$. If $\gamma_2 \neq 0$, then the Latent ignorability condition is violated because $Z \not\perp\!\!\!\perp (A,Y)|U$. If it were possible to infer from the distribution of the observables the direction of bias $\hat{\theta} - \theta$, then there exists some function g of the joint distribution F of (A,Y,Z) such that $g(F)[\hat{\theta} - \theta] > 0$. To prove that this is false, it suffices to show that for some F, the bias $\hat{\theta} - \theta$ may be positive and negative, depending on unobservables, so that for all g, we have $g(F)[\hat{\theta} - \theta] \not> 0$.

Consider the models of the previous section with parameters set to the following values to yield multivariate normal distributions G, H:

	G	H		G	H
$\mathrm{Var}(U)$	1	1	α_1	1	1
$\mathbb{E}[U]$	0	0	β_1^*	1	1
$\mathrm{Var}(\epsilon)$	1	1	γ_1	-5	1
α_0	0	0	θ	1	-1
$\mathrm{Var}(\delta)$	1	1	γ_2	3	1
β_0^*	0	0			
$\mathrm{Var}(\varepsilon)$	1	9			
γ_0	0	0			

Given zero means of A, Y, Z, the corresponding covariance matrices

$$\mathrm{Cov}(G) = \begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 2 & 2 \\ -1 & 0 & 2 & 12 \end{bmatrix}, \qquad \mathrm{Cov}(H) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 2 & 2 \\ 1 & 0 & 2 & 12 \end{bmatrix}$$

imply the same distribution for (A, Y, Z), despite the fact that the true effects θ have opposite signs. This shows that in general the direction of bias cannot be identified.

S9.2 Proofs to theorems in section 9.3

Proof to Theorem 9.1. For all a,

$$\begin{aligned} g^{-1}(\beta_0 + \beta_2 y) &= \mathbb{E}[Z|A=0, Y=y] && \text{(by NCO model)} \\ &= \mathbb{E}[Z|A=0, Y(0)=y] && \text{(by consistency)} \\ &= \mathbb{E}[Z|A=a, Y(0)=y] \\ & \qquad \text{(by exposure-NCO independence given counterfactual outcome)} \\ &= \mathbb{E}[Z|A=a, Y(0)+\theta A - \theta A = y] \\ &= \mathbb{E}[Z|A=a, Y(a) = y + \theta a] && \text{(by rank preservation)} \\ &= \mathbb{E}[Z|A=a, Y = y + \theta a] && \text{(by consistency)} \\ &= g^{-1}(\beta_0 + \beta_1 a + \beta_2(y + \theta a)) \\ &= g^{-1}(\beta_0 + (\beta_1 + \beta_2\theta)a + \beta_2 y), \end{aligned}$$

so that, for $a = 1$,

$$\beta_0 + \beta_2 y = \beta_0 + (\beta_1 + \beta_2\theta) + \beta_2 y,$$

$$\theta = -\beta_1/\beta_2.$$

□

Proof to Theorem 9.2.

$$\begin{aligned} Z &= \beta_0 + \beta_1 Y(0) + \rho(A - \mathbb{E}[A|Y(0)]) + \chi \quad \text{(by linear NCO model)} \\ &= \beta_0 + \beta_1 Y(0) + \rho(A - \alpha_0 - \alpha_1 Y(0)) + \chi \\ &\qquad \text{(by linear exposure model)} \\ &= \beta_0 + \beta_1(Y(A) - \theta A) + \rho(A - \alpha_0 - \alpha_1 Y(A) + \alpha_1 \theta A) + \chi \\ &\qquad \text{(by rank preservation)} \\ &= (\beta_0 - \rho\alpha_0) + (\rho + [\rho\alpha_1 - \beta_1]\theta)A + (\beta_1 - \rho\alpha_1)Y + \chi, \\ &\qquad \text{(by consistency)} \\ \text{and } \mathbb{E}[Z|A,Y] &= (\beta_0 - \rho\alpha_0 + \mathbb{E}[\chi]) + (\rho + [\rho\alpha_1 - \beta_1]\theta)A + (\beta_1 - \rho\alpha_1)Y, \\ &\qquad \text{(by linear NCO model)} \end{aligned}$$

so that

$$\beta_1^* = \rho + (\rho\alpha_1 - \beta_1)\theta \text{ and } \beta_2^* = \beta_1 - \rho\alpha_1,$$

and, in turn, $\theta = (\beta_1^* - \rho)/\beta_2^*$. □

Proof to Theorem 9.3. We have

$$\begin{aligned} &\mathbb{E}[Z|A = 1-a] \\ &\quad = \mathbb{E}\{\mathbb{E}[Z|A = 1-a, Y(a)]|A = 1-a\} \\ &\quad = \mathbb{E}\{\mathbb{E}[Z|A = a, Y(a)]|A = 1-a\} \\ &\qquad\quad \text{(by exposure-NCO independence given counterfactual outcome)} \\ &\quad = \mathbb{E}[Z|A = a, Y = 0]\Pr(Y(a) = 0|A = 1-a) + \\ &\qquad + \mathbb{E}[Z|A = a, Y = 1]\Pr(Y(a) = 1|A = 1-a) \qquad \text{(by consistency)} \\ &\quad = \mathbb{E}[Z|A = a, Y = 0] + \{\mathbb{E}[Z|A = a, Y = 1] - \mathbb{E}[Z|A = a, Y = 0]\} \\ &\qquad \times \Pr(Y(a) = 1|A = 1-a), \end{aligned}$$

so that

$$\Pr(Y(a) = 1|A = 1-a) = \frac{\mathbb{E}[Z|A = 1-a] - \mathbb{E}[Z|A = a, Y = 0]}{\mathbb{E}[Z|A = a, Y = 1] - \mathbb{E}[Z|A = a, Y = 0]}.$$

It follows that

$$\mathbb{E}[Y(a)] = \mathbb{E}[Y(a)|A = a]\Pr(A = a) + \mathbb{E}[Y(a)|A = 1-a]\Pr(A = 1-a)$$

$$
\begin{aligned}
&= \mathbb{E}[Y|A=a]\Pr(A=a) + \mathbb{E}[Y(a)|A=1-a]\Pr(A=1-a) \\
&\hspace{20em}\text{(by consistency)} \\
&= \mathbb{E}[Y|A=a]\Pr(A=a) \\
&\quad + \frac{\mathbb{E}[Z|A=1-a] - \mathbb{E}[Z|A=a,Y=0]}{\mathbb{E}[Z|A=a,Y=1] - \mathbb{E}[Z|A=a,Y=0]}\Pr(A=1-a).
\end{aligned}
$$

□

Proof to Theorem 9.4.

$$
\begin{aligned}
&\mathbb{E}[Y(1)-Y(0)|A=1] \\
&\quad = \mathbb{E}[Y(1)|A=1] - \mathbb{E}[Y(0)|A=1] \\
&\quad = \mathbb{E}[Y|A=1] - \mathbb{E}[Y(0)|A=1] && \text{(by consistency)} \\
&\quad = \mathbb{E}[Y|A=1] - (\mathbb{E}[Y(0)|A=1] - \mathbb{E}[Y(0)|A=0]) - \mathbb{E}[Y(0)|A=0] \\
&\quad = \mathbb{E}[Y|A=1] - (\mathbb{E}[N|A=1] - \mathbb{E}[N|A=0]) - \mathbb{E}[Y(0)|A=0] \\
&\hspace{20em}\text{(by additive equi-confounding)} \\
&\quad = (\mathbb{E}[Y|A=1] - \mathbb{E}[Y|A=0]) - (\mathbb{E}[N|A=1] - \mathbb{E}[N|A=0]) \\
&\hspace{20em}\text{(by consistency)}
\end{aligned}
$$

□

Proof to Theorem 9.5. By quantile-quantile equi-confounding, we have, for all $p \in [0,1]$,

$$
\begin{aligned}
F_0(F_1^{-1}(p)) &= G_0(G_1^{-1}(p)), \\
F_0^{-1}(F_0(F_1^{-1}(p))) &= F_0^{-1}(G_0(G_1^{-1}(p))), \\
F_1^{-1}(p) &= F_0^{-1}(G_0(G_1^{-1}(p))). && \text{(under strictly monotonic } F_1\text{)}
\end{aligned}
$$

Note that the right-hand side of the above equality is a functional of observables because $F_0(y) = \Pr(Y(0) \le y|A=0) = \Pr(Y \le y|A=0)$ by consistency. Now, letting $V \sim \text{Uniform}[0,1]$, we have that $F_1^{-1}(V) \sim Y(0)|A=1$ by the Probability Integral Transform theorem, and so

$$
\mathbb{E}[Y(0)|A=1] = \mathbb{E}[F_0^{-1}(G_0(G_1^{-1}(V)))].
$$

□

Proof to Theorem 9.6. Let h be the function satisfying $\mathbb{E}[Y|A=a,U] = \mathbb{E}[h(Z)|A=a,U]$ with probability 1 (and note that this function exists by the confounding bridge assumption). Let $\mathcal{U} = \{u : \mathbb{E}[Y|A = a, U = u] = \mathbb{E}[h(Z)|A = a, U = u]\}$, so that $\Pr(U \in \mathcal{U}) = 1$ and

$$\begin{aligned}
\mathbb{E}[Y(a)] &= \mathbb{E}[Y(a)|U \in \mathcal{U}] \\
&= \mathbb{E}\{\mathbb{E}[Y(a)|U]|U \in \mathcal{U}\} \\
&= \mathbb{E}\{\mathbb{E}[Y(a)|A = a, U]|U \in \mathcal{U}\} \\
&\qquad\qquad \text{(since } Y(a) \perp\!\!\!\perp A|U \text{ by latent ignorability)} \\
&= \mathbb{E}\{\mathbb{E}[Y|A = a, U]|U \in \mathcal{U}\} \qquad \text{(by consistency)} \\
&= \mathbb{E}\{\mathbb{E}[h(Z)|A = a, U]|U \in \mathcal{U}\} \\
&= \mathbb{E}\{\mathbb{E}[h(Z)|U]|U \in \mathcal{U}\} \qquad \text{(since } Z \perp\!\!\!\perp A|U \text{ by latent ignorability)} \\
&= \mathbb{E}[h(Z)|U \in \mathcal{U}] \\
&= \mathbb{E}[h(Z)].
\end{aligned}$$

Next, note that by the confounding bridge assumption, for all $U \in \mathcal{U}$,

$$\begin{aligned}
\mathbb{E}[Y|A = a, U] &= \mathbb{E}[h(Z)|A = a, U] \\
\mathbb{E}[Y|A = a, B, U] &= \mathbb{E}[h(Z)|A = a, B, U], \qquad \text{(by latent ignorability)}
\end{aligned}$$

so that

$$\begin{aligned}
\mathbb{E}\{\mathbb{E}[Y|A = a, B, U]|A = a, B\} &= \mathbb{E}\{\mathbb{E}[h(Z)|A = a, B, U]|A = a, B\} \\
\mathbb{E}[Y|A = a, B] &= \mathbb{E}[h(Z)|A = a, B], \\
\mathbb{E}[Y - h(Z)|A = a, B] &= 0.
\end{aligned}$$

Let $\mathcal{H}(a)$ be the collection of all h' satisfying $\mathbb{E}[Y - h'(Z)|A = a, B] = 0$ with probability 1. Now, for any $h' \in \mathcal{H}(a)$, we must have

$$\mathbb{E}[h(Z) - h'(Z)|A = a, B] = 0.$$

But from completeness, with $g(Z) = h(Z) - h'(Z)$, it follows that $h(Z) = h'(Z)$ with probability 1. This concludes the proof. $\square$

Proof to Theorem 9.7. Since $Z \perp\!\!\!\perp (A, B)|U$, we have $\Pr(\boldsymbol{Z}|A = a, \boldsymbol{B}) = \Pr(\boldsymbol{Z}|\boldsymbol{U})\Pr(\boldsymbol{U}|A = a, \boldsymbol{B})$. Since matrices $\Pr(\boldsymbol{Z}|\boldsymbol{U})$ and $\Pr(\boldsymbol{U}|A = a, \boldsymbol{B})$ are of full rank, $\Pr(\boldsymbol{Z}|A = a, \boldsymbol{B})$ is of full rank and has left or right inverse $\Pr(\boldsymbol{Z}|A = a, \boldsymbol{B})^{-1}$. Let $h(\boldsymbol{Z}) = \mathbb{E}[Y|A = a, \boldsymbol{B}]\Pr(\boldsymbol{Z}|A = a, \boldsymbol{B})^{-1}$ and observe that

$$h(\boldsymbol{Z}) = \mathbb{E}[Y(a)|A = a, \boldsymbol{B}]\Pr(\boldsymbol{Z}|A = a, \boldsymbol{B})^{-1} \qquad \text{(by consistency)}$$

$$\begin{aligned}
&= \mathbb{E}[Y(a)|\boldsymbol{U}]\Pr(\boldsymbol{U}|A=a,\boldsymbol{B})\Pr(\boldsymbol{Z}|A=a,\boldsymbol{B})^{-1} \\
&\hfill \text{(since } Y(a) \perp\!\!\!\perp (A,B)|U) \\
&= \mathbb{E}[Y(a)|\boldsymbol{U}]\Pr(\boldsymbol{U}|A=a,\boldsymbol{B})[\Pr(\boldsymbol{Z}|\boldsymbol{U})\Pr(\boldsymbol{U}|A=a,\boldsymbol{B})]^{-1} \\
&\hfill \text{(since } Z \perp\!\!\!\perp (A,B)|U) \\
&= \mathbb{E}[Y(a)|\boldsymbol{U}]\Pr(\boldsymbol{U}|A=a,\boldsymbol{B})\Pr(\boldsymbol{U}|A=a,\boldsymbol{B})^{-1}\Pr(\boldsymbol{Z}|\boldsymbol{U})^{-1} \\
&= \mathbb{E}[Y(a)|\boldsymbol{U}]\Pr(\boldsymbol{Z}|\boldsymbol{U})^{-1}.
\end{aligned}$$

It follows that $\mathbb{E}[Y(a)|\boldsymbol{U}] = h(\boldsymbol{Z})\Pr(\boldsymbol{Z}|\boldsymbol{U})$ and in turn $\mathbb{E}[Y(a)] = \mathbb{E}[Y(a)|\boldsymbol{U}]\Pr(\boldsymbol{U}) = h(\boldsymbol{Z})\Pr(\boldsymbol{Z})$, as desired.

□

S9.3 Derivation of expressions in section 9.3.1

S9.3.1 Implications of models (9.1)

Expression of the COCA

An implementation of the COCA by ordinary least squares under the linear NCO model $\mathbb{E}[Z|A,Y] = \beta_0 + \beta_1 A + \beta_2 Y$, identifies the following quantity

$$\begin{aligned}
\hat{\theta} &= -\frac{\hat{\beta}_1}{\hat{\beta}_2} \\
&= -\frac{\mathrm{Cov}(A,Z)\mathrm{Var}(Y) - \mathrm{Cov}(Y,Z)\mathrm{Cov}(A,Y)}{\mathrm{Cov}(Y,Z)\mathrm{Var}(A) - \mathrm{Cov}(A,Z)\mathrm{Cov}(A,Y)},
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathrm{Var}(A)\alpha_1^2 + \sigma_Y^2 + \sigma_\theta^2\mathrm{Var}(A) + \sigma_\theta^2\mathbb{E}[A]^2 + \mathbb{E}[\theta]^2\mathrm{Var}(A) \\
&\quad + 2\mathrm{Var}(A)\alpha_1\mathbb{E}[\theta], \\
\mathrm{Var}(Z) &= \mathrm{Var}(A)\alpha_1^2\gamma_1^2 + \gamma_1^2\sigma_Y^2 + \sigma_Z^2, \\
\mathrm{Cov}(A,Y) &= \mathrm{Var}(A)\alpha_1 + \mathrm{Var}(A)\mathbb{E}[\theta], \\
\mathrm{Cov}(A,Z) &= \mathrm{Var}(A)\alpha_1\gamma_1, \\
\mathrm{Cov}(Y,Z) &= \gamma_1(\mathrm{Var}(A)\alpha_1^2 + \sigma_Y^2 + \mathrm{Var}(A)\alpha_1\mathbb{E}[\theta]).
\end{aligned}$$

Deterministic relation between $\mathbb{E}[\theta]$ and $\mathrm{Var}[\theta]$ given observed data distribution

From the expressions of the variances and covariates above, for arbitrary $\mathbb{E}[\theta], \mathrm{Var}(A)$, it follows that

$$
\begin{aligned}
\alpha_1 &= \frac{\mathrm{Cov}(A,Y) - \mathrm{Var}(A)\mathbb{E}[\theta]}{\mathrm{Var}(A)},\\
\alpha_0 &= \mathbb{E}[Y] - (\alpha_1 + \mathbb{E}[\theta])\mathbb{E}[A],\\
\gamma_1 &= \frac{\mathrm{Cov}(A,Z)}{\mathrm{Var}(A)\alpha_1},\\
\gamma_0 &= \mathbb{E}[Z] - (\alpha_0\gamma_1 + \alpha_1\gamma_1\mathbb{E}[A]),\\
\sigma_Y^2 &= \frac{\mathrm{Cov}(Y,Z) - \gamma_1(\mathrm{Var}(A)\alpha_1^2 + \mathrm{Var}(A)\alpha_1\mathbb{E}[\theta])}{\gamma_1},\\
\sigma_\theta^2 &= \frac{\mathrm{Var}(Y) - [\mathrm{Var}(A)\alpha_1^2 + \sigma_Y^2 + \mathrm{Var}(A)\mathbb{E}[\theta]^2 + 2\mathrm{Var}(A)\alpha_1\mathbb{E}[\theta]]}{\mathrm{Var}(A) + \mathbb{E}[A]^2},\\
\sigma_Z^2 &= \mathrm{Var}(Z) - (\mathrm{Var}(A)\alpha_1^2\gamma_1^2 + \gamma_1^2\sigma_Y^2),
\end{aligned}
$$

provided that $\mathrm{Var}(A), \alpha_1, \gamma_1 \neq 0$, and $\sigma_Y^2, \sigma_\theta^2, \sigma_Z^2 \geq 0$. Note that the right-hand sides of every equality are expressed only in terms of functionals of the available data distribution and the left-hand sides of the equalities above it. It follows that we have a deterministic relationship between $\mathrm{Var}(\theta)$ and $\mathbb{E}[\theta]$ given the observed data distribution of (A, Y, Z). In fact, the relationship is linear:

$$
\begin{aligned}
\sigma_\theta^2 &= \frac{\mathrm{Var}(Y)\mathrm{Cov}(A,Z) - \mathrm{Cov}(A,Y)\mathrm{Cov}(Y,Z)}{(\mathrm{Var}(A) + \mathbb{E}[A]^2)\mathrm{Cov}(A,Z)}\\
&\quad - \frac{\mathrm{Cov}(A,Y)\mathrm{Cov}(A,Z) - \mathrm{Var}(A)\mathrm{Cov}(Y,Z)}{(\mathrm{Var}(A) + \mathbb{E}[A]^2)\mathrm{Cov}(A,Z)}\mathbb{E}[\theta].\\
&= \frac{\mathrm{Var}(A)\mathrm{Var}(Y) - \mathrm{Cov}(A,Y)^2}{(\mathrm{Var}(A) + \mathbb{E}[A]^2)\mathrm{Cov}(A,Z)}(\hat{\beta}_1 - \hat{\beta}_2\mathbb{E}[\theta]).
\end{aligned}
$$

The distribution of $Z|A, Y$

First note that

$$
\begin{aligned}
\mathbb{E}[\theta|Y, A=0] &= \mathbb{E}[\theta|Y(0), A=0] && \text{(by consistency)}\\
&= \mathbb{E}[\theta|A=0] && \text{(since } Y(0) \perp\!\!\!\perp \theta|A)\\
&= \mathbb{E}[\theta]. && \text{(since } \theta \perp\!\!\!\perp A)
\end{aligned}
$$

Next, for arbitrary a, consider $\mathbb{E}[\theta|Y, A = a]$ and note that $(\theta, Y)|A = a$ takes a bivariate normal distribution with means

$$\begin{aligned}
\mathbb{E}[\theta|A = a] &= \mathbb{E}[\theta], \\
\mathbb{E}[Y|A = a] &= \mathbb{E}[Y(A)|A = a] && \text{(by consistency)} \\
&= \mathbb{E}[Y(0) + \theta A|A = a] \\
&= \alpha_0 + \alpha_1 a + \mathbb{E}[\theta],
\end{aligned}$$

variances

$$\begin{aligned}
\mathrm{Var}(\theta|A = a) &= \sigma_\theta^2, && \text{(since } \theta \perp\!\!\!\perp A\text{)} \\
\mathrm{Var}(Y|A = a) &= \mathrm{Var}(Y(A)|A = a) && \text{(by consistency)} \\
&= \mathrm{Var}(Y(0) + \theta A|A = a) \\
&= \mathrm{Var}(Y(0)|A = a) + \mathrm{Var}(\theta) && \text{(since } Y(0) \perp\!\!\!\perp \theta|A \text{ and } \theta \perp\!\!\!\perp A\text{)} \\
&= \sigma_Y^2 + \sigma_\theta^2,
\end{aligned}$$

and correlation

$$\begin{aligned}
&\mathrm{Cor}(Y, \theta|A = a) \\
&\quad = \sqrt{\frac{\mathrm{Cov}^2(Y, \theta|A = a)}{\mathrm{Var}(Y|A = a)\mathrm{Var}(\theta|A = a)}} \\
&\quad = \sqrt{\frac{\mathbb{E}[(Y - \mathbb{E}[Y|A = a])(\theta - \mathbb{E}[\theta|A = a])|A = a]^2}{\mathrm{Var}(Y|A = a)\mathrm{Var}(\theta|A = a)}} \\
&\quad = \sqrt{\frac{\mathbb{E}[(Y(A) - \mathbb{E}[Y(A)|A = a])(\theta - \mathbb{E}[\theta|A = a])|A = a]^2}{\mathrm{Var}(Y|A = a)\mathrm{Var}(\theta|A = a)}} && \text{(by consistency)} \\
&\quad = \sqrt{\frac{\mathbb{E}[(Y(0) + \theta A - \mathbb{E}[Y(0) + \theta A|A = a])(\theta - \mathbb{E}[\theta|A = a])|A = a]^2}{\mathrm{Var}(Y|A = a)\mathrm{Var}(\theta|A = a)}} \\
&\quad = \sqrt{\frac{\begin{array}{l}\mathbb{E}[((Y(0) - \mathbb{E}[Y(0)|A = a]) \\ \quad + (\theta A - \mathbb{E}[\theta|A = a]))(\theta - \mathbb{E}[\theta|A = a])|A = a]^2\end{array}}{\mathrm{Var}(Y|A = a)\mathrm{Var}(\theta|A = a)}} \\
&\quad = \sqrt{\frac{[\mathrm{Cov}(Y(0), \theta|A = a) + \mathrm{Cov}(\theta A, \theta|A = a)]^2}{\mathrm{Var}(Y|A = a)\mathrm{Var}(\theta|A = a)}} \\
&\quad = \sqrt{\frac{a^2\mathrm{Var}(\theta|A = a)^2}{\mathrm{Var}(Y|A = a)\mathrm{Var}(\theta|A = a)}} && \text{(since } Y(0) \perp\!\!\!\perp \theta|A\text{)}
\end{aligned}$$

$$= a\sqrt{\frac{\sigma_\theta^2}{\sigma_Y^2 + \sigma_\theta^2}}.$$

Therefore,

$$\mathbb{E}[\theta|Y, A = a] = \mathbb{E}[\theta] + \sqrt{\frac{\sigma_\theta^2}{\sigma_Y^2 + \sigma_\theta^2}}\frac{\sigma_\theta^2}{\sigma_Y^2}[-(\alpha_0 + \mathbb{E}[\theta])a - \alpha_1 a^2 + aY]$$

(DeGroot and Schervisch, 2012, Theorem 5.10.4, p. 340).

Hence,

$$\mathbb{E}[\theta|Y, A] = \mathbb{E}[\theta] + \sqrt{\frac{\sigma_\theta^2}{\sigma_Y^2 + \sigma_\theta^2}}\frac{\sigma_\theta^2}{\sigma_Y^2}[-(\alpha_0 + \mathbb{E}[\theta])A - \alpha_1 A^2 + AY]$$

and

$$\begin{aligned} Z &= \gamma_0 + \gamma_1 Y(0) + \varepsilon,\ \varepsilon|(A, \theta, Y(0)) \sim \text{Normal}(0, \sigma_Z^2) \\ &= \gamma_0 + \gamma_1(Y(A) - \theta A) + \varepsilon \\ &= \gamma_0 + \gamma_1(Y - \theta A) + \varepsilon && \text{(by consistency)} \\ &= \gamma_0 - \gamma_1\theta A + \gamma_1 Y + \varepsilon, \end{aligned}$$

so that $Z|A, Y$ has a normal distribution with mean

$$\begin{aligned} \mathbb{E}[Z|A, Y] &= \gamma_0 - \gamma_1\mathbb{E}[\theta|Y, A]A + \gamma_1 Y \\ &= \gamma_0 - \gamma_1 A\left[\mathbb{E}[\theta] + \sqrt{\frac{\sigma_\theta^2}{\sigma_Y^2 + \sigma_\theta^2}}\frac{\sigma_\theta^2}{\sigma_Y^2}[-(\alpha_0 + \mathbb{E}[\theta])A - \alpha_1 A^2 + AY]\right] \\ &\quad + \gamma_1 Y \\ &= \gamma_0 - \gamma_1\mathbb{E}[\theta]A + \sqrt{\frac{\sigma_\theta^2}{\sigma_Y^2 + \sigma_\theta^2}}\frac{\sigma_\theta^2}{\sigma_Y^2}[(\alpha_0 + \mathbb{E}[\theta])\gamma_1 A^2 \\ &\quad + \alpha_1\gamma_1 A^3 - \gamma_1 A^2 Y] + \gamma_1 Y \\ &= \beta_0^* + \beta_1^* A + \beta_2^* A^2 + \beta_3^* A^3 + \beta_4^* Y + \beta_5^* A^2 Y, \end{aligned}$$

where

$$\begin{aligned} \beta_0^* &= \gamma_0, \\ \beta_1^* &= -\gamma_1\mathbb{E}[\theta], \\ \beta_2^* &= \sqrt{\frac{\sigma_\theta^2}{\sigma_Y^2 + \sigma_\theta^2}}\frac{\sigma_\theta^2}{\sigma_Y^2}(\alpha_0 + \mathbb{E}[\theta])\gamma_1, \end{aligned}$$

$$\beta_3^* = \sqrt{\frac{\sigma_\theta^2}{\sigma_Y^2 + \sigma_\theta^2}\frac{\sigma_\theta^2}{\sigma_Y^2}}\alpha_1\gamma_1,$$
$$\beta_4^* = \gamma_1,$$
$$\beta_5^* = -\sqrt{\frac{\sigma_\theta^2}{\sigma_Y^2 + \sigma_\theta^2}\frac{\sigma_\theta^2}{\sigma_Y^2}}\gamma_1,$$

so $\mathbb{E}[\theta] = -\beta_1^*/\beta_4^*$ if $\beta_4^* \neq 0$. Therefore, with a continuous primary outcome and non-binary exposure, the rank preservation assumption can sometimes be dropped whilst maintaining identifiability. If A is binary, we have $\mathbb{E}[Z|A,Y] = \beta_0^* + (\beta_1^* + \beta_2^* + \beta_3^*)A + \beta_4^* Y + \beta_5^* AY$, where

$$\begin{aligned}(\beta_1^* + \beta_2^* + \beta_3^*) &= \gamma_1\sqrt{\frac{\sigma_\theta^2}{\sigma_Y^2 + \sigma_\theta^2}\frac{\sigma_\theta^2}{\sigma_Y^2}}(\alpha_0 + \alpha_1) + \gamma_1\left(\sqrt{\frac{\sigma_\theta^2}{\sigma_Y^2 + \sigma_\theta^2}\frac{\sigma_\theta^2}{\sigma_Y^2}} - 1\right)\mathbb{E}[\theta] \\ &= -\beta_5^*(\alpha_0 + \alpha_1) - (\beta_4^* + \beta_5^*)\mathbb{E}[\theta].\end{aligned}$$

This suggests a test for violations of rank preservation since the interaction term coefficient β_5^* is zero if and only if $\mathrm{Var}(\theta) = 0$ or $\beta_4^* = 0$. Provided that $\beta_4^* \neq 0$, a valid test of the null hypothesis $\beta_5^* = 0$ is thus a valid test of rank preservation under the above models.

S9.3.2 Implications of models (9.3)

Under models 9.3, we have the following variances and covariances:

$$\begin{aligned}\mathrm{Var}(A) &= \alpha_1^2\mathrm{Var}(U_1) + \alpha_2^2\mathrm{Var}(U_2) + \sigma_A^2, \\ \mathrm{Var}(Y) &= (1 + \theta\alpha_1)^2\mathrm{Var}(U_1) + (1 + \theta\alpha_2)^2\mathrm{Var}(U_2) + \theta^2\sigma_A^2, \\ \mathrm{Var}(Z) &= (\alpha_1')^2\mathrm{Var}(U_1) + (\alpha_2')^2\mathrm{Var}(U_1) + \sigma_Z^2, \\ \mathrm{Cov}(A,Y) &= (1 + \theta\alpha_1)\alpha_1\mathrm{Var}(U_1) + (1 + \theta\alpha_2)\alpha_2\mathrm{Var}(U_2) + \theta\sigma_A^2, \\ \mathrm{Cov}(A,Z) &= \alpha_1\alpha_1'\mathrm{Var}(U_1) + \alpha_2\alpha_2'\mathrm{Var}(U_2), \\ \mathrm{Cov}(Y,Z) &= (1 + \theta\alpha_1)\alpha_1'\mathrm{Var}(U_1) + (1 + \theta\alpha_2)\alpha_2'\mathrm{Var}(U_2)\end{aligned}$$

and means

$$\begin{aligned}\mathbb{E}[A] &= \alpha_0 + \alpha_1\mathbb{E}[U_1] + \alpha_2\mathbb{E}[U_2], \\ \mathbb{E}[Y] &= \theta\alpha_0 + (1 + \theta\alpha_1)\mathbb{E}[U_1] + (1 + \theta\alpha_2)\mathbb{E}[U_2], \\ \mathbb{E}[Z] &= \alpha_0 + \alpha_1\mathbb{E}[U_1] + \alpha_2\mathbb{E}[U_2].\end{aligned}$$

S9.3.3 Partial identification in the presence of classical measurement error in the outcome

Theorem. *Suppose the following conditions hold:*
- Rank preservation: $Y(A) = Y(0) + \theta A$, θ *constant.*
- Exposure-NCO independence given counterfactual outcome: $Z \perp\!\!\!\perp A | Y(0)$.
- NCO model: $Z = \beta_0^* + \beta_1^* Y(0) + \varepsilon$, $\varepsilon \perp\!\!\!\perp (A, Y(0)), \mathbb{E}[\varepsilon] = 0$.
- Classical measurement error: $Y = Y(A) + U$, $U \perp\!\!\!\perp (A, Y(0), Z)$, $\mathbb{E}[U] = 0$.

Then,

$$\theta \in \left[\hat{\theta}, \hat{\theta}\left(1 - R^2 \frac{1}{1 - \mathrm{Cor}^2(A,Y)}\right) + R^2 \frac{\mathrm{Var}(Y)}{\mathrm{Cov}(A,Y)}\left(1 - \frac{1}{1 - \mathrm{Cor}^2(A,Y)}\right)\right],$$

where $R^2 = 1 - \mathbb{E}[\mathrm{Var}(Y|A)]/\mathrm{Var}(Y)$ *is the proportion of variance of* Y *explained by* A*, and* $\hat{\theta} = -\hat{\beta}_1/\hat{\beta}_2$ *and* $\hat{\beta}_1$ *and* $\hat{\beta}_1$ *are the ordinary least squares coefficients for* A *and* Y *in a linear regression of* Z *on* A *and* Y.

Proof. We have that

$$\begin{aligned} Z &= \beta_0^* + \beta_1^* Y(0) + \varepsilon && \text{(by NCO model)} \\ &= \beta_0^* + \beta_1^*(Y(A) - \theta A) + \varepsilon && \text{(by rank preservation)} \\ &= \beta_0^* + \beta_1^*(Y - U - \theta A) + \varepsilon && \text{(under classical measurement error)} \\ &= \beta_0^* + \beta_1^* Y - \beta_1^* U - \beta_1^* \theta A + \varepsilon, \end{aligned}$$

where $\varepsilon \perp\!\!\!\perp (Y, A, U)$ (since $U \perp\!\!\!\perp \varepsilon | (A, Y(0))$ and $\varepsilon \perp\!\!\!\perp (A, Y(0))$, so that $\varepsilon \perp\!\!\!\perp (Y(0), A, U)$).

Now, let

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(A,Z)\mathrm{Var}(Y) - \mathrm{Cov}(Y,Z)\mathrm{Cov}(A,Y)}{\mathrm{Var}(A)\mathrm{Var}(Y) - \mathrm{Cov}^2(A,Y)},$$
$$\hat{\beta}_2 = \frac{\mathrm{Cov}(Y,Z)\mathrm{Var}(A) - \mathrm{Cov}(A,Z)\mathrm{Cov}(A,Y)}{\mathrm{Var}(A)\mathrm{Var}(Y) - \mathrm{Cov}^2(A,Y)},$$

the ordinary least squares coefficients in a linear regression of Z on A and Y. We have

$$\mathrm{Cov}(A,Z) = \beta_1^*(\mathrm{Cov}(A,Y) - \theta \mathrm{Var}(A)),$$
$$\mathrm{Cov}(Y,Z) = \beta_1^*(\mathrm{Var}(Y) - \mathrm{Var}(U) - \theta \mathrm{Cov}(A,Y)),$$

so that

$$\hat{\beta}_1 = \beta_1^* \left(\frac{\text{Cov}(A,Y)\text{Var}(U)}{\text{Var}(A)\text{Var}(Y) - \text{Cov}^2(A,Y)} - \theta \right),$$

$$\hat{\beta}_2 = \beta_1^* \left(1 - \frac{\text{Var}(A)\text{Var}(U)}{\text{Var}(A)\text{Var}(Y) - \text{Cov}^2(A,Y)} \right)$$

and in turn

$$\begin{aligned}
\hat{\theta} &= -\frac{\hat{\beta}_1}{\hat{\beta}_2} \\
&= -\frac{\text{Cov}(A,Y)\text{Var}(U) - \theta(\text{Var}(A)\text{Var}(Y) - \text{Cov}^2(A,Y))}{\text{Var}(A)\text{Var}(U) - (\text{Var}(A)\text{Var}(Y) - \text{Cov}^2(A,Y))}, \\
\theta &= \hat{\theta}\left(1 - \frac{\text{Var}(A)\text{Var}(U)}{\text{Var}(A)\text{Var}(Y) - \text{Cov}^2(A,Y)} \right) + \frac{\text{Cov}(A,Y)\text{Var}(U)}{\text{Var}(A)\text{Var}(Y) - \text{Cov}^2(A,Y)} \\
&= \hat{\theta}\left(1 - \frac{\text{Var}(U)}{\text{Var}(Y)} \frac{1}{1 - \text{Cor}^2(A,Y)} \right) \\
&\quad + \frac{\text{Var}(U)}{\text{Var}(Y)} \frac{\text{Var}(Y)}{\text{Cov}(A,Y)} \left(1 - \frac{1}{1 - \text{Cor}^2(A,Y)} \right).
\end{aligned}$$

By the law of total (conditional) variance,

$$\begin{aligned}
\text{Var}(Y) &= \mathbb{E}[\text{Var}(Y|A)] + \text{Var}(\mathbb{E}[Y|A]) \\
&= \mathbb{E}[\text{Var}(Y|A, Y(0))|A] + \mathbb{E}[\text{Var}(\mathbb{E}[Y|A, Y(0)])|A] + \text{Var}(\mathbb{E}[Y|A]) \\
&= \text{Var}(U) + \mathbb{E}[\text{Var}(\mathbb{E}[Y|A, Y(0)])|A] + \text{Var}(\mathbb{E}[Y|A]).
\end{aligned}$$

Now, define $R^2 = (\text{Var}(Y) - \mathbb{E}[\text{Var}(Y|A)])/\text{Var}(Y)$, the proportion of variance of Y explained by A and observe that

$$R^2 \geq \frac{\text{Var}(U)}{\text{Var}(Y)} \geq 0.$$

Next, define

$$\tilde{\theta}(\lambda) = \hat{\theta}\left(1 - \lambda \frac{1}{1 - \text{Cor}^2(A,Y)} \right) + \lambda \frac{\text{Var}(Y)}{\text{Cov}(A,Y)} \left(1 - \frac{1}{1 - \text{Cor}^2(A,Y)} \right)$$

and note that, because the first derivative of $\tilde{\theta}$ is invariant to changes in λ, $\tilde{\theta}$ is monotonic. Hence

$$\theta \in [\tilde{\theta}(0), \tilde{\theta}(R^2)].$$

$\square$

References

DeGroot, M. and M. Schervish (2012): *Probability and Statistics*, Boston: Pearson, 4th edition.