TECHNICAL REPORT
R-493
July 2022

Original Article

Sociological Methods & Research
1–34

DOI: 10.1177/00491241221099552
journals.sagepub.com/home/smr
SAGE

A Crash Course in Good and Bad Controls

Carlos Cinelli[1], **Andrew Forney**[2], **and Judea Pearl**[3]

Abstract

Many students of statistics and econometrics express frustration with the way a problem known as "bad control" is treated in the traditional literature. The issue arises when the addition of a variable to a regression equation produces an unintended discrepancy between the regression coefficient and the effect that the coefficient is intended to represent. Avoiding such discrepancies presents a challenge to all analysts in the data intensive sciences. This note describes graphical tools for understanding, visualizing, and resolving the problem through a series of illustrative examples. By making this "crash course" accessible to instructors and practitioners, we hope to avail these tools to a broader community of scientists concerned with the causal interpretation of regression models.

Keywords

causal inference, bad controls, back-door criterion, DAG, regression

[1] Department of Statistics, University of Washington, Seattle
[2] Department of Computer Science, Loyola Marymount University, Los Angeles
[3] Department of Computer Science, University of California, Los Angeles

Corresponding Author:
Carlos Cinelli, Department of Statistics, University of Washington, Seattle, WA 98195, USA.
Email: cinelli@uw.edu

Introduction

Students, data analysts, and empirical social scientists have likely encountered the problem of "bad controls" (Angrist and Pischke, 2009, 2014). The problem arises when an analyst needs to decide whether or not the addition of a variable to a regression equation helps getting estimates closer to the parameter of interest. Analysts have long known that some variables, when added to the regression equation, can produce unintended discrepancies between the regression coefficient and the effect that the coefficient is expected to represent. Such variables have become known as "bad controls," to be distinguished from "good controls" (also known as "confounders" or "deconfounders") which are variables that must be added to the regression equation to eliminate what came to be known as "omitted variable bias"(Angrist and Pischke, 2009; Steiner and Kim, 2016; Cinelli and Hazlett, 2020a, 2020b).

The problem of "bad controls" however, has not received *systematic* attention in the standard statistics and econometrics literature. While most of the widely adopted textbooks discuss the problem of omitting "relevant" variables, they do not provide guidance on deciding which variables are relevant, nor which variables, if included in the regression, could induce, or worsen existing biases.[1] Researchers exposed only to this literature may get the impression that adding "more controls" to a regression model is always better. The few exceptions that do discuss the problem of "bad controls" unfortunately cover only a narrow aspect of the problem (e.g. Angrist and Pischke, 2009, 2014; Wooldridge, 2010; Imbens and Rubin, 2015; Gelman, Hill and Vehtari, 2020). Typical is the discussion found in Angrist and Pischke (2009, p.64)

> Some variables are bad controls and should not be included in a regression model, even when their inclusion might be expected to change the short regression coefficients. Bad controls are variables that are themselves outcome variables in the notional experiment at hand. That is, bad controls might just as well be dependent variables too. Good controls are variables that we can think of having been fixed at the time the regressor of interest was determined.

Here, "good controls" are defined as variables that are thought to be unaffected by the treatment, whereas "bad controls" are variables that could be in principle affected by the treatment. Similar discussion can be found in Rosenbaum (2002) and Rubin (2009), for qualifying a variable for inclusion in propensity score analysis, as well as in Wooldridge (2005). Some authors (e.g, Wooldridge, 2010; Gelman, Hill and Vehtari, 2020) briefly warn about the potential of bias amplification of certain pre-treatment variables, but do not elaborate further.

Although an improvement over an absence of discussion, these conditions are neither necessary nor sufficient for deciding whether a variable is a good control.

Recent advances in graphical models have produced simple criteria to distinguish "good" from "bad" controls; these range from necessary and sufficient conditions for deciding which set of variables should be adjusted for to identify the causal effect of interest (e.g, the back-door criterion and adjustment criterion in Pearl (1995), Shpitser, VanderWeele and Robins (2012) and Perkovic et al. (2018)), to deciding which, among a set of valid adjustment sets, would yield more precise estimates (Kuroki and Miyakawa, 2003; Kuroki and Cai, 2004; Hahn, 2004; White and Lu, 2011; Henckel, Perković and Maathuis, 2022; Rotnitzky and Smucler, 2020; Witte et al., 2020). The purpose of this note is to provide practicing analysts a concise, simple, and *visual* summary of these criteria through illustrative examples.

Preliminaries and basic terminology

Causal diagrams, and more specifically directed acyclic graphs (DAGs), have become popular in the social and health sciences for explaining and resolving difficult problems of causal inference in a rigorous, yet accessible manner. Many introductions to DAGs have now been published in a number of academic fields, such as sociology (Elwert, 2013; Morgan and Winship, 2015), economics (Hünermund and Bareinboim, 2019; Cunningham, 2021), psychology (Rohrer, 2018), epidemiology (Greenland, Pearl and Robins, 1999; Hernán and Robins, 2020) and statistics (Pearl et al., 2009; Pearl, Glymour and Jewell, 2016). Here we assume that readers are familiar with the basic notions of causal inference, DAGs, and in particular "path-blocking" as well as back-door paths. For those who need to refresh these notions, we provide a gentle introduction in the appendix. Still, given the simplicity of our illustrative examples, even the uninitiated reader will be able to understand and benefit from the main lessons of this crash course.

Briefly, causal DAGs provide a parsimonious representation of the qualitative aspects of the data generating process. Letters (e.g, X) represent random variables, and arrows, such as $X \rightarrow Y$, denote a (possible) direct causal effect of X on Y. No assumptions need to be made regarding the functional form of the causal relationships, nor about the distribution of variables. For this crash course, it is important to recall the three main sources of association that form the building blocks of a DAG, and when these are closed or opened:

1. *Mediators*, or chains, are patterns of the form $X \rightarrow Z \rightarrow Y$, meaning that X *causally* affects Y through the mediator Z. Conditioning on Z in a chain *blocks* (closes) this flow of association.

2. *Common causes*, or forks, are patterns of the form $X \leftarrow Z \rightarrow Y$, meaning X and Y share a common cause (a confounder) Z, thus inducing a *non-causal* association between both variables. Conditioning on Z in a fork *blocks* this flow of association.
3. *Common effects*, or colliders, are patterns of the form $X \rightarrow Z \leftarrow Y$, meaning that both X and Y share a common effect Z. Contrary to the other two variables, by default a common effect does not induce an association between X and Y. However, conditioning on Z induces a *non-causal* association between both variables.

Moreover, one important fact to keep in mind is that controlling for a descendant of a variable is equivalent to "partially" controlling for that variable. Any arbitrary path p from X to Y (consisting of a sequence of mediators, common causes, or colliders) will be blocked conditional on Z if, and only if, Z is a common cause or mediator along the path, or if p contains a collider and Z is *not* that collider, nor any of its descendants. We say that Z d-separates X from Y if Z blocks (closes) all paths from X to Y; d-separation implies that Y and X are conditionally independent given Z.

Note that causal paths from X to Y are paths of the form $X \rightarrow \ldots \rightarrow Y$, namely, those consisting of a sequence of (possibly empty) mediators. All other paths are non-causal, and may induce "spurious" associations between X and Y. In particular, for a given variable X, we call "back-door" paths those confounding paths that begin with an arrow pointing into X. If we are interested, then, in estimating the causal effect of X on Y, our task is conceptually simple: we must block all spurious paths between X and Y, and we must not perturb any of the causal paths between them. This will be our guiding principle for deciding whether or not Z should be included in the regression equation, and it characterizes the essence of the graphical conditions known as the *back-door* criterion and the *adjustment* criterion (Pearl, 1995; Shpitser, VanderWeele and Robins, 2012; Perkovic et al., 2018). Readers can find the formal statements of these graphical criteria in the appendix.

Illustrative examples

In the following set of models, the target of our analysis is the *average causal effect* (ACE) of a treatment X on an outcome Y, which stands for the expected increase of Y in response to a unit increase in X due to an *intervention*. Observed variables will be designated by black dots and unobserved variables by white empty circles. Variable Z, highlighted in red, will represent the variable whose inclusion in the regression equation is to be decided, with "good control" standing for bias reduction, "bad control" standing for bias increase, and "neutral control" when the addition of Z neither increases nor decreases the asymptotic bias. For this last case, we will also make brief remarks about how Z could affect the precision of the ACE estimate. Readers accustomed with the potential outcomes framework should know that deciding whether Z is a "good control" is equivalent to deciding whether ignorability of treatment assignment holds, conditional on Z. Readers who prefer to see algebraic derivations can find in the appendix analytical expressions for each graph, under the assumption of linearity[2] (the problem of "bad controls" is, however, non-parametric, i.e, it holds regardless of functional form assumptions).

Models 1, 2 and 3 – Good Controls (blocking back-door paths)

In Model 1 (Figure 1a), Z stands for a common cause of both X and Y. Once we control for Z, we block the back-door path from X to Y, producing an unbiased estimate of the ACE. In Models 2 and 3, Z is not a common cause of both X and Y, and therefore, not a traditional "confounder" as in Model 1. Nevertheless, controlling for Z blocks the back-door path from X to Y due to the unobserved confounder U, and again, produces an unbiased estimate of the ACE.

Models 4, 5 and 6 – Good Controls (blocking back-door paths)

When thinking about possible threats of confounding, modelers need to keep in mind that common causes of X and any mediator (between X and Y) also confound the effect of X on Y. Therefore, Models 4, 5 and 6 (Figure 2) are analogous to Models 1, 2 and 3—controlling for Z blocks the back-door path from X to Y and produces an unbiased estimate of the ACE.

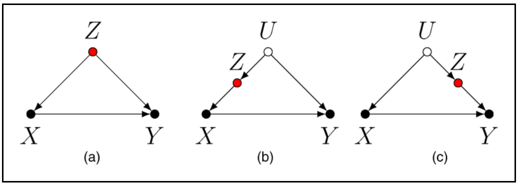


Figure 1. (a) Model 1; (b) Model 2; (c) Model 3.

Model 7 – Bad Control (M-bias)

We now encounter our first "bad control" (Figure 3). Here Z is correlated with the treatment and the outcome and it is also a "pre-treatment" variable. Traditional econometrics textbooks usually deem pre-treatment variables "good controls" (Angrist and Pischke, 2009, 2014; Imbens and Rubin, 2015). Careful analysis, however, reveals that Z is a "bad control." Controlling for Z will induce bias by opening the back-door path $X \leftarrow U1 \rightarrow Z \leftarrow U2 \rightarrow Y$, thus spoiling a previously unbiased estimate of the ACE. This structure is known as the "M-bias," and has spurred several controversies. Readers can find further discussion in Pearl (2009a, p. 186), Shrier (2009), Pearl (2009c, 2009b), Sjölander (2009), Rubin (2009), Ding and Miratrix (2015), and Pearl (2015).

Undecidable—"Damned if you do, damned if you don't." Consider a variation of Model 7 (Figure 4) such that Z has a direct effect on Y, as the one presented in Figure 4. Note that now we have an open back-door path, $X \leftarrow U_1 \rightarrow Z \rightarrow Y$, and the unadjusted estimate is no longer unbiased. While adjusting for Z closes this back-door path, it also opens back-door the path $X \leftarrow U_1 \rightarrow Z \leftarrow U_2 \rightarrow Y$, as we had in our previous example. In either case, the causal effect is not identified, and whether adjusting for Z reduces or increases the absolute value of the bias cannot be determined without further assumptions (see appendix). In this case, progress can

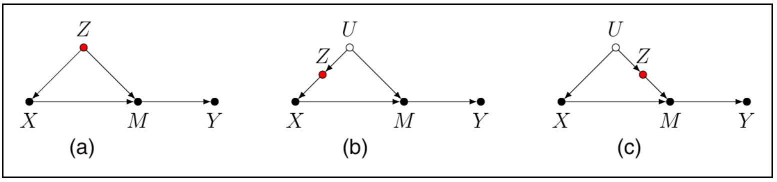


Figure 2. (a) Model 4; (b) Model 5; (c) Model 6.

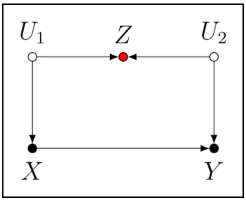


Figure 3. Model 7

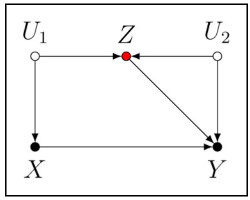


Figure 4. Variation of Model 7

be made with sensitivity analyses, by, for instance, positing plausible bounds on the the strength of the direct effect of Z on Y, or on the strength of the effects of the latent variables (Cinelli et al., 2019; Cinelli and Hazlett, 2020a).

Model 8 – Neutral Control (possibly good for precision)

In Model 8 (Figure 5), Z is not a confounder nor does it block any back-door paths. Likewise, controlling for Z does not open any back-door paths from X to Y. Thus, in terms of asymptotic bias, Z is a "neutral control." Analysis shows, however, that controlling for Z reduces the variation of the outcome variable Y, and helps to improve the precision of the ACE estimate in finite samples (Hahn, 2004; White and Lu, 2011; Henckel, Perković and Maathuis, 2022; Rotnitzky and Smucler, 2020).

Model 9 – Neutral Control (possibly bad for precision)

Similar to the previous case, in Model 9 (Figure 6) Z is "neutral" in terms of bias reduction. However, controlling for Z will reduce the variation of the treatment variable X and so may hurt the precision of the estimate of the ACE in finite samples (Henckel, Perković and Maathuis, 2022, Corollary 3.4). As a general rule of thumb, parents of X which are not necessary for identification are harmful for the asymptotic variance of the estimator; on the other hand, parents of Y which do not spoil identification are beneficial. See Henckel, Perković and Maathuis (2022) for recent developments in graphical criteria for efficient estimation via adjustment in linear models. Remarkably, these conditions also have been shown to hold in non-parametric models for a broad class of non-parametric estimators (Rotnitzky and Smucler, 2020).

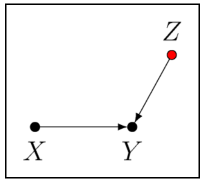


Figure 5. Model 8

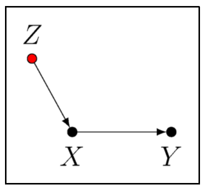


Figure 6. Model 9

Model 10 – Bad Control (bias amplification)

We now encounter our second "pre-treatment" "bad control" (Figure 7), due to a phenomenon called "bias amplification" (Bhattacharya and Vogt, 2007; Wooldridge, 2009; Pearl, 2011, 2010, 2013; Middleton et al., 2016; Steiner and Kim, 2016). Naive control for Z in this model will not only fail to deconfound the effect of X on Y, but, in linear models, will amplify any existing bias.

Models 11 and 12 – Bad Controls (overcontrol bias)

If our target quantity is the ACE, we want to leave all channels through which the causal effect flows "untouched." In Model 11 (Figure 8a), Z is a mediator of the causal effect of X on Y. Controlling for Z will block the very effect we want to estimate (the *total* effect of X on Y), thus biasing our estimates (this is usually known as "overcontrol bias"). In Model 12 (Figure 8b), although Z is not itself a mediator of the causal effect of X on Y, controlling for Z is equivalent to partially controlling for the mediator M, and will thus bias our estimates. Models 11 and 12 violate the back-door criterion (Pearl, 2009a), which excludes controls that are descendants of the treatment along paths to the outcome. Note that the same conclusions would hold if we had an extra direct causal path $X \rightarrow Y$.

Total versus direct effects. The previous considerations assume the researcher is interested in the *total* effect of X on Y, as given by the ACE. If, instead, interest lies in the *controlled direct effect* (CDE) of X on Y (i.e, the effect of X while holding Z constant by intervention, see Pearl (2009a, 2011) as well as the appendix), then adjusting for Z in Model 11 (Figure 8a) would indeed be appropriate. However, consider a variation of Model 11 (Figure 9) with an unobserved confounder of Z and Y, denoted by U, as shown in Figure 9. First notice that U *does not* confound the effect of X on Y, and thus our ACE estimate remains unbiased as it were in Model 11, so long as we do not adjust for Z. On the other hand, here adjusting for Z now opens the colliding path $X \rightarrow Z \leftarrow U \rightarrow Y$, thus biasing the CDE estimate.

Model 13 – Neutral Control (possibly good for precision)

At first look, Model 13 (Figure 10) might seem similar to Model 12, and one may think that adjusting for Z would bias the effect estimate, by restricting variations of the mediator M. However, the key difference here is that Z is a cause, not an effect, of the mediator (and, consequently, also a cause of Y). Thus, Model 13 is analogous to Model 8, and so controlling for Z will be neutral in terms of bias and may increase the precision of the ACE estimate in finite samples. Readers can find further discussion of this case in Pearl (2013).

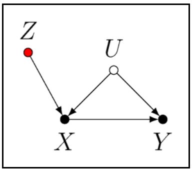


Figure 7. Model 10

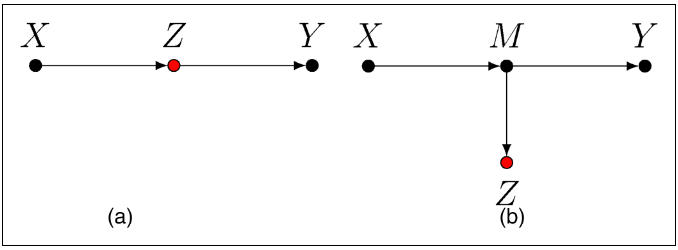


Figure 8. (a) Model 11; (b) Model 12

Models 14 and 15 – Neutral Controls (possibly helpful in the case of selection bias)

Contrary to folklore, not all "post-treatment" variables are inherently bad controls. In Models 14 and 15 (Figure 11) controlling for Z does not open any confounding paths between X and Y. Thus, Z is neutral in terms of bias. However, controlling for Z does reduce the variation of the treatment variable X and so may hurt the precision of the ACE estimate in finite samples. Additionally, in Model 15, suppose one has only samples with $W = w$ recorded (a case of selection bias[3], which we explain next). In this case, controlling for Z can help to obtain the W-specific effect of X on Y by blocking the colliding path due to W. In linear models, controlling for Z actually fully recovers the ACE (see appendix).

Models 16 and 17 – Bad Controls (selection bias)

Contrary to Models 14 and 15, here (Figure 12) controlling for Z is no longer harmless, and induces what is classically known as "selection bias" or "collider stratification bias." Adjusting for Z in Model 16 opens the colliding path $X \rightarrow Z \leftarrow U \rightarrow Y$ and so biases the ACE. In Model 17, adjusting for Z not only opens the path $X \rightarrow Z \leftarrow Y$, but also the colliding path due to the latent parents of Y, thus biasing the ACE and motivating our final example.

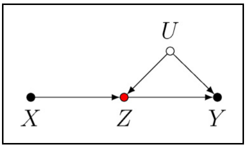


Figure 9. Variation of Model 11

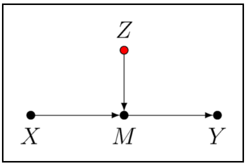


Figure 10. Model 13

Model 18 – Bad Control (case-control bias)

In our last example (Figure 13), Z is not in the causal pathway from X to Y, Z is not a direct cause of X, and Z is connected to Y. Thus, one might surmise that, as in Model 8, controlling for Z is harmless for identification, and perhaps beneficial for finite sample efficiency. However, controlling for the effects of the outcome Y will induce bias in the estimate of the ACE, even without the direct arrow $X \rightarrow Z$, thus making Z a "bad control." This happens because Z is in fact a descendant of a collider: the outcome Y itself. A visual explanation of this phenomenon using "virtual colliders" can be found in Pearl (2009a, Sec. 11.3). The same phenomenon can also be explained by explicitly drawing the potential outcomes on the DAG (see both explanations in the appendix). Model 18 is special case of selection bias usually known as "case-control" bias. Finally, although controlling for Z will generally bias numerical estimates of the ACE, it does have an exception when X has no causal effect on Y. In this scenario, X is still d-separated from Y even after conditioning on Z. Thus, adjusting for Z is valid for testing whether the effect of X on Y is *zero*.

Bad controls in applied research

Despite their simplicity, these illustrative examples should provide practitioners with a principled framework to understand many problems found in real world applications. To demonstrate, we now briefly present three cases of bad controls discussed in applied research, coming from diverse areas such as epidemiology, sociology, and economics.

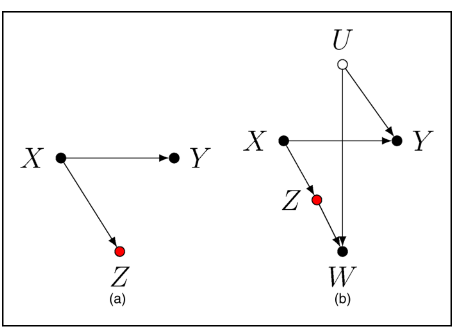


Figure 11. (a) Model 14; (b) Model 15

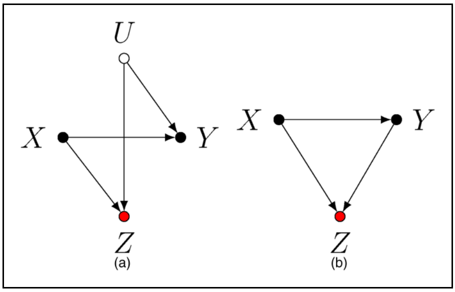


Figure 12. (a) Model 16; (b) Model 17

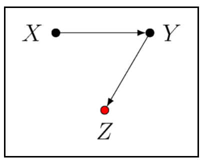


Figure 13. Model 18

The birth-weight paradox (Hernández-Díaz, Schisterman and Hernán, 2006). Infants born to smokers were found to have higher risks of mortality than infants born to non-smokers. However, among infants with low birth-weight (LBW), this relationship was reversed. This reversal of effects has created many controversies in epidemiology—does it mean that maternal smoking is beneficial for LBW infants? A plausible reason for such a finding could simply be collider stratification bias, as shown in Model 16. Here X is maternal smoking, Y infant mortality, Z birth-weight, and U stands for unobserved risk-factors (such as birth-defects and malnutrition), that could also affect birth-weight. Note that stratifying the analysis by birth-weight would induce a spurious association between smoking and mortality due to the competing risk-factors. LBW infants of non-smokers need to have alternative causes for their LBW (such as malnutrition), and such causes could also lead to higher mortality.

Homophily bias in social network analysis (Elwert and Winship, 2014). An important task in the causal inference of social networks is to estimate the causal effects of social contagion, also known as "interpersonal effects." However, social ties in the analysis of social networks may be pretreatment colliders as exemplified in the "M-bias" structure of Model

7. Suppose we are interested in assessing whether the civic engagement of individual 1 (X) leads to the civic engagement of individual 2 in the subsequent time period (Y). Let Z denote whether such individuals are friends, and U_1 and U_2 denote the personal characteristics (such as altruism) of individuals 1 and 2, respectively. Here, the social tie Z is a collider, and computing the association of Y and X between friends ($Z = 1$) would bias the interpersonal causal effects in civic engagement.

The Antebellum Puzzle (Schneider, 2020). An interesting puzzle of economic history is the fact that, during the nineteenth century in Britain and the United States, the average height of adult men fell even though the economic conditions of these countries improved alongside childhood nutrition. One possible explanation for such a paradoxical finding is selection bias in the forms of Models 17 and 18 wherein researchers using data from individuals enlisted in the military or in prison are effectively conditioning on colliders. For military records, consider Model 18, and let X denote childhood nutrition, Y adult height, and Z an indicator of whether the individual was enlisted in the military. The causal path from Y to Z represents the fact that taller men may have better opportunities in the civilian market, and thus shorter men were more likely to enlist. Restricting the analysis to those enlisted in the military is therefore equivalent to controlling for Z, and leads to selection bias. Now for prison records, consider Model 17, and let Z be an indicator of whether the individual was arrested. Here one could argue that both childhood nutrition and adult height have pathways to committing a crime through socio-economic opportunities, thus again leading to selection bias.

These examples are by no means exhaustive. Readers can find other interesting cases across applied sciences, such as: the threats of collider bias in understanding risk factors of COVID-19 (Griffith et al., 2020); the “Obesity paradox,” in which obesity appears to benefit individuals who survive heart failure (Banack and Kaufman, 2013); and examples of “bad controls” due to adjustment of mediators and colliders in multi-generational mobility (Breen, 2018), anesthesiology research (Gaskell and Sleigh, 2020) or animal science (Bello et al., 2018). Further discussion of bad controls in theoretical and applied works can be found in Pearl and Mackenzie (2018).

Multiple controls

When considering multiple controls, the status of a single control as “good” or “bad” may change depending on the context of the other variables under consideration. Nevertheless, the main lessons from our illustrative examples remain. A set of control variables $\mathbf{Z}$ will be "good" if: (i) it blocks all non-causal paths from the treatment to the outcome; (ii) it leaves any mediating paths from the treatment to the outcome "untouched" (since we are interested in the total effect); and, (iii) it does not open new spurious paths between the treatment and the outcome (e.g., due to colliders). As to efficiency considerations, we should give preference to those variables "closer" to the outcome, in opposition to those closer to the treatment—so long as, of course, this does not spoil identification.

Finally, we remind readers that, when considering models with more complicated structures, one can always resort to specialized computer programs. Open-source software implementing algorithms for selecting adjustment sets can be found in the R packages `pcalg` (Kalisch et al., 2012), `dagitty` (Textor et al., 2016)[4], and `causaleffect` (Tikka and Karvanen, 2017). Users familiar with the software `SAS` may find the procedure `CAUSALGRAPH` useful (Thompson, 2019). A web application implementing the methods discussed in Bareinboim and Pearl (2016) is also available.[5] In other words, given a causal diagram, the problem of deciding which variables are good or bad controls has been automatized.

Beyond adjustment

Here we have focused on the identification of causal effects through simple covariate adjustment, classifying Z as a "good" or "bad" control according to this criterion. However, other identification opportunities may be available. For instance, going back again to Model 10, Z is what is usually known as an "instrumental variable." In this case, while Z is indeed a "bad control," it can still be used as an instrument to bound or point identify causal effects under certain parametric assumptions, albeit using a different formula (Wright, 1928; Bowden and Turkington, 1990; Balke and Pearl, 1994; Angrist, Imbens and Rubin, 1996; Balke and Pearl, 1997; Brito and Pearl, 2002). More generally, the *do*-calculus provides a complete solution for the task of non-parametric identification of treatment effects in causal DAGs, beyond the simple adjustment formula (Pearl, 1995; Shpitser and Pearl, 2008; Pearl, 2009a). In certain instances, such as the "front-door" criterion, this allows exploiting post-treatment variables for identification (Pearl, 2009a). Further details on the *do*-calculus should be the topic of a separate crash course.

Sensitivity analysis

In real world applications, it can be the case that the causal effect of X on Y cannot be identified from the DAG structure alone. When that happens, without further assumptions, it is usually not possible to determine whether including Z in the regression equation will reduce or increase the absolute value of the bias, as we have seen in the example of Figure 4. In such cases, claims about the causal effect of X on Y must rely on knowledge beyond the constraints of the DAG, such as plausibility judgments (i) on the direct effect of observed variables, (ii) on the strength of association of latent variables with X and Y, or (iii) on the relative importance of unobserved confounders as compared to observed confounders (Cinelli, Pearl and Chen, 2018; Cinelli et al., 2019; Cinelli and Hazlett, 2020a, 2020b; Zhang et al., 2021). A suite of sensitivity analysis tools to examine the robustness of linear regression estimates to omitted variable biases (OVB) can be found in the package `sensemakr` for R, Stata and Python (Cinelli, Ferwerda and Hazlett, 2020; LaPierre et al., 2022). An interactive web application is also available.[6] Generalization of OVB results to fully nonparametric models, using Debiased Machine Learning, is developed in Chernozhukov et al. (2021).

Concluding remarks

In this note, we demonstrated through illustrative examples how simple graphical criteria can be used to decide when a variable should (or should not) be included in a regression equation—and thus whether it can be deemed a "good" or "bad" control. Many of these examples act as cautionary notes against prevailing practices: for instance, Models 7 to 10 reveal that one should be cautious of the general recommendation, usually derived from propensity score logic, of conditioning on *all* pre-treatment predictors of the treatment assignment[7]; whereas Models 14 and 15 show that not all "post-treatment" variables are "bad-controls," and some may even help with identification.

In all cases, structural knowledge is indispensable for deciding whether a variable is a good or bad control, and graphical models provide a natural language for articulating such knowledge, as well as efficient tools for examining its logical ramifications. We have found that an example-based approach to "bad controls," such as the one presented here, can serve as a powerful instructional device to supplement more extended and formal discussions of the problem. By making this "crash course" accessible to instructors and practitioners, we hope to avail these tools to a broader community of scientists concerned with the causal interpretation of regression models.

Appendix

This appendix provides a short introduction to the notions of causal models, causal diagrams and "path-blocking" for the identification of causal effects via adjustment. Readers can find more extensive discussions in Pearl (2009a); Pearl, Glymour and Jewell (2016) and Pearl and Mackenzie (2018).

Structural causal models and causal diagrams

In order to decide whether there is a discrepancy between a certain regression equation (an associational quantity), and a target "causal effect" (a causal quantity), we need to mathematically *define* what this causal effect is. And to do that, we first need the concept of a *causal model*. We briefly introduce *structural causal models* (SCM) (Pearl, 2009a) with an example.

Consider the SCM M shown in Figure 14a. The variables $\mathbf{V} = \{Z, X, Y\}$ are called the *endogenous* variables, and stand for those variables that the investigator chose to model their cause-effect relationships; the variables $\mathbf{U} = \{U_z, U_x, U_y\}$ are called the *exogenous* variables and represent everything else that the investigator chose *not* to explicitly model (these are also usually called *disturbances*). The functions $\mathcal{F} = \{f_z, f_x, f_y\}$ are called *structural equations*, and each function represents a causal process that *assigns* to its respective endogenous variable a value based on the values of the other variables. We use the assignment symbol ($\leftarrow$) to emphasize the *asymmetry* in a causal relationship, flowing from cause to effect. Finally, the exogenous variables have an associated probability distribution $P(\mathbf{U})$ summarizing their uncertainty. In this particular example, we assume the exogenous variables are mutually independent (but in general, this need not be the case). The SCM M induces a joint distribution on the endogenous variables $P(\mathbf{V})$, which we denote by *observational distribution*. In observational studies, the investigator only has access to samples of $P(\mathbf{V})$.

Every SCM has an associated graph G, usually called its *causal diagram*. In the types of models we consider here, which do not exhibit cycles, the causal diagram will be a directed acyclic graph (DAG). The causal diagram of our example is shown in Figure 14b. The graph G contains one node for each variable in M, and a directed arrow $V_i \rightarrow V_j$ whenever V_i appears in the structural equation of V_j, meaning that V_i is a *direct cause* of V_j. Here we explicitly show the exogenous variables, but, conventionally, these are omitted from the graph for brevity. When the exogenous variables are omitted from the diagram, a dashed bidirected arrow $V_i \leftrightarrow V_j$ should be added whenever the exogenous variables entering f_{v_i} and f_{v_j} are *not* independent.

Interventions and causal effects

Interventions are modeled by modifying mechanisms of the SCM. For example, the act $do(X = x)$ in the model of Figure 14a amounts to replacing the original mechanism $X \leftarrow f_x(Z, U_x)$ with a new mechanism in which X is externally forced to attain the value x, i.e., $X \leftarrow x$. This results in the modified SCM M_x of Figure 15a.

The model M_x induces an *interventional distribution* on the endogenous variables, denoted by $P(\mathbf{V} \mid do(X = x))$. With the concept of an intervention in mind, we can now define the average causal effect (i.e, the expected increase of Y in response to a unit increase in X due to an *intervention*) as the average contrast of Y under two distinct interventions:

$$\mathrm{ACE(x) = E[Y \mid do(x + 1)] - E[Y \mid do(x)]}$$

In general the ACE varies depending on levels of x, but in linear models, as we show below, the ACE reduces to a single number.

Other causal effects can be defined with the same model modification logic. For instance, the controlled direct effect (or CDE, i.e, the expected increase of Y per unit of a controlled increase in X, while holding Z constant) is defined as the difference:

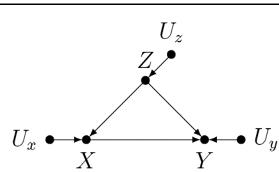
$$\mathrm{CDE(x, z) = E[Y \mid do(x + 1), do(z)] - E[Y \mid do(x), do(z)]}$$

Potential outcomes

Potential outcomes $\mathbf{V}_x$ are defined as the solution of the endogenous variables $\mathbf{V}$ in the modified model M_x. Thus, $P(\mathbf{V} \mid do(X = x))$ can be equivalently

$$M = \begin{cases} Z & \leftarrow f_z(U_z) \\ X & \leftarrow f_x(Z, U_x) \\ Y & \leftarrow f_y(X, Z, U_y) \\ \mathbf{U} & \sim P(\mathbf{U}) \end{cases}$$

(a)



(b)

Figure 14. Structural Causal Model *M* and its associated graph *G*

written as $P(\mathbf{V}_x)$ (likewise, we could have written all variables in M_x and G_x as Z_x, X_x and Y_x). As such, the ACE can be equivalently written as $ACE(x) = E[Y_{x+1}] - E[Y_x]$.

Causal and non-causal paths: chains, forks and colliders

Concretely, let us suppose that the structural equations of our example are linear, that is, $Z \leftarrow U_z$, $X \leftarrow \lambda_{zx}Z + U_x$, $Y \leftarrow \lambda_{xy}X + \lambda_{zy}Z + U_y$. Further assume that the disturbances $\mathbf{U}$ are normally distributed, and that the random variables X, Z, Y have mean zero and unit variance. Then the ACE evaluates to:

$$\mathrm{ACE(x) = E[Y \mid do(x+1)] - E[Y \mid do(x)]} = \lambda_{\mathrm{xy}}$$

To contrast, now let us compute the regression coefficient of Y on X, denoted by β_{yx}

$$\beta_{yx} = \frac{\mathrm{Cov(Y, X)}}{\mathrm{Var(X)}} = \lambda_{xy} + \lambda_{zx}\lambda_{zy}$$

Note how the regression coefficient $\beta_{yx} = \lambda_{xy} + \lambda_{zx}\lambda_{zy}$ differs from the $\mathrm{ACE} = \lambda_{\mathrm{xy}}$. This happens because the observed association of X and Y mixes both the *causal* association (the path $X \rightarrow Y$), and the *non-causal* association due to the confounder Z (the path $X \leftarrow Z \rightarrow Y$). We call such confounding paths, that start with an arrow pointing to X, "back-door paths." Note, however, that the regression coefficient of Y on X adjusting for Z (denoted by $\beta_{yx.z}$) evaluates to (a derivation is provided later, in Equations A.6 to A.10)

$$\beta_{yx.z} = \lambda_{xy}$$

That is, controlling for Z in this model effectively blocks the back-door path,

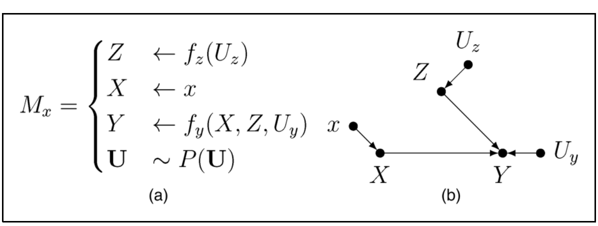


Figure 15. Effect of intervention $do(X = x)$

and recovers the ACE.

In general, how does path blocking work in a graphical model? To answer this question, we need to understand the three main patterns of a causal diagram, which help us characterize when paths (consisting of sequences of the following triplets) of the graph are blocked or open.

- **Chains (mediators).** Chains are patterns of the form $X \rightarrow Z \rightarrow Y$, meaning that X causally affects Y through the mediator Z. Conditioning on Z in a chain blocks this flow of association.
- **Forks (common causes).** Forks are patterns of the form $X \leftarrow Z \rightarrow Y$, meaning X and Y share a common cause (a confounder) Z, thus inducing a *non-causal* association between both variables. Conditioning on Z in a fork blocks this flow of association.
- **Colliders (common effects).** Colliders are patterns of the form $X \rightarrow Z \leftarrow Y$, meaning that both X and Y share a common effect Z. Contrary to the other two patterns, this path is closed by default—conditioning on Z *opens* the path and induces a *non-causal* association between X and Y.

A final rule to keep in mind is that controlling for a descendant of a variable is equivalent to "partially" controlling for that variable. Thus, controlling for a descendant of a mediator or a confounder partially blocks the flow of association, whereas controlling for a descendant of a collider partially opens the flow of association.

We can now judge whether any path p in a graph, no matter how complicated, is blocked by a set $\mathbf{Z}$. This happens if, and only if: (i) p contains a chain or a fork, such that the middle node is in $\mathbf{Z}$; *or*, (ii) p contains a collider, such that neither the middle node, nor any of its descendants, are in $\mathbf{Z}$.

The back-door and the adjustment criteria

Armed with these tools, the DAG reveals which set of variables $\mathbf{Z}$ blocks the correct paths for valid estimation of the ACE. We would like to find a set $\mathbf{Z}$, such that,

- it blocks all spurious paths from X to Y;
- it *does not* (partially) block any of the causal paths from X to Y; and,
- it does not (partially) open other spurious paths.

The above conditions characterize the so-called *back-door* criterion, later generalized by the *adjustment* criterion (Pearl, 1995; Shpitser, VanderWeele and Robins, 2012; Perkovic et al., 2018). If we can find such a set of controls $\mathbf{Z} = \{Z_1, , Z_k\}$, then the interventional expectation of Y can be computed from the observational distribution as

$$E[Y \mid do(X = x)] = E[E[Y \mid X = x, \mathbf{Z}]] \qquad (A.1)$$

Readers accustomed to potential outcomes should note that, if $\mathbf{Z}$ satisfies the adjustment criterion, then conditional ignorability holds, ie., $Y_x \perp X \mid \mathbf{Z}$.

Formal statements. For completeness, we now provide the formal statements of the back-door and adjustment criteria.

Definition 1(Back-door criterion (Pearl, 2009a)). A set of variables Z satisfies the back-door criterion relative to (X, Y) in a DAG G if:

- No node in Z is a descendant of X; and,
- Z blocks every path between X and Y that contains an arrow into X.

The adjustment criterion was later devised to explicitly handle cases in which Z may contain descendants of X.

Definition 2(Adjustment criterion (Shpitser, VanderWeele and Robins, 2012; Perkovic et al., 2018)). A set of variables Z satisfies the adjustment criterion relative to (X, Y) in a DAG G if:

- In the dag $G_{\bar{x}}$ (the DAG G with the arrows incoming into X removed), no element in Z is a descendant of any $W \in X$ which lies on a proper causal path from X to Y.
- All proper non-causal paths in G from X to Y are blocked by Z.

Linear versus non-linear models

The previous identification result is non-parametric, and it involves *two* expectations. First we compute the conditional expectation $E[Y \mid X = x, \mathbf{Z} = \mathbf{z}]$, then we *average* this conditional expectation over $P(\mathbf{Z})$. If, however, the conditional expectation function $E[Y \mid X = x, \mathbf{Z} = \mathbf{z}]$ is linear, the expression simplifies to

$$E[E[Y \mid X = x, \mathbf{Z}]] = \alpha + \beta_{yx.\mathbf{z}} x + \sum_{j=1}^{k} \beta_{yz_j.x\mathbf{z}_{-\mathbf{j}}} E[Z_j]$$

Where α is a constant and $\mathbf{Z}_{-\mathbf{j}}$ denotes the set $\mathbf{Z}$ excluding Z_j. Therefore, under the parameteric assumption of linearity, the ACE simply equals the regression coefficient $\beta_{yx.\mathbf{Z}}$, and no averaging over the distribution of $\mathbf{Z}$ is necessary (similar result can be obtained if the conditional expectation is linearly separable on X). If, however, the conditional expectation is not linear, the regression coefficient $\beta_{yx.\mathbf{Z}}$ targets a different causal quantity, which may be an incomplete summary of the ACE (see, e.g., Angrist and Pischke, 2009). In such cases, users should resort back to the proper adjustment formula as given by Equation A.1.

Virtual colliders and d-separation

Finally, we explain both d-separation and virtual colliders using the case of Model 18. Rewrite Model 18 showing the exogenous variables explicitly, as in Figure 16a. We can now clearly see the colliding path $X \rightarrow Y \leftarrow U_y$. Conditioning on Z, a descendant of Y, thus partially opens this path, and creates a spurious association between X and U_y, the disturbance of Y, making Z a "bad control." Another approach to see why Z is a bad control is to explicitly draw the potential outcome Y_x in the DAG, as shown in Figure 16b. As explained, recall that Y_x is the solution of Y in the modified model M_x. This results in $Y_x = f_Y(x, U_y)$, a function of the random variable U_y. Therefore, conditioning on Z, a descendant of Y, partially opens the path $X \rightarrow Y \leftarrow U_y \rightarrow Y_x$, and thus $Y_x \not\perp X \mid \mathbf{Z}$.

Now let us consider the case in which the arrow $X \rightarrow Y$ is removed (zero causal effect of X on Y). First recall that two nodes X and Y are *d-separated* conditional on $\mathbf{Z}$ if the set $\mathbf{Z}$ blocks every path from X to Y in the graph. If X and Y are *d-separated* conditional on $\mathbf{Z}$, this implies the conditional independence $Y \perp X \mid \mathbf{Z}$. In Model 18, when there is no path from X to Y, conditioning on Z also does not open any other paths between these two variables. Hence, X is still d-separated from Y even after conditioning on Z, and the conditional independence $Y \perp X \mid \mathbf{Z}$ holds.

Analytical expressions for linear models

Here we provide algebraic derivations for each illustrative model under the assumption of *linearity* of the structural equations. Before proceeding, we remind readers that adjusting for "bad controls" still lead to bias in non-parametric models. Furthermore, overt selection (rather than adjustment for) bad controls will also lead to bias (although the size and sign of the bias may differ).

Without loss of generality, we assume random variables have been standardized to have mean zero and unit variance. We use σ_{yx} to denote the covariance of Y and X and, like before, $\beta_{yx.z}$ to denote the partial regression coefficient of the regression of Y on X controlling for Z. The partial regression coefficient $\beta_{yx.z}$ can be written in terms of the covariances as (Cramér, 1946),

$$\beta_{yx.z} = \frac{\sigma_{yx} - \sigma_{xz}\sigma_{yz}}{1 - \sigma_{xz}^2} \tag{A.2}$$

In linear structural causal models, each edge $V_i \rightarrow V_j$ of a causal DAG can be mapped to a single structural coefficient $\lambda_{v_i v_j}$ representing the strength of the direct effect of V_i on V_j. We can use Wright's path-tracing rules (Wright, 1921) to equate the covariance $\sigma_{v_i v_j}$ of any two variables V_i and V_j, to the sum of products of structural coefficients along *unblocked* paths between V_i and V_j.

For instance, in Model 1, path-tracing results in the following covariances,

$$\sigma_{yx} = \lambda_{xy} + \lambda_{zx}\lambda_{zy} \tag{A.3}$$

$$\sigma_{xz} = \lambda_{zx} \tag{A.4}$$

$$\sigma_{yz} = \lambda_{zy} + \lambda_{zx}\lambda_{xy} \tag{A.5}$$

With the help of Wright's rules and Equation A.2, one can easily proceed with the algebraic derivation for each model. For the sake of brevity, we provide the full derivation for Model 1, and for the remaining models (except Model 15) only the final result is presented (since the derivations would be very

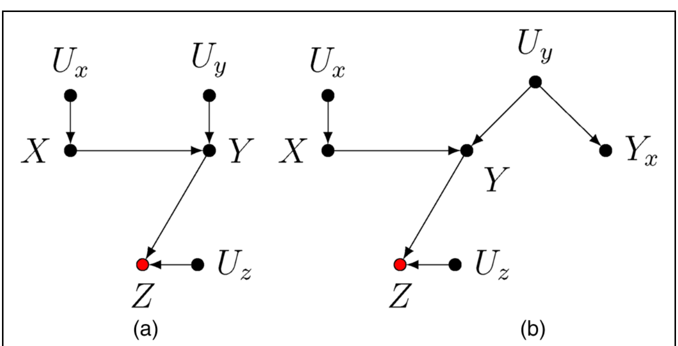


Figure 16. Model 18 explained

similar).

Model 1. In Model 1, the average causal effect of X on Y equals ACE $= \lambda_{xy}$. The unadjusted regression coefficient equals $\beta_{yx} = \sigma_{yx} = \lambda_{xy} + \lambda_{zx}\lambda_{zy}$, and the partial regression coefficient equals,

$$\beta_{yx.z} = \frac{\sigma_{yx} - \sigma_{xz}\sigma_{yz}}{1 - \sigma_{xz}^2} \tag{A.6}$$

$$= \frac{\lambda_{xy} + \lambda_{zx}\lambda_{zy} - \lambda_{zx}(\lambda_{zy} + \lambda_{zx}\lambda_{xy})}{1 - \lambda_{zx}^2} \tag{A.7}$$

$$= \frac{\lambda_{xy} + \lambda_{zx}\lambda_{zy} - \lambda_{zx}\lambda_{zy} - \lambda_{zx}^2\lambda_{xy}}{1 - \lambda_{zx}^2} \tag{A.8}$$

$$= \frac{(1 - \lambda_{zx}^2)\lambda_{xy}}{1 - \lambda_{zx}^2} \tag{A.9}$$

$$= \lambda_{xy} \tag{A.10}$$

Model 2. Path-tracing leads to the following covariances,

$$\sigma_{yx} = \lambda_{xy} + \lambda_{zx}\lambda_{uz}\lambda_{uy} \tag{A.11}$$

$$\sigma_{xz} = \lambda_{zx} \tag{A.12}$$

$$\sigma_{yz} = \lambda_{zx}\lambda_{xy} + \lambda_{uz}\lambda_{uy} \tag{A.13}$$

We have: ACE $= \lambda_{xy}$, $\beta_{yx} = \lambda_{xy} + \lambda_{zx}\lambda_{uz}\lambda_{uy}$, and $\beta_{yx.z} = \lambda_{xy}$.

Model 3. Path-tracing leads to the following covariances,

$$\sigma_{yx} = \lambda_{xy} + \lambda_{ux}\lambda_{uz}\lambda_{zy} \tag{A.14}$$

$$\sigma_{xz} = \lambda_{ux}\lambda_{uz} \tag{A.15}$$

$$\sigma_{yz} = \lambda_{zy} + \lambda_{uz}\lambda_{ux}\lambda_{uy} \tag{A.16}$$

We have: ACE $= \lambda_{xy}$, $\beta_{yx} = \lambda_{xy} + \lambda_{ux}\lambda_{uz}\lambda_{zy}$, and $\beta_{yx.z} = \lambda_{xy}$.

Model 4. Path-tracing leads to the following covariances,

$$\sigma_{yx} = \lambda_{xm}\lambda_{my} + \lambda_{zx}\lambda_{zm}\lambda_{my} \tag{A.17}$$

$$\sigma_{xz} = \lambda_{zx} \tag{A.18}$$

$$\sigma_{yz} = \lambda_{zm}\lambda_{my} + \lambda_{zx}\lambda_{xm}\lambda_{my} \tag{A.19}$$

We have: ACE $= \lambda_{xm}\lambda_{my}$, $\beta_{yx} = \lambda_{xm}\lambda_{my} + \lambda_{zx}\lambda_{zm}\lambda_{my}$, and $\beta_{yx.z} = \lambda_{xm}\lambda_{my}$.

Model 5. Path-tracing leads to the following covariances,

$$\sigma_{yx} = \lambda_{xm}\lambda_{my} + \lambda_{zx}\lambda_{uz}\lambda_{um}\lambda_{my} \tag{A.20}$$

$$\sigma_{xz} = \lambda_{zx} \tag{A.21}$$

$$\sigma_{yz} = \lambda_{uz}\lambda_{um}\lambda_{my} + \lambda_{zx}\lambda_{xm}\lambda_{my} \tag{A.22}$$

We have: ACE $= \lambda_{xm}\lambda_{my}$, $\beta_{yx} = \lambda_{xm}\lambda_{my} + \lambda_{zx}\lambda_{uz}\lambda_{um}\lambda_{my}$, and $\beta_{yx.z} = \lambda_{xm}\lambda_{my}$.

Model 6. Path-tracing leads to the following covariances,

$$\sigma_{yx} = \lambda_{xm}\lambda_{my} + \lambda_{ux}\lambda_{uz}\lambda_{zm}\lambda_{my} \tag{A.23}$$

$$\sigma_{xz} = \lambda_{ux}\lambda_{uz} \tag{A.24}$$

$$\sigma_{yz} = \lambda_{zm}\lambda_{my} + \lambda_{uz}\lambda_{ux}\lambda_{xm}\lambda_{my} \tag{A.25}$$

We have: ACE $= \lambda_{xm}\lambda_{my}$, $\beta_{yx} = \lambda_{xm}\lambda_{my} + \lambda_{ux}\lambda_{uz}\lambda_{zm}\lambda_{my}$, and $\beta_{yx.z} = \lambda_{xm}\lambda_{my}$.

Model 7. Path-tracing leads to the following covariances,

$$\sigma_{yx} = \lambda_{xy} \tag{A.26}$$

$$\sigma_{xz} = \lambda_{u_1x}\lambda_{u_1z} \tag{A.27}$$

$$\sigma_{yz} = \lambda_{u_2z}\lambda_{u_2y} + \lambda_{u_1z}\lambda_{u_1x}\lambda_{xy} \tag{A.28}$$

We have: ACE $= \lambda_{xy}$, $\beta_{yx} = \lambda_{xy}$, and $\beta_{yx.z} = \lambda_{xy} - \frac{\lambda_{u_1x}\lambda_{u_1z}\lambda_{u_2z}\lambda_{u_2y}}{1-(\lambda_{u_1x}\lambda_{u_1z})^2}$.

Variation of Model 7. Path-tracing leads to the following covariances,

$$\sigma_{yx} = \lambda_{xy} + \lambda_{u_1x}\lambda_{u_1z}\lambda_{zy} \tag{A.29}$$

$$\sigma_{xz} = \lambda_{u_1x}\lambda_{u_1z} \tag{A.30}$$

$$\sigma_{yz} = \lambda_{zy} + \lambda_{u_2z}\lambda_{u_2y} + \lambda_{u_1z}\lambda_{u_1x}\lambda_{xy} \tag{A.31}$$

We have: ACE $= \lambda_{xy}$, $\beta_{yx} = \lambda_{xy} + \lambda_{u_1x}\lambda_{u_1z}\lambda_{zy}$, and $\beta_{yx.z} = \lambda_{xy} - \frac{\lambda_{u_1x}\lambda_{u_1z}\lambda_{u_2z}\lambda_{u_2y}}{1-(\lambda_{u_1x}\lambda_{u_1z})^2}$.
Thus, depending on the parameterization of the model, the absolute value of the bias of $\beta_{yx.z}$ can be greater than that of β_{yx}.

Model 8. Path-tracing leads to the following covariances,

$$\sigma_{yx} = \lambda_{xy} \tag{A.32}$$

$$\sigma_{xz} = 0 \tag{A.33}$$

$$\sigma_{yz} = \lambda_{zy} \tag{A.34}$$

We have: $\text{ACE} = \lambda_{xy}$, $\beta_{yx} = \lambda_{xy}$, and $\beta_{yx.z} = \lambda_{xy}$.

Model 9. Path-tracing leads to the following covariances,

$$\sigma_{yx} = \lambda_{xy} \tag{A.35}$$

$$\sigma_{xz} = \lambda_{zx} \tag{A.36}$$

$$\sigma_{yz} = \lambda_{zx}\lambda_{xy} \tag{A.37}$$

We have: $\text{ACE} = \lambda_{xy}$, $\beta_{yx} = \lambda_{xy}$, and $\beta_{yx.z} = \lambda_{xy}$.

Model 10. Path-tracing leads to the following covariances,

$$\sigma_{yx} = \lambda_{xy} + \lambda_{ux}\lambda_{uy} \tag{A.38}$$

$$\sigma_{xz} = \lambda_{zx} \tag{A.39}$$

$$\sigma_{yz} = \lambda_{zx}\lambda_{xy} \tag{A.40}$$

We have: $\text{ACE} = \lambda_{xy}$, $\beta_{yx} = \lambda_{xy} + \lambda_{ux}\lambda_{uy}$, and $\beta_{yx.z} = \lambda_{xy} + \frac{\lambda_{ux}\lambda_{uy}}{1-\lambda_{zx}^2}$. Since $0 < 1 - \lambda_{zx}^2 < 1$, the absolute value of the bias of $\beta_{yx.z}$ is always greater than that of β_{yx}.

Model 11. Path-tracing leads to the following covariances,

$$\sigma_{yx} = \lambda_{xz}\lambda_{zy} \tag{A.41}$$

$$\sigma_{xz} = \lambda_{xz} \tag{A.42}$$

$$\sigma_{yz} = \lambda_{zy} \tag{A.43}$$

We have: $\text{ACE} = \lambda_{xz}\lambda_{zy}$, $\text{CDE} = 0$, $\beta_{yx} = \lambda_{xz}\lambda_{zy}$, $\beta_{yx.z} = 0$.

Model 12. Path-tracing leads to the following covariances,

$$\sigma_{yx} = \lambda_{xm}\lambda_{my} \tag{A.44}$$

$$\sigma_{xz} = \lambda_{xm}\lambda_{mz} \tag{A.45}$$

$$\sigma_{yz} = \lambda_{mz}\lambda_{my} \tag{A.46}$$

We have: ACE $= \lambda_{xm}\lambda_{my}$, CDE $= 0$, $\beta_{yx} = \lambda_{xm}\lambda_{my}$, and $\beta_{yx.z} = \lambda_{xm}\lambda_{my} \times (\frac{1-\lambda_{mz}^2}{1-\lambda_{xm}^2\lambda_{mz}^2})$.

Variation of Model 11. Path-tracing leads to the following covariances,

$$\sigma_{yx} = \lambda_{xz}\lambda_{zy} \tag{A.47}$$

$$\sigma_{xz} = \lambda_{xz} \tag{A.48}$$

$$\sigma_{yz} = \lambda_{zy} + \lambda_{uz}\lambda_{uy} \tag{A.49}$$

We have: ACE $= \lambda_{xz}\lambda_{zy}$, CDE $= 0$, $\beta_{yx} = \lambda_{xz}\lambda_{zy}$, and $\beta_{yx.z} = -\frac{\lambda_{xz}\lambda_{uz}\lambda_{uy}}{1-\lambda_{xz}^2}$.

Model 13. Path-tracing leads to the following covariances,

$$\sigma_{yx} = \lambda_{xm}\lambda_{my} \tag{A.50}$$

$$\sigma_{xz} = 0 \tag{A.51}$$

$$\sigma_{yz} = \lambda_{zm}\lambda_{my} \tag{A.52}$$

We have: ACE $= \lambda_{xm}\lambda_{my}$, $\beta_{yx} = \lambda_{xm}\lambda_{my}$, and $\beta_{yx.z} = \lambda_{xm}\lambda_{my}$.

Model 14. Path-tracing leads to the following covariances,

$$\sigma_{yx} = \lambda_{xy} \tag{A.53}$$

$$\sigma_{xz} = \lambda_{xz} \tag{A.54}$$

$$\sigma_{yz} = \lambda_{xz}\lambda_{xy} \tag{A.55}$$

We have: ACE $= \lambda_{xy}$, $\beta_{yx} = \lambda_{xy}$, and $\beta_{yx.z} = \lambda_{xy}$.

Model 15. Model 15 requires a more elaborate derivation since we need to consider four variables. Here we only have samples with $W = w$ recorded. In linear models, this is equivalent to having adjusted for W in a regression model. Thus, this means the researcher does not have access to the regression coefficients β_{yx} nor $\beta_{yx.z}$, but rather $\beta_{yx.w}$ and $\beta_{yx.wz}$ (that is, W is *always* conditioned on, due to sample selection). Path-tracing leads to the following covariances (there is no need to compute all of them to solve this problem, but we show them here for completeness),

$$\sigma_{yx} = \lambda_{xy} \tag{A.56}$$

$$\sigma_{xz} = \lambda_{xz} \tag{A.57}$$

$$\sigma_{xw} = \lambda_{xz}\lambda_{zw} \tag{A.58}$$

$$\sigma_{yz} = \lambda_{xz}\lambda_{xy} \tag{A.59}$$

$$\sigma_{yw} = \lambda_{uy}\lambda_{uw} + \lambda_{xy}\lambda_{xz}\lambda_{zw} \tag{A.60}$$

$$\sigma_{zw} = \lambda_{zw} \tag{A.61}$$

The average causal effect is $\mathrm{ACE} = \lambda_{\mathrm{xy}}$. By Equation A.2, the regression coefficient without adjusting for Z, $\beta_{yx.w}$, equals

$$\beta_{yx.w} = \lambda_{xy} - \frac{\lambda_{xz}\lambda_{zw}\lambda_{uw}\lambda_{uy}}{1-(\lambda_{xz}\lambda_{zw})^2} \tag{A.62}$$

Now we must compute the regression coefficient adjusting for Z, $\beta_{yx.wz}$. Following Cramér (1946) $\beta_{yx.wz}$ can be written as

$$\beta_{yx.wz} = \frac{\rho_{yx.z} - \rho_{xw.z}\rho_{yw.z}}{(1-\rho^2_{xw.z})^{1/2}} \times \frac{\sigma_{y.z}}{\sigma_{x.z}} \tag{A.63}$$

Where

$$\rho_{yx.z} = \frac{\sigma_{yx} - \sigma_{xz}\sigma_{yz}}{(1-\sigma^2_{xz})^{1/2}(1-\sigma^2_{yz})^{1/2}} \tag{A.64}$$

denotes the partial correlation of Y with X after adjusting for Z and

$$\sigma_{y.z} = (1-\sigma^2_{yz})^{1/2}$$

denotes the partial standard deviation of Y after adjusting for Z. Since X is d-separated from W given Z, we know that $\rho_{xw.z} = 0$ (we can also verify this by checking that $\beta_{xw.z} = 0$). Also note that $\sigma_{yz} = \lambda_{xz}\lambda_{xy} = \sigma_{xz}\sigma_{yx}$. We thus obtain

$$\beta_{yx.wz} = \rho_{yx.z} \times \frac{(1-\sigma^2_{yz})^{1/2}}{(1-\sigma^2_{xz})^{1/2}} \tag{A.65}$$

$$= \frac{\sigma_{yx} - \sigma_{xz}\sigma_{yz}}{(1-\sigma^2_{xz})^{1/2}(1-\sigma^2_{yz})^{1/2}} \times \frac{1-\sigma^2_{yz}}{1-\sigma^2_{xz}} \tag{A.66}$$

$$= \frac{\sigma_{yx} - \sigma_{xz}^2 \sigma_{yx}}{1 - \sigma_{xz}^2} \tag{A.67}$$

$$= \sigma_{yx} \frac{1 - \sigma_{xz}^2}{1 - \sigma_{xz}^2} = \sigma_{yx} = \lambda_{xy} \tag{A.68}$$

Model 16. Path-tracing leads to the following covariances,

$$\sigma_{yx} = \lambda_{xy} \tag{A.69}$$

$$\sigma_{xz} = \lambda_{xz} \tag{A.70}$$

$$\sigma_{yz} = \lambda_{xz}\lambda_{xy} + \lambda_{uz}\lambda_{uy} \tag{A.71}$$

We have: ACE $= \lambda_{xy}$, $\beta_{yx} = \lambda_{xy}$, and $\beta_{yx.z} = \lambda_{xy} - \frac{\lambda_{xz}\lambda_{uz}\lambda_{uy}}{1-\lambda_{xz}^2}$.

Model 17. Path-tracing leads to the following covariances,

$$\sigma_{yx} = \lambda_{xy} \tag{A.72}$$

$$\sigma_{xz} = \lambda_{xz} + \lambda_{xy}\lambda_{yz} \tag{A.73}$$

$$\sigma_{yz} = \lambda_{yz} + \lambda_{xz}\lambda_{xy} \tag{A.74}$$

We have: ACE $= \lambda_{xy}$, $\beta_{yx} = \lambda_{xy}$, and $\beta_{yx.z} = \lambda_{xy} \times \frac{(1-\lambda_{xz}\sigma_{xz})}{1-\sigma_{xz}^2} - \frac{\sigma_{xz}\lambda_{yz}}{1-\sigma_{xz}^2}$.

Model 18. Path-tracing leads to the following covariances,

$$\sigma_{yx} = \lambda_{xy} \tag{A.75}$$

$$\sigma_{xz} = \lambda_{xy}\lambda_{yz} \tag{A.76}$$

$$\sigma_{yz} = \lambda_{yz} \tag{A.77}$$

We have: ACE $= \lambda_{xy}$, $\beta_{yx} = \lambda_{xy}$, and $\beta_{yx.z} = \lambda_{xy} \times \frac{1-\lambda_{yz}^2}{1-\lambda_{xy}^2\lambda_{yz}^2}$.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research was supported in parts by grants from the National Science Foundation [#IIS-2106908], Office of Naval Research [#N00014-17-S-12091 and #N00014-21-1-2351], and Toyota Research Institute of North America [#PO000897].

ORCID iDs

Carlos Cinelli https://orcid.org/0000-0002-2021-7739
Andrew Forney https://orcid.org/0000-0002-6366-1290

Notes

1. See Chen and Pearl (2013) for a critical appraisal of econometrics textbooks, and Bollen and Pearl (2013) for eight misconceptions that still prevail in statistics and the social sciences. Warnings regarding the adjustment of post-treatment variables date back to at least Rosenbaum (1984), but a systematic solution to the problem of covariate selection was not available before the development of causal graphical models.
2. R code with numerical simulations for all examples can be found in: https://www.kaggle.com/code/carloscinelli/crash-course-in-good-and-bad-controls-linear-r.
3. Some economists may denote confounding bias as "selection bias," meaning preferential selection to treatment (Angrist and Pischke, 2009, 2014). Here selection bias means preferential selection into the available data.
4. Also available online in www.dagitty.net.
5. Available online at www.causalfusion.net.
6. Available online at https://carloscinelli.shinyapps.io/robustness_value/.
7. For instance, examples of such recommendations can be found in Rosenbaum (2002, p.76), Rubin (2009), Imbens and Rubin (2015, p.265), Dorie et al. (2016, p.3453).

References

Angrist, J. and J.-S. Pischke 2009. *Mostly Harmless Econometrics: An Empiricists Guide*. Princeton: Princeton University Press.

Angrist, J. D., G. W. Imbens, and D. B Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American statistical Association* 91 (434): 444-55.

Angrist, J. D. and J. -S. Pischke 2014. *Mastering 'metrics: The Path From Cause to Effect*Princeton University Press.

Balke, A. and J Pearl. 1994. "Counterfactual probabilities: Computational methods, bounds and applications". In *Uncertainty Proceedings 1994*, pages 46–54. Elsevier.

Balke, A. and J Pearl. 1997. "Bounds on Treatment Effects from Studies with Imperfect Compliance." *Journal of the American Statistical Association* 92 (439): 1171-6.

Banack, H. R. and J. S Kaufman. 2013. "The "obesity Paradox" Explained." *Epidemiology (Cambridge, Mass.)* 24 (3): 461-2.

Bareinboim, E. and J Pearl. 2016. "Causal Inference and the Data-fusion Problem." *Proceedings of the National Academy of Sciences* 113 (27): 7345-52.

Bello, N. M., V. C. Ferreira, D. Gianola, and G. J Rosa. 2018. "Conceptual Framework for Investigating Causal Effects from Observational Data in Livestock." *Journal of animal science* 96 (10): 4045-62.

Bhattacharya, J. and W. B Vogt. 2007. Do instrumental variables belong in propensity scores? Technical report, National Bureau of Economic Research.

Bollen, K. A. and J Pearl. 2013. "Eight myths about causality and structural equation models". In *Handbook of causal analysis for social research*, pages 301–328. Springer.

Bowden, R. J. and D. A Turkington. 1990. *Instrumental Variables*Cambridge university press.

Breen, R. 2018. "Some Methodological Problems in the Study of Multigenerational Mobility." *European Sociological Review* 34 (6): 603-11.

Brito, C. and J Pearl. 2002. "Generalized instrumental variables". In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, pages 85–93. Morgan Kaufmann Publishers Inc.

Chen, B. and J Pearl. 2013. "Regression and Causation: A Critical Examination of Six Econometrics Textbooks." *Real-World Economics Review*

Chernozhukov, V., C. Cinelli, W. Newey, A. Sharma, and V Syrgkanis. 2021. "Omitted variable bias in machine learned causal models." arXiv preprint arXiv:2112.13398.

Cinelli, C., J. Ferwerda, and C Hazlett. 2020. "sensemakr: Sensitivity analysis tools for OLS in R and Stata." Available at SSRN 3588978.

Cinelli, C. and C Hazlett. 2020a. "Making Sense of Sensitivity: Extending Omitted Variable Bias." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82 (1): 39-67.

Cinelli, C. and C Hazlett. 2020b. "An Omitted Variable Bias Framework for Sensitivity Analysis of Instrumental Variables." *Work. Pap*

Cinelli, C., D. Kumor, B. Chen, J. Pearl, and E Bareinboim. 2019. "Sensitivity Analysis of Linear Structural Causal Models." *International Conference on Machine Learning*

Cinelli, C., J. Pearl, and B Chen. 2018. When confounders are confounded: Naive benchmarking in sensitivity analysis.

Cramér, H. 1946. *Mathematical Methods of Statistics*Princeton University Press.

Cunningham, S. 2021. *Causal Inference: The Mixtape*Yale University Press.

Ding, P. and L. W Miratrix. 2015. "To Adjust Or Not to Adjust? Sensitivity Analysis of M-bias and Butterfly-bias." *Journal of Causal Inference* 3 (1): 41-57.

Dorie, V., M. Harada, N. B. Carnegie, and J Hill. 2016. "A Flexible, Interpretable Framework for Assessing Sensitivity to Unmeasured Confounding." *Statistics in medicine* 35 (20): 3453-70.

Elwert, F. 2013. Graphical causal models. In *Handbook of causal analysis for social research*, pages 245–273. Springer.

Elwert, F. and C Winship. 2014. "Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable." *Annual review of sociology* 40:31-53.

Gaskell, A. L. and J. W Sleigh. 2020. "An Introduction to Causal Diagrams for Anesthesiology Research." *Anesthesiology* 132 (5): 951-67.

Gelman, A., J. Hill, and A Vehtari. 2020. *Regression and Other Stories*Cambridge University Press.

Greenland, S., J. Pearl, and J. M Robins. 1999. "Causal Diagrams for Epidemiologic Research." *Epidemiology (Cambridge, Mass.)*37-48.

Griffith, G. J., T. T. Morris, M. J. Tudball, A. Herbert, G. Mancano, L. Pike, G. C. Sharp, J. Sterne, T. M. Palmer, G. D. Smith, and al et. 2020. "Collider Bias Undermines Our Understanding of Covid-19 Disease Risk and Severity." *Nature communications* 11 (1): 1-12.

Hahn, J. 2004. "Functional Restriction and Efficiency in Causal Inference." *Review of Economics and Statistics* 86 (1): 73-6.

Henckel, L., E. Perković, and M. H Maathuis. 2022. "Graphical Criteria for Efficient Total Effect Estimation Via Adjustment in Causal Linear Models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 84 (2): 579-99.

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

Hernández-Díaz, S., E. F. Schisterman, and M. A Hernán. 2006. "The Birth Weight "paradox" Uncovered?" *American journal of epidemiology* 164 (11): 1115-20.

Hünermund, P. and E Bareinboim. 2019. "Causal inference and data-fusion in econometrics." *arXiv preprint arXiv:1912.09104*.

Imbens, G. W. and D. B Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*Cambridge University Press.

Kalisch, M., M. Mächler, D. Colombo, M. H. Maathuis, and P Bühlmann. 2012. "Causal Inference Using Graphical Models with the R Package Pcalg." *Journal of Statistical Software* 47 (11): 1-26.

Kuroki, M. and Z Cai. 2004. "Selection of identifiability criteria for total effects by using path diagrams." In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pages 333–340.

Kuroki, M. and M Miyakawa. 2003. “Covariate Selection for Estimating the Causal Effect of Control Plans by Using Causal Diagrams.” *Journal of the Royal Statistical Society: Series B* 65 (1): 209-22.

LaPierre, N., B. L. Hill, Z. Zhang, and C Cinelli. 2022. sensemakr: Sensitivity analysis tools for OLS in python. https://github.com/KennyZhang-17/PySensemakr.

Middleton, J. A., M. A. Scott, R. Diakow, and J. L Hill. 2016. “Bias Amplification and Bias Unmasking.” *Political Analysis* 24 (3): 307-23.

Morgan, S. L. and C Winship. 2015. *Counterfactuals and Causal Inference* Cambridge University Press.

Pearl, J. 1995. “Causal Diagrams for Empirical Research.” *Biometrika* 82 (4): 669-88.

Pearl, J. 2009a. *Causality*Cambridge University Press.

Pearl, J. 2009b. Letter to the editor: Remarks on the method of propensity score. *Statistics in Medicine*, 28:1420–1423. URL: https://ucla.in/2NbS14j.

Pearl, J. 2009c. Myth, confusion, and science in causal analysis. *UCLA Cognitive Systems Laboratory*, Technical Report (R-348). URL: https://ucla.in/2EihVyD.

Pearl, J. 2010. On a class of bias-amplifying variables that endanger effect estimates. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, pages 417–424. URL: https://ucla.in/2N8mBMg.

Pearl, J. 2011. “Invited Commentary: Understanding Bias Amplification.” *American journal of epidemiology* 174 (11): 1223-7. doi: URL: https://ucla.in/2PORDX2.

Pearl, J. 2013. “Linear Models: A Useful “microscope” for Causal Analysis.” *Journal of Causal Inference* 1 (1): 155-70. doi: URL: https://ucla.in/2LcpmHz.

Pearl, J. 2015. “Comment on Ding and Miratrix:“to Adjust Or Not to Adjust?”” *Journal of Causal Inference* 3 (1): 59-60. doi: URL: https://ucla.in/2PgOWNd.

Pearl, J. et al. 2009. “Causal Inference in Statistics: An Overview.” *Statistics surveys* 3:96-146.

Pearl, J., M. Glymour, and N. P Jewell. 2016. *Causal Inference in Statistics: a Primer*John Wiley & Sons.

Pearl, J. and D Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect* Hachette UK.

Perkovic, E., J. Textor, M. Kalisch, and M. H Maathuis. 2018. “Complete Graphical Characterization and Construction of Adjustment Sets in Markov Equivalence Classes of Ancestral Graphs.” *Journal of Machine Learning Research* 18

Rohrer, J. M. 2018. “Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data.” *Advances in Methods and Practices in Psychological Science* 1 (1): 27-42.

Rosenbaum, P. R. 1984. “The Consequences of Adjustment for a Concomitant Variable that Has Been Affected by the Treatment.” *Journal of the Royal Statistical Society: Series A (General)* 147 (5): 656-66.

Rosenbaum, P. R. 2002. *Observational Studies*Springer.
Rotnitzky, A. and E Smucler. 2020. “Efficient Adjustment Sets for Population Average Causal Treatment Effect Estimation in Graphical Models.” *Journal of Machine Learning Research* 21 (188): 1-86.
Rubin, D. B. 2009. “Should Observational Studies Be Designed to Allow Lack of Balance in Covariate Distributions Across Treatment Groups?” *Statistics in Medicine* 28 (9): 1420-3.
Schneider, E. B. 2020. “Collider Bias in Economic History Research.” *Explorations in Economic History* 78:101356.
Shpitser, I. and J Pearl. 2008. “Complete Identification Methods for the Causal Hierarchy.” *Journal of Machine Learning Research* 9 (Sep): 1941-79.
Shpitser, I., T. VanderWeele, and J. M Robins. 2012. On the validity of covariate adjustment for estimating causal effects. *arXiv preprint arXiv:1203.3515*.
Shrier, I. 2009. “Propensity Scores.” *Statistics in Medicine* 28 (8): 1317-8.
Sjölander, A. 2009. “Propensity Scores and M-structures.” *Statistics in medicine* 28 (9): 1416-20.
Steiner, P. M. and Y Kim. 2016. “The Mechanics of Omitted Variable Bias: Bias Amplification and Cancellation of Offsetting Biases.” *Journal of causal inference* 4 (2).
Textor, J., B. van der Zander, M. S. Gilthorpe, M. Liśkiewicz, and G. T Ellison. 2016. “Robust Causal Inference Using Directed Acyclic Graphs: the R Package ‘dagitty’.” *International journal of epidemiology* 45 (6): 1887-94.
Thompson, C. 2019. Causal graph analysis with the causalgraph procedure. In *Proceedings of SAS Global Forum*.
Tikka, S. and J Karvanen. 2017. “Identifying Causal Effects with the R Package Causaleffect.” *Journal of Statistical Software* 76
White, H. and X Lu. 2011. “Causal Diagrams for Treatment Effect Estimation with Application to Efficient Covariate Selection.” *Review of Economics and Statistics* 93 (4): 1453-9.
Witte, J., L. Henckel, M. H. Maathuis, and V Didelez. 2020. “On Efficient Adjustment in Causal Graphs.” *Journal of Machine Learning Research* 21:246.
Wooldridge, J. 2009. Should instrumental variables be used as matching variables. Technical report, Citeseer.
Wooldridge, J. M. 2005. “Violating Ignorability of Treatment by Controlling for Too Many Factors.” *Econometric Theory* 21 (5): 1026-8.
Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*MIT press.
Wright, P. G. 1928. *Tariff on Animal and Vegetable Oils*. New York: Macmillan Company.
Wright, S. 1921. “Correlation and Causation.” *Journal of agricultural research* 20 (7): 557-85.

Zhang, C., C. Cinelli, B. Chen, and J Pearl. 2021. Exploiting equality constraints in causal inference. In *International Conference on Artificial Intelligence and Statistics*, pages 1630–1638. PMLR.

Author Biographies

Carlos Cinelli is an assistant professor of statistics at the University of Washington. His research focuses on developing new causal and statistical methods for transparent and robust causal claims in the empirical sciences.

Andrew Forney is currently an assistant professor of computer science at Loyola Marymount University. His research primarily focuses on the intertwining of topics in reinforcement learning and causal inference towards crafting increasingly dynamic and cognitive autonomous agents. With backgrounds in both computing and psychology, he hopes to concert the tools of each discipline to better understand human thought and to likewise better approximate it.

Judea Pearl is Chancellor's professor of computer science and statistics at the UCLA. He is a graduate of the Technion, Israel, and has joined the faculty of the UCLA in 1970, where he currently directs the Cognitive Systems Laboratory and conducts research in artificial intelligence, human cognition and philosophy of science. Pearl has authored Heuristics (1983), Probabilistic Reasoning (1988) and Causality (2000, 2009), winner of the London School of Economics Lakatosh Award. In 2018, he authored The Book of Why (with D. Mackenzie), which brings causal reasoning to the general public. Pearl is a member of the National Academy of Sciences and a fellow of the cognitive science society and the Association for the Advancement of Artificial Intelligence. In 2012, he won the Technion's Harvey Prize and the ACM Alan Turing Award.