

Nonparametric Identification of Causal Effects under Temporal Dependence

Allan Dafoe¹

Abstract

Social scientists routinely address temporal dependence by adopting a simple technical fix. However, the correct identification strategy for a causal effect depends on causal assumptions. These need to be explicated and justified; almost no studies do so. This article addresses this shortcoming by offering a precise general statement of the (nonparametric) causal assumptions required to identify causal effects under temporal dependence. In particular, this article clarifies when one should condition or not condition on lagged dependent variables (LDVs) to identify causal effects: one should not condition on LDVs, if there is no reverse causation *and* no outcome auto-causation; one should condition on LDVs if there are no unobserved common causes of treatment and the lagged outcome, *or* no unobserved persistent causes of the outcome. When only one of these is true (with one exception), the incorrect decision will induce bias. Absent a well-justified identification strategy, inferences should be appropriately qualified.

Keywords

temporal dependence, time series, lagged dependent variables, causal graphs, democratic peace

¹ Department of Political Science, Yale University, New Haven, CT, USA

Corresponding Author:

Allan Dafoe, Department of Political Science, Yale University, New Haven, CT 06520, USA.
Email: allan.dafoe@yale.edu

Introduction

Temporal dependence poses a problem for causal inference. Temporal dependence is the property that, after conditioning on covariates, observations are not independent across time; for parametric estimation, this means that the error term is not independent across time.¹ Concerns about temporal dependence are prevalent in studies of time series (cross-sectional) data; some of the most highly cited works in political methodology concern this issue (Beck and Katz 1995; Beck, Katz, and Tucker 1998; Box-Steffensmeier and Jones 2004).² Failure to properly account for temporal dependence can lead to biased, usually too small, standard errors. Further, depending on the cause of the temporal dependence, estimates of causal effects³ can be biased and inconsistent. It is, therefore, imperative that scholars who seek to draw causal inferences understand well the implications of temporal dependence.

One especially common response to temporal dependence involves *conditioning* on (possibly transformed) lagged dependent variables (LDVs). *Conditioning* is the process of adjusting an estimate based on the values of a set of covariates (the “conditioning set”). This adjustment is usually done by including control variables in a regression model, though adjustment can also be done using other methods such as stratification, matching, and inverse-probability weighting. A lagged dependent variable is the dependent variable from a previous time period. Event history models based on “gap time”⁴ implicitly condition on (transformed) LDVs (see Online Appendix A1).

In some areas of social science, such as in international relations, almost all work conditions on (transformed) LDVs; it is taken for granted that one should “control for temporal dependence.” For example, there was a recent exchange in *Political Analysis* about the appropriate way to control for temporal dependence, with Carter and Signorino (2010) recommending cubic polynomials of time since the last event against cubic splines of time since last event (Beck et al. 1998); these are each a specific transformation of the LDVs, see Online Appendix A1. Assumed by this conversation was that controlling for some function of the LDVs would improve estimates of the causal parameters of interest. This article contributes to this conversation by making clear that this is not necessarily the case.⁵ In fact, for the case of the study of the democratic peace, which this article considers, the identification assumptions for excluding LDVs are arguably more plausible than the identification assumptions for including them.

Estimates are not necessarily improved by controlling for temporal dependence, no matter what specification is used. Whether estimates will be improved or harmed depends on causal assumptions that need to be explicated and justified.⁶ This article will systematically show what nonparametric assumptions are needed to justify using (transformed) LDVs for causal identification. This article joins other work that has pointed out some of the causal assumptions required of different temporal specifications (Glynn and Quinn 2013; Keele and Kelly 2006; Morgan and Winship 2015:chap. 11; Wilson and Butler 2007) and of the dangers of controlling for LDVs (Achen 2000). More broadly, it speaks to the importance of understanding the assumptions underlying our causal estimates (Collier, Sekhon, and Stark 2010; Dunning 2008; Sekhon 2009; Shalizi and Thomas 2011).

Nonparametric identification involves identifying causal effects without any knowledge of the functional form of the causal processes. Nonparametric identification is a useful place to start for a few reasons. First, the question of nonparametric identification is in one sense prior to the question of parametric identification, since if an effect can be nonparametrically identified then it can be parametrically identified, and we need not rely on strong parametric assumptions to do so. Second, in social science we rarely have confidence about the functional form of causal processes; accordingly, inferences that don't depend on parametric assumptions are often more credible. Third, insights arising from studying nonparametric identification provide a foundation for understanding the challenges for identification in many of the species of parametric models, including better understanding of the existence and character of particular kinds of biases, such as omitted variable, included variable, selection, and measurement biases (works doing so using similar tools as this article include Elwert and Winship 2014; Glynn and Gerring 2013; Glynn and Kashin 2013; Glynn and Quinn 2013; Morgan and Winship 2015; Pearl 2009b).

Scholars can determine the appropriate conditioning set for nonparametric identification using the tools of structural causal models (Pearl 2009b). Using these tools, this article establishes that nonparametric identification is not possible for a general temporal causal process (process \mathcal{D}). Not conditioning on the LDV leads to confounding biases; conditioning on the LDV leads to collider biases. Further, this result generalizes to the broader class of temporal causal processes, \mathcal{D}_G , whose causal graph contains process \mathcal{D} as a subgraph.

This article then articulates the minimal assumptions on process \mathcal{D} , and therefore also necessary assumptions on \mathcal{D}_G , required to nonparametrically

identify the causal effect using backdoor adjustment.⁷ The principle take-away result is that we should not condition on LDVs if we can assume *both* that there is no reverse causation (no effect of Y_{t-k} , the lagged outcome, on treatment) *and* no outcome autocausation (no effect of Y_{t-k} on Y_t not mediated by treatment); we should condition on LDVs if we can assume *either* that there are no unobserved common causes of treatment and the lagged outcome, *or* no unobserved persistent causes of the outcome. However, the causal processes implied by the above sets of minimal identification assumptions are observationally indistinguishable; the data alone can not tell us which conditioning strategy is appropriate. This article then illustrates how to think about these empirical commitments using the study of the democratic peace as an example.

To summarize, temporal dependence is a symptom of potentially deep problems for causal inference. Overcoming these problems requires making causal assumptions. These assumptions should be explicated and justified.⁸ This article will show how to do so.

Temporal Dependence in Social Science

Temporal dependence is the property that, after conditioning on covariates, observations of the same unit are dependent across time. In the context of parametric models, temporal dependence is the property that the error term is dependent across time. Temporal dependence is typically diagnosed by looking for autocorrelation in the residuals of a regression (e.g., Greene 2012:§20.7). Since most statistical procedures assume independence of observations, conditional on some covariates,⁹ the belief that data are temporally dependent often means that a central assumption is violated.

Temporal dependence can arise from many causal processes, such as omitted unit-specific effects, omitted temporally correlated variables, autocorrelated disturbances, and outcome autocausation (an effect of $Y_{i,t-k}$ on $Y_{i,t}$ not going through treatment).¹⁰ Some of these causes of temporal dependence are relatively benign, making estimates of standard errors inconsistent, but not biasing estimates of the causal parameters of interest. Other causes of temporal dependence will be more harmful, making it impossible to identify (and hence have consistent estimators for) causal parameters of interest. A large toolbox of solutions have been developed to address particular sources of temporal dependence, which include: estimating a multilevel model to account for cross-sectional unit-specific effects (Gelman and Hill 2006; Greene 2012:§11.4-11.5; Wooldridge 2010:chap. 10-11); conditioning on lags of the causal factor or other variables; conditioning on lags of the

outcome, sometimes with a specific structure as with first-difference models (Allison 1990; Morgan and Winship 2015:chap. 11); quasi-differencing the outcome (Cochrane and Orcutt 1949; Prais and Winsten 1954); and estimating the variance–covariance matrix so as to allow for serial correlation such as with the Newey–West robust consistent estimator for serial correlation (Newey and West 1987; also see Beck and Katz 1995; Freedman 2006; King and Roberts 2015). For reviews, see Beck and Katz (2011), Greene (2012), Hamilton (1994), King (1998), Wilson and Butler (2007), and Wooldridge (2009, 2010).

Each of these methods is justified by showing that they lead to estimators with desirable properties *for particular causal processes*. Thus, a crucial step in the use of any method for causal inference is the articulation and justification of the causal assumptions underlying the method. In practice, however, most articles seem to treat temporal dependence as a problem amenable to a “technical fix,” offering no statement or defense of the causal conditions required for their method (Wilson and Butler 2007). Part of the difficulty is that the identification assumptions for many methods are bundled together and expressed as a parametric model, which are hard to unpack and judge in terms of the relative plausibility of particular causal assumptions. By contrast, social scientists are most able to evaluate claims about the existence of causal effects, sometimes the sign of effects, and sometimes the rough magnitude of effects but rarely the functional form of effects.

This article addresses this shortcoming by offering a precise statement of the nonparametric causal commitments required to identify causal effects under temporal dependence and to justify the decision to condition or not to condition on LDVs. To do so, we now introduce the problem of nonparametric identification and the tool of causal graphs for analyzing it.

Nonparametric Identification under Temporal Dependence

Social scientists want consistent estimators for causal effects.¹¹ In order to have a consistent estimator for a causal effect, we must be able to identify it. A causal effect is *identifiable* if we could determine its value, given an “unlimited number of observations” (Manski 2007:3).¹² An effect is nonparametrically identifiable if we can identify the effect without knowledge about the functional form of causal processes.¹³

Nonparametric identification can be evaluated using a set of methods involving causal graphs, referred to by some as *structural causal models*

(Pearl 2009a). Structural causal models provide a means for evaluating the circumstances under which conditioning on a particular set of covariates could, given unlimited data, nonparametrically recover causal effects. Structural causal models are similar to structural equation models but rather than identify effects using parametric assumptions, they identify effects using assumptions about the nonexistence of certain causal effects that make the causal model *Markovian* and *acyclic* (see below; Pearl 2009a:§3.2.3). If we have additional insight about functional form then we can supplement these tools with those structural features, but typically in social science we are not confident about functional form. Structural causal models are also consistent with the potential outcomes framework.¹⁴

For textbook references on nonparametric identification using causal graphs, and closely related concepts, see Hernan and Robins (2015), Morgan and Winship (2015), Pearl (2009a), Shalizi (*in press*:chap. 20-24), Spirtes, Glymour, and Scheines (2001). Because these methods are relatively new to social science, the following subsection (and Online Appendix A) introduces some key concepts.

Nonparametric Identification Using Causal Graphs

Following the Neyman–Rubin model of potential outcomes (see Online Appendix A.2), the effect of W is identifiable if treatment assignment is conditionally *ignorable* (Rosenbaum and Rubin 1983; Stone 1993): $Y(w) \perp\!\!\!\perp W|C$ and $P(W = w|C) > 0$ for all w (the latter expression called *positivity*, see Hernan and Robins 2015:chap. 3). C is a set of covariates, such that after conditioning on C treatment assignment is independent of the potential values of the outcome. In order to find such a set of covariates C for any particular situation, we can describe the causal process (nonparametrically) using causal graphs and then use the following graphical tools to find C .

A directed graph is a set of vertices (\mathcal{V}) and a set of directed edges (\mathcal{E}). For example, graph $\mathcal{D}_{A_4 \wedge A_5}$, depicted in Figure 1, denotes $\{\mathcal{V} = \{c', e', Y', W', Y\}, \mathcal{E} = \{c'W, c'Y', e'Y', e'Y, WY\}\}$ (this graph and its subscripts are explained below).

A *directed acyclic graph* (DAG), like graph $\mathcal{D}_{A_4 \wedge A_5}$, is a directed graph with no cycles: no sequence of directed edges that connect a vertex to itself (Pearl 2009b:§1.2.1). A directed edge between two vertices W and Y in a causal DAG, denoted inline as $W \rightarrow Y$, implies a direct causal effect of W on Y (more on this below).¹⁵ A *path* is a sequence of connected edges. *Parents* of vertex set W are those nodes that have directed edges going into W . Denote

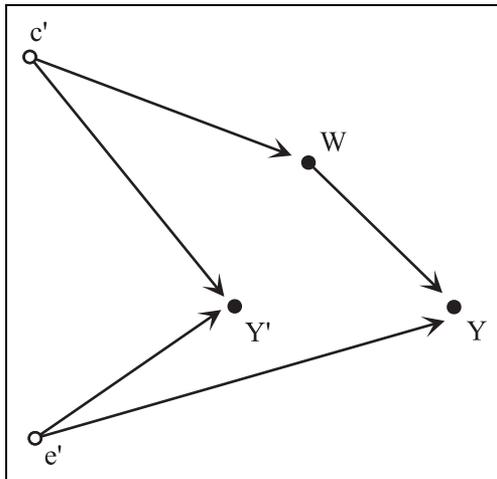


Figure 1. Graph $D_{A_4 \wedge A_5}$.

these $PA(W)$. In this case, $PA(W) = \{c'\}$. *Descendants* of W , $D(W)$, are nodes that have a sequence of directed edges going from W to them; in this case $D(W) = \{Y\}$. *Nondescendants* of W , $ND(W)$, are those nodes that don't have a sequence of directed edges going from W to them; in this case, $ND(W) = \{c', Y', e'\}$. Other kinship terminology can be similarly used, such as *children*, *ancestors*, and *spouses*.

In order to apply DAGs to any given problem, we need to know when we can represent a causal process using a causal graph. A nonparametric process¹⁶ D can be analyzed using a DAG G , if the probability function on observable variables P that D generates is *Markov relative to G* , which means that every variable in P is independent of its nondescendants in G , conditional on its parents (Pearl 2009b:theorem 1.2.7). A process D and probability distribution P that are Markov relative to a DAG G are said to satisfy the *Parental Markov Condition*. G is a *causal DAG* if G represents a causal process that is Markov relative to G . In simple terms, a causal process D can be depicted using a causal DAG if no variable can affect itself (no cycles), and the only causal effects not depicted are independent disturbances.

To establish ignorable treatment assignment, we want to determine whether treatment is independent of the potential outcomes. To do so, we will use a property called *d-separation* (d is for directional), which refers to whether two variables are not causally connected to each other. If two

variables in a causal DAG are *d-separated*, then they will be statistically independent. On the other hand, if two variables are *d-connected*, then they will typically¹⁷ be statistically dependent. See the Online Appendix A.3 for the formal definition of *d-separation*.

There are three basic kinds of causal relationships that will *d-connect* two variables. (1) Two variables will be *d-connected* if one is the direct cause of the other; for example, in Figure 1, c' and W are *d-connected* because $c' \rightarrow W$, (2) two variables will be *d-connected* if they share a common cause that is not conditioned upon; in Figure 1, W and Y' are *d-connected* because $Y' \leftarrow c' \rightarrow W$, and (3) two variables will be *d-connected* if they both affect a variable that is conditioned upon; in Figure 1, c' and e' are marginally *d-separated*, but become *d-connected* if Y' is conditioned upon. This can be denoted in line as $c' \rightarrow \boxed{Y'} \leftarrow e'$; the Y' denotes that $\boxed{Y'}$ is conditioned upon, in which case we say that Y' is an activated collider between c' and e' , more on this below.

In general, any two variables will be *d-connected* if they are on a path of *d-connected* variables. For example, in Figure 1, Y' and Y are *d-connected* through two paths: $Y' \leftarrow c' \rightarrow W \rightarrow Y$ and $Y' \leftarrow e' \rightarrow Y$. If we conditioned on Y' , then W and Y would be *d-connected* through Y' because $W \leftarrow c' \rightarrow \boxed{Y'} \leftarrow e' \rightarrow Y$.¹⁸ An indirect causal effect is a sequence of directed edges with all arrows pointing in the same direction; denote an indirect effect as $W \rightarrow Y \Leftrightarrow W \rightarrow m_1 \rightarrow \dots \rightarrow m_k \rightarrow Y$ for some set of mediators $\{m_1, \dots, m_k\} \subseteq D(W) \cap A(Y)$, where $A(Y)$ denotes the set of ancestors of Y . Note that this definition includes the direct effect.

A *collider* is a variable that is affected by two other variables (Y' in $c' \rightarrow Y' \leftarrow e'$), so called because the causal arrows collide into it. An unactivated collider is a collider that has not been conditioned upon; an activated collider is a collider that has been conditioned upon. Dependence does not flow through an unactivated collider, whereas it does flow through an activated collider. To see this, consider the following example. Suppose permanent residence to a country (Y') is awarded in part to those who receive high numbers in a random lottery (c') and in part to those with valuable skills (e'), then $c' \rightarrow Y' \leftarrow e'$ and clearly c' and e' are marginally independent of each other even though they both cause Y' . However, if we activate the collider (condition on it) then we will make c' and e' dependent. Suppose we only look at individuals who won permanent residence (condition on $Y' = 1$). Then those with low lottery numbers (c') who won residency ($Y' = 1$) will tend to have more valuable skills (e') because they needed those skills to win residency. Though c' and e' are marginally independent, they will be correlated within levels of Y' .

To summarize, dependence in a causal DAG can (and typically will) flow through a path that consists only of unblocked chains of causation, unblocked common causes, and activated colliders. Dependence will not flow through unactivated colliders. If there is no path d -connecting two variables, then they are d -separated and statistically independent.

Finally, to identify the causal effect of W on Y we want to isolate all causal paths $W \rightarrow Y$, so called *front-door paths* because they involve causal paths that begin with an arrow pointing out of W (out the “front-door” of W). To do so, we need to block all confounding causal paths between W and Y , so called *back-door paths* because they involve an arrow pointing into W (into the “back-door” of W). If we are able to find a set of observed variables C that blocks all back-door paths, then we can identify the effect of W on Y , and we say that C satisfies the *back-door criterion* relative to (W, Y) (see Online Appendix A.4 for the formal definition). When there exists a C that satisfies the back-door criterion relative to (W, Y) the causal effect is identifiable using the *back-door adjustment* formula (see Online Appendix A.4). However, when selecting C we have to be careful not to condition on colliders that would unblock a back-door path, inducing bias. This gives rise to the counterintuitive result that conditioning on a pretreatment covariate can induce bias to what was otherwise an unbiased estimator. This is known as M bias because the causal graph makes an “ M ” (Greenland, 2003; Ding and Miratrix, 2015; Thoemmes, 2015).¹⁹ In fact, it is precisely because of M bias that it is sometimes problematic to condition on LDVs, as examined below.

We can now put all of this together. If a causal process D (and induced probability distribution P) satisfy the parental Markov condition, then D can be represented by causal DAG G . Using G , we can look for a set of covariates C that will block all spurious (back-door) causal paths between W and Y , leaving only the desired (front-door) causal effects of W on Y . However, we have to be careful in selecting C that it does not activate a collider that opens a back-door path. If we find such a C that satisfies the back-door criterion relative to (W, Y) then, given infinite data and positivity, we can recover the causal effect—the full marginal distributions of the potential outcomes—by conditioning on C and using the back-door adjustment formula.

A General Causal Process: Process \mathcal{D}

This section will introduce a general temporal causal process, denoted process \mathcal{D} . To do so, I adopt as few assumptions about the causal process as possible and I enumerate all assumptions that I make. Assumptions in a causal DAG consist of statements about the absence of causation between observed

variables, the absence of unobserved shared causes between observed variables, and the absence of other observed variables that are d -connected to at least two variables.

For ease of exposition, we first assume that there are only three observable variables of interest: the lagged outcome Y' , the treatment variable W (the causal factor of interest), and the outcome Y .

(\mathcal{A}_1) observables: $\mathcal{V}_O = \{Y', W, Y\}$ where \mathcal{V}_O is the set of observable variables.

Later, we will weaken this assumption, demonstrating that the main result generalizes to causal processes with many observable variables, so long as the new variables do not provide exhaustive mechanisms for certain effects in process \mathcal{D} .²⁰

In order for W to cause Y , W must occur prior to Y , therefore (\mathcal{A}_1) Y is post-treatment, which implies that $Y \not\rightarrow W$ (or equivalently, $W \in ND(Y)$).

For ease of exposition, I will also assume that Y' occurs prior to W . Modifying this assumption will not change any of the main results.

(\mathcal{A}_2) Y' is pretreatment: $W \not\rightarrow Y'$

In order for it to be feasible to identify the effect of W on Y , we must assume: (\mathcal{A}_3) No fundamental confounding: $\nexists \varepsilon$ s.t. $W \leftarrow \varepsilon \rightarrow Y$ and ε is unobservable.

In process \mathcal{D} , all pairs of causal factors are allowed to have unobserved common causes, except for W and Y (by \mathcal{A}_3). c' denotes all unobserved common causes of W and Y' : $W \leftarrow c' \rightarrow Y'$. e' denotes all unobserved common causes of Y' and Y (persistent causes of the outcome): $Y' \leftarrow e' \rightarrow Y$. Process \mathcal{D} can then be represented by the set of expressions below, where \leftarrow denotes a deterministic causal relationship.²¹ $f_i(\cdot)$ represents some unknown function, and v_i denotes an independent disturbance and serves as a notational reminder that the distribution of the observed variables is probabilistic.

This process could generate data in a time series, in a cross section, or both, so long as the observations are sufficiently independent that an infinite draw of them would identify all relevant quantities. For a depiction of how this process could look like when strung together in a time series, see Online Appendix A.5. For discussion of other issues, causal quantities, and identification strategies in time-series data, see Blackwell (2013), Blackwell and Glynn (2014), Pearl and Robins (1995), Robins, Hernan, and Brumback (2000).

Because \mathcal{D} has independent errors, and all causality is directed acyclic, process \mathcal{D} satisfies the parental Markov condition and can therefore be represented

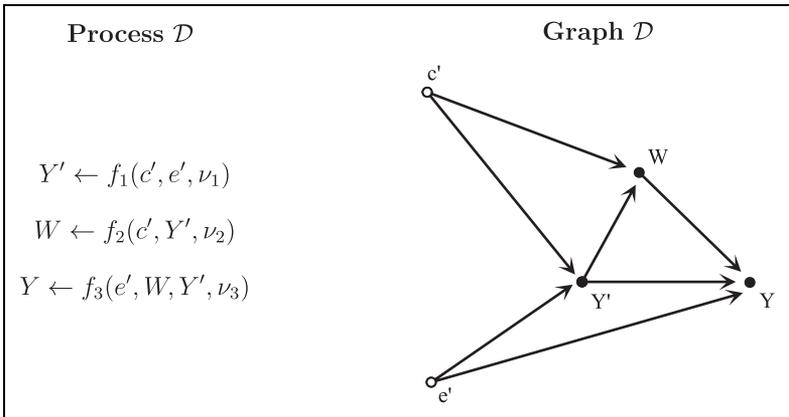


Figure 2. Process \mathcal{D} and Graph \mathcal{D} .

by a DAG, specifically Graph \mathcal{D} in Figure 2 and analyzed using the tools of structural causal models (Pearl 2009b:theorem 1.2.7; Shalizi **in press**:§20.2).

Backdoor Paths in Process \mathcal{D}

To identify the causal effect (through backdoor adjustment) in process \mathcal{D} , we want to find a conditioning set C that satisfies the backdoor criterion. For a given conditioning set, the backdoor criterion will not be satisfied if there is an open backdoor path, which is a path of dependence (d -connected nodes) connecting W and Y , starting with an arrow pointing into W (in the “back door” of W).

In process \mathcal{D} there are only two possible conditioning sets: the set containing the LDV, $C = \{Y'\}$, and the empty set, $C = \{\}$. When conditioning on the LDV, $C = \{Y'\}$, there are open backdoor paths: $B_{coll} = \{W \leftarrow c' \rightarrow \boxed{Y'} \leftarrow e' \rightarrow Y\}$ (where the box denotes conditioning). These are denoted as B_{coll} because they involve a bias arising from conditioning on a collider (the LDV); “ B ” is used to serve as a reminder that these refer to “backdoor paths” that, when open, will induce a “Bias”.

There are also open backdoor paths when not conditioning on the LDV, $C = \{\}$. Denote these as B_{con} because they involve a confounding bias that could be blocked through the LDV. $B_{con} = \{B_{RC1}, B_{RC2}, B_{AC}\}$, $B_{RC1} = \{W \leftarrow Y' \rightarrow Y\}$, $B_{RC2} = \{W \leftarrow Y' \leftarrow e' \rightarrow Y\}$, and $B_{AC} = \{W \leftarrow c' \rightarrow Y' \rightarrow Y\}$. RC refers to paths involving “reverse causality” (Y' affects W). AC refers to paths not involving reverse causation but

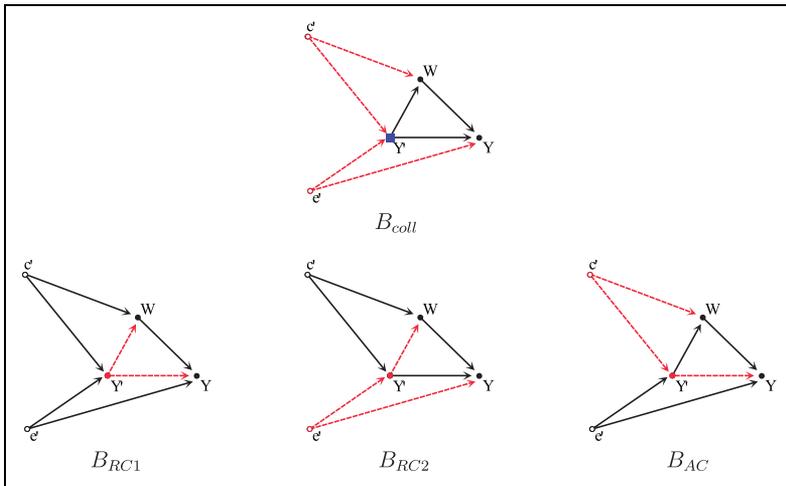


Figure 3. Sources of bias. The red dashed arrows denote the biasing paths. The top figure illustrates the collider biases B_{coll} that arise when conditioning on Y' . The blue square denotes conditioning on this variable. The bottom figures illustrate the confounding biases B_{con} that arise when not conditioning on Y' .

involving outcome **autocausation**.²² These backdoor paths are depicted with dashed red edges in Figure 3.

Therefore, there is no conditioning set that satisfies the backdoor criterion for process \mathcal{D} . There are confounding biases when not conditioning on the LDV; there are collider biases when conditioning on the LDV. Coupled with the fact that there is no exhaustive mediator of the treatment effect, this gives us the first result (the proof is in Online Appendix A.6).

Proposition 1: The effect of W on Y is not identifiable in process \mathcal{D} .

Generalization to Causal Processes with Other Observable Variables

For process \mathcal{D} , there is no identification strategy. To what extent will this result generalize to other nonparametric causal processes in which there are additional observable variables? Informally, the answer is that it will generalize so long as the new variables do not provide exhaustive mechanisms for any of the causal paths in process \mathcal{D} (formally, the answer is that the result will generalize so long as there remains a *hedge* in the new graph, see

Shpitser and Pearl 2006). When exhaustive mechanisms exist, various solutions become possible.

To see this, note that the fundamental problem for nonparametric identification in process \mathcal{D} is that when we don't condition on the LDV there is a spurious association through B_{con} and if we do condition on the LDV we induce spurious association through B_{coll} . Adding observable variables to process \mathcal{D} can not change this basic dilemma so long as they don't exhaustively mediate all of the causal connections on one of these sets of paths of spurious association.²³ We can therefore generalize Proposition 1 to the set of nonparametric causal processes, denoted \mathcal{D}_G for general, that have graph \mathcal{D} as a subgraph (proof in Online Appendix A.7):

Theorem 1. The effect of W on Y is not nonparametrically identifiable in any causal process, \mathcal{D}_G , whose graph has graph \mathcal{D} as a subgraph.²⁴

Nonparametric identification through backdoor adjustment can only become less attainable in \mathcal{D}_G : for any conditioning set C_1 , backdoor paths that are open in process \mathcal{D} will also be open in \mathcal{D}_G but there may now also be additional open backdoor paths in \mathcal{D}_G .

To summarize, Theorem 1 shows that causal effects are not identified for a large class of nonparametric causal processes. Implicit to this theorem is the insight that any time series methods that can identify the effect of W in \mathcal{D}_G must be identifying off of parametric assumptions (typically, additive effects in a linear link function) or other assumptions not made in \mathcal{D}_G . For example, as noted by Beck and Katz (2011, 339), Hamilton (1994:226) shows how in a linear version of \mathcal{D} , in which e' consists of autoregressive AR (1) disturbances, the effect of W can be identified using an ADL(2,1) model (autoregressive distributed lag model of order 2 in autoregression and order 1 in distributed lags). That is, the analyst estimates $Y_{i,t} = \alpha_0 + \alpha_1 Y_{i,t-1} + \alpha_2 Y_{i,t-2} + \alpha_3 W_{i,t} + \alpha_4 W_{i,t-1} + v_{i,t}$ and α_3 will consistently estimate the (assumed constant) effect of W_t on Y_t . This identification strategy breaks down under weakening of the parametric assumptions, such as by allowing the autocorrelation in disturbances to include a moving average term or by allowing the link function to be nonlinear.

Methodological texts typically focus on special cases of \mathcal{D}_G , often parametric special cases, in which identification is possible. However, such a focus may mislead scholars by understating the difficulty of eliminating biases. Social phenomena are complex and our knowledge of them limited. Even when we restrict ourselves to causal processes that are not fundamentally confounded (\mathcal{A}_3), the causal process may still not permit causal

identification, no matter what conditioning strategy we use. When not conditioning on LDVs, we will have bias from B_{con} ; when conditioning on LDVs, we will have bias from B_{coll} . In such circumstances, a pragmatic way forward will be to evaluate our estimators under different presumed causal processes (e.g., Beck and Katz 1995, 1996; Bertrand, Duflo, and Mullainathan 2002; Box-Steffensmeir, Boef, and Joyce 2007; Franzese and Hays 2007; Freedman 2005; Glynn and Quinn 2013; Keele 2010; Keele and Kelly 2006) to sign and bound these biases (Blackwell 2014; Rosenbaum 2002) and qualify our inferences appropriately.

If practitioners do move to a special case of \mathcal{D}_G in which causal identification is possible, the move should be done with caution, detailing, and justifying the additional assumptions employed for causal identification. To assist scholars in understanding the assumptions underlying claims to causal identification, the following section will systematically state the minimal nonparametric assumptions needed for identification in process \mathcal{D} .

Assumptions for Identification

Nonparametric identification in process \mathcal{D} , and therefore also \mathcal{D}_G , is impossible because of the two sources of bias, B_{coll} and B_{con} . Nonparametric identification (through backdoor adjustment) requires that one of these backdoor paths be absent. This section will introduce the assumptions we can make on process \mathcal{D} and will offer some substantive discussion to illustrate how scholars might want to evaluate their plausibility. This section will then systematically summarize the minimal nonparametric conditions²⁵ that must be assumed for process \mathcal{D} (and \mathcal{D}_G) to rule out these backdoor paths so as to permit identification.

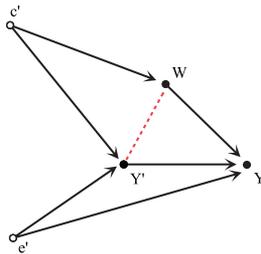
Discussion of Identification Conditions

To identify the causal effect (via backdoor adjustment), we need to assume away some of the causal connections (edges) so as to break one of the backdoor paths. There are four sets of causal effects (edges) in process \mathcal{D} that we could assume away. I discuss these roughly in order of my assessment of their plausibility for causal processes of interest to social scientists. I will also not evaluate whether these conditions are strictly true, since for nonexperimental social phenomena it is hard to rule out any causal connection. Instead I will evaluate whether the conditions could be approximately true, so that any net effects are likely to be small, as smaller effects will typically lead to smaller biases.

To illustrate how these conditions can be evaluated I will discuss them in the context of the study of the democratic peace (Dafoe 2011; Hayes 2012;

Russett 1993). The causal proposition at the heart of the democratic peace is that some set of characteristics of democratic regimes promotes peace amongst democracies (or inversely, some characteristics of autocratic regimes promote conflict against democracies). Evaluating these conditions is not an easy task as it requires assessing the plausibility of several specific kinds of causal connections. Accordingly, the following discussion is merely an illustration of how scholars can approach the task: what an argument for or against a condition would look like. Future research could devote more concentrated efforts into evaluating each of these conditions for specific fields. Online Appendix A.9 also briefly considers two other examples.

(\mathcal{A}_4) No “reverse causation”: $Y' \not\rightarrow W$. (\mathcal{A}_4) No reverse causation implies that the (lagged) outcome (e.g., whether a war occurred last year) does not have an effect on the causal factor of interest (whether the government is democratic). The red dashed edge in the accompanying figure depicts the effect that is assumed to be absent.

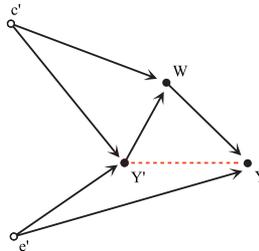


To argue that this condition is likely to be false requires arguing that reverse causation is likely: that war occurrence affects regime type. To make such an argument, we could appeal to examples where the occurrence of war had a profound effect on regime type. For example, after World War II, the United States and Soviet Union each promoted democratic and communist governments in their respective spheres of influence. We could also appeal to scholarship that has investigated this specific causal connection. For example, Thompson (1996) argues for reverse causation: zones of peace promote democracy. However, interpreting claims about causal connections are subtle because different causal connections often look similar to each other. Thompson’s argument illustrates this, since Thompson (1996) also invokes “aspirations to regional hegemony” and “domestic concentrations of economic and political power” as common causes of regional conflict and autocracy, which would be better characterized as common causes of W and Y' ($Y' \leftarrow c' \rightarrow W$) not as $Y' \rightarrow W$.

To argue that condition \mathcal{A}_4 is likely to be true requires arguing that reverse causation is unlikely: that war occurrence does not have systematic effects on regime type. To do so, we could argue that political institutions are deeply rooted and highly persistent (Acemoglu et al. 2008), immune to all but the most penetrating conflicts. Thus, we could say that \mathcal{A}_4 is more likely to be true for the study of smaller level conflicts, like militarized interstate disputes, than for the study of large wars. In response to the World War II example, we could argue that the effects seem as likely to be positive (more democratic) as negative (less democratic), so that the average effect is approximately zero. We could also invoke other theory and research that finds no evidence of an effect of war occurrence on regime type (Mousseau and Shi 1999).

In my assessment, (\mathcal{A}_4) no reverse causation seems to be approximately true for the democratic peace, at least as applied to low level conflict events.

(\mathcal{A}_5) *No outcome autocausation (not mediated by treatment):* $Y' \nrightarrow Y$. \mathcal{A}_5 assumes that the realization of the outcome (war this year) does not have a causal effect on the outcome in future time periods (war in the future), other than through treatment (democracy).

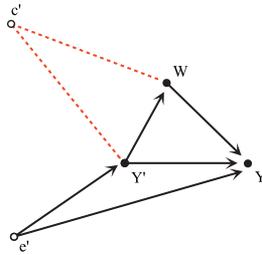


We could argue that this condition is likely to be false for the democratic peace by pointing out the various ways that war occurrence makes war more (or less) likely in the future. The occurrence of war could make war more likely through the hardening of hatred between peoples, the strengthening of hawkish domestic coalitions, and the centralization of authority. War could make future war less likely through the depletion of military and industrial resources, the exhaustion of resolve, the revelation of information, or the resolution of disputes.

We could argue that this condition is likely to be (approximately) true by showing how these purported mechanisms are weak, or that the positive and negative effects balance so that the average effect is close to zero. Note that outcome autocausation can not be identified using autocorrelation in the outcome unless we believe we have controlled for all persistent causes of the outcome (all e').

In my assessment no outcome autocausation (\mathcal{A}_5) is somewhat plausible for the democratic peace.

(\mathcal{A}_6) *No unobserved common causes of W and Y' .* $\forall c'$, either $c' \nrightarrow Y'$ OR $c' \nrightarrow W$. \mathcal{A}_6 assumes that there are no unobserved factors that affect both W (democracy) and Y' (lagged peace).



Whereas \mathcal{A}_4 and \mathcal{A}_5 involved assuming no causal effect between known and observed factors, \mathcal{A}_6 and \mathcal{A}_7 will require making assumptions about the absence of causal effects on a subset of all other factors, known and unknown. For this reason \mathcal{A}_6 and \mathcal{A}_7 are much harder to empirically corroborate; they cannot, even in principle, be definitively evaluated through experiments, and our ability to interrogate them is constrained by our imagination about possible causes.

(\mathcal{A}_6 (W and Y' do not have any common unobserved causes) is very similar to the condition of no confounding (W and Y do not have any common unobserved causes). As with the proposition of no confounding, we can affirm \mathcal{A}_6 if we have confidence that treatment is as if randomly assigned, independently across time, which implies \mathcal{A}_6 and \mathcal{A}_4 .²⁶

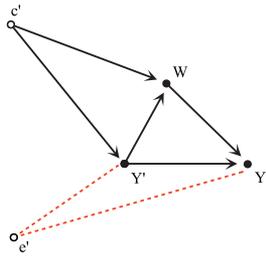
We could argue that \mathcal{A}_6 is approximately true for the democratic peace in a similar manner as we would argue that democracy is not fundamentally confounded with peace. We could argue that we have adequately conditioned on the most important common causes of democracy and (lagged) peace. We could diagnose this claim through placebo tests, for example, by looking to see whether some pretreatment variable that doesn't affect democracy is uncorrelated with democracy. We could argue that \mathcal{A}_6 is false by pointing out important common causes that are not adequately controlled and through failed placebo tests. In my assessment, \mathcal{A}_6 is not approximately true for the democratic peace: There are many common causes of democracy and lagged peace that have not been (and cannot be) adequately conditioned away (for a *descriptive* perspective on the democratic peace, see Dafoe 2011; Dafoe, Oneal, and Russett 2013).

Unlike \mathcal{A}_4 and \mathcal{A}_5 , \mathcal{A}_6 can be made true by improved theory and empirics. Specifically, if we know all of the common causes of W and Y' and have

sufficiently good measures of them, then we can make \mathcal{A}_6 true by conditioning on them. In addition, if we can find *exhaustive isolated mechanisms* for the causal effect of any particular c' on Y' or c' on W then we can condition on them to block this source of bias. Graphically, to be an *exhaustive* mechanism requires that on every path $Y' \leftarrow c' \rightarrow W$ there is an observed mediator. To be an *isolated* mechanism requires that this mediator not be a collider on another backdoor path (e.g., it is not a consequence of e').²⁷

One especially promising mechanism is W' , the lagged treatment, because it is often plausible that much of the dependence between W and Y' flows through the dependence between W and W' . For example, suppose $W_{i,t} = \sum_{k=0}^t \varepsilon_{i,t-k}$, where $\varepsilon_{i,t}$ is assigned randomly independently per unit and per time period. This would be the case in a panel study of the effect of total lottery winnings, where $\varepsilon_{i,t}$ would be the per period lottery winnings and $W_{i,t}$ would be the total lottery winnings. Even though $\varepsilon_{i,t}$ could be independently and randomly assigned (over time and across subjects), $W_{i,t}$ would not be independent over time. In such a context, conditioning on the lagged treatment (W') would make \mathcal{A}_6 true. Given the generality of the situation in which some dependence between W and Y' flows through W' , it seems like a good general recommendation to consider controlling for lags of the treatment. This recommendation is especially made by Glynn and Quinn (2013) and is found (in its parametric form) in the recommendation to employ (auto regressive) distributed lag or error correction models that include the lagged independent variables (Beck and Katz 2011; De Boef and Keele 2008). Nevertheless, conditioning on W' is not guaranteed to make \mathcal{A}_6 true, since there could still be other unobserved common causes of Y' and W .²⁸

(\mathcal{A}_7) *No unobserved persistent causes of the outcome.* $\forall e'$, either $e' \rightarrow Y'$ OR $e' \rightarrow Y$. \mathcal{A}_7 assumes that there are no unobserved factors that affect both Y' (lagged peace) and Y (peace).



Like \mathcal{A}_6 , \mathcal{A}_7 is a very strong assumption since it assumes away a constellation of causal effects for all other known and unknown factors, empirical

interrogation of \mathcal{A}_7 is limited by our imagination and patience, and \mathcal{A}_7 cannot be experimentally verified or disproven, even in principle. \mathcal{A}_7 is especially implausible since we can be confident that we have not observed all causes of the outcome, and causes of the outcome, like most phenomena, are usually temporally persistent. For the democratic peace, we could argue against \mathcal{A}_7 by pointing out the many temporally persistent causes of peace and war that cannot be adequately controlled for: unresolved disputes over territory, proximity, enduring rivalries, deep-rooted amity or enmity, historical grievances, and geopolitical insecurity. \mathcal{A}_7 becomes more plausible to the extent that we think we are aware of and can adequately measure the most important persistent causes of the outcome. I regard \mathcal{A}_7 as false for the democratic peace.

Minimal Identification Assumptions

The following theorem will now state the minimal²⁹ assumptions that we need to make about process \mathcal{D} for identification. These sets of assumptions will correspond to the maximal³⁰ subgraphs of process \mathcal{D} that permit identification for a particular conditioning set. These assumptions achieve identification by eliminating either B_{coll} or B_{con} .

Theorem 2 *Minimal Assumptions for Identification*

Possible nonparametric assumptions on Process \mathcal{D} are:

(\mathcal{A}_4) No “Reverse Causation”: $Y' \nrightarrow W$

(\mathcal{A}_5) No Outcome Autocausation (not Mediated through Treatment): $Y' \nrightarrow Y$

(\mathcal{A}_6) No Unobserved Common Causes of W and Y' : $\forall c'$, either $c' \nrightarrow Y'$ OR $c' \nrightarrow W$.

(\mathcal{A}_7) No Unobserved Persistent Causes of the Outcome: $\forall e'$, either $e' \nrightarrow Y'$ OR $e' \nrightarrow Y$.

The maximal subgraphs of Process \mathcal{D} that permit identification for a particular conditioning set are:

By eliminating B_{coll} : $\mathcal{A}_{coll}^I = \mathcal{A}_6$ OR \mathcal{A}_7

By eliminating B_{con} : $\mathcal{A}_{con}^I = (\mathcal{A}_5$ AND $\mathcal{A}_7)$ OR $(\mathcal{A}_4$ AND $\mathcal{A}_5)$ OR $(\mathcal{A}_4$ AND $\mathcal{A}_6)$

Given temporal dependence, B_{con} simplifies to: $(\mathcal{A}_4$ AND $\mathcal{A}_5)$ OR $(\mathcal{A}_4$ AND $\mathcal{A}_6)$

When B_{coll} are absent, nonparametric identification is possible with conditioning set $\mathbf{C} = \{Y'\}$. When B_{con} are absent, nonparametric identification is possible with $\mathbf{C} = \{\}$.

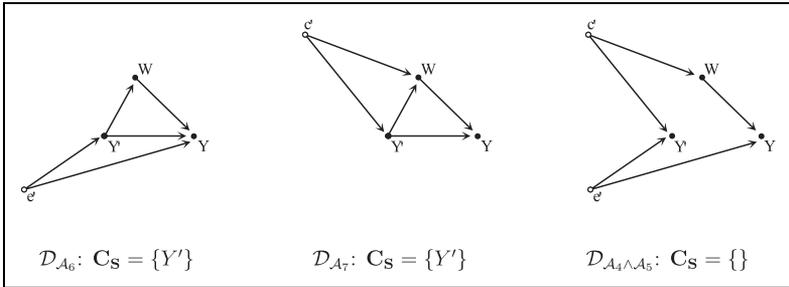


Figure 4. The three maximal subgraphs of D that permit identification. Subscript denotes the (positive) assumptions that define the graph. C_S denotes the sufficient conditioning set.

Proof. To prove Theorem 2, we state all sets of assumptions that block B_{coll} (and then B_{con}) and simplify.

Recall that $B_{coll} = \{W \leftarrow c' \rightarrow \overline{Y'} \leftarrow e' \rightarrow Y\}$. Therefore, B_{coll} will be eliminated if we assume: A_6 OR A_7 .

Recall that $B_{con} = \{B_{RC1}, B_{RC2}, B_{AC}\}$; $B_{RC1} = \{W \leftarrow Y' \rightarrow Y\}$. $B_{RC2} = \{W \leftarrow Y' \leftarrow e' \rightarrow Y\}$; $B_{AC} = \{W \leftarrow c' \rightarrow Y' \rightarrow Y\}$. Therefore, B_{con} will be eliminated if we assume: $(A_4$ OR $A_5)$ AND $(A_4$ OR $A_7)$ AND $(A_5$ OR $A_6)$

$\Rightarrow (A_4$ AND $A_5)$ OR $(A_4$ AND $A_6)$ OR $(A_5$ AND $A_7)$. The last condition, $(A_5$ AND $A_7)$, implies that there is no temporal dependence (that $Y' \perp\!\!\!\perp Y|W$). Given that we observed temporal dependence ($Y' \not\perp\!\!\!\perp Y|W$), or believe it is present, we can simplify the above condition to: $(A_4$ AND $A_5)$ OR $(A_4$ AND $A_6)$. □

The causal processes implied by these assumptions can be represented by the following graphs, denoted by their assumptions in subscripts as follows (\wedge denotes the AND logical operators, \vee the OR logical operator): D_{A_6} , D_{A_7} , $D_{A_4 \wedge A_5}$, and $D_{A_4 \wedge A_6}$. Since $D_{A_4 \wedge A_6}$ is a special case of (a subgraph of) D_{A_6} , I will not examine it further. For D_{A_6} and D_{A_7} , $\{Y'\}$ is a sufficient conditioning set and $\{\}$ is not a sufficient conditioning set. For $D_{A_4 \wedge A_5}$, $\{\}$ is a sufficient conditioning set and $\{Y'\}$ is not a sufficient conditioning set. These graphs are the only three maximal subgraphs of D that permit identification (depicted in Figure 4).

Observational Indistinguishability

Suppose that we are willing to assume that we are studying one of the causal processes that permit identification, but we are not sure which.

Formally, we assume $\mathcal{A}^{\mathcal{I}} = \mathcal{A}_{coll}^{\mathcal{I}} \text{ OR } \mathcal{A}_{con}^{\mathcal{I}} = \mathcal{A}_6 \text{ OR } \mathcal{A}_7 \text{ OR } (\mathcal{A}_4 \text{ AND } \mathcal{A}_5)$. Can we use the data to inform which conditioning strategy we should use? Is there a diagnostic for determining whether we should include or exclude the LDV?

While diagnostics often are available in the parametric settings (e.g., De Boef and Keele 2008), in this nonparametric setting and absent additional assumptions there are no empirical implications that can be used to determine the underlying causal process. The set of empirical predictions of causal processes $\mathcal{D}_{\mathcal{A}_6}$, $\mathcal{D}_{\mathcal{A}_7}$, and $\mathcal{D}_{\mathcal{A}_4 \wedge \mathcal{A}_5}$ are equivalent. This can be seen by listing all implied conditional independence or dependence relations for each causal process. These are listed in Online Appendix A.8 and can be summarized simply: Each observed variable will be dependent on each other variable, unconditionally and conditional on the third variable. This can also be seen through an application of (Pearl 2009b:theorem 1.2.8; Verma and Pearl 1990).

Generalizing to Process $\mathcal{D}_{\mathcal{G}}$

These identification assumptions for process \mathcal{D} have a natural generalization to process $\mathcal{D}_{\mathcal{G}}$. The above showed that the minimal nonparametric assumptions sufficient for identification in process \mathcal{D} are $\mathcal{A}^{\mathcal{I}}$. Similarly, any set of nonparametric assumptions sufficient for identification in process $\mathcal{D}_{\mathcal{G}}$ must include $\mathcal{A}^{\mathcal{I}}$. $\mathcal{A}^{\mathcal{I}}$ are necessary, but possibly not sufficient, assumptions for identification in process $\mathcal{D}_{\mathcal{G}}$. This follows from the fact that graph $\mathcal{D}_{\mathcal{G}}$, being a supergraph of graph \mathcal{D} , will contain all the backdoor paths in graph \mathcal{D} , and possibly more. Therefore, any set of sufficient identifying assumptions for process $\mathcal{D}_{\mathcal{G}}$ must include those necessary to rule out the backdoor paths in process \mathcal{D} .

Conclusion

The results of this article are summarized in the Venn diagram in Figure 5. For process \mathcal{D} , depicted by the rectangle, there are two biasing paths B_{coll} and B_{con} that make nonparametric identification impossible. If we are willing to assume $\mathcal{A}_{con}^{\mathcal{I}} = (\mathcal{A}_4 \text{ AND } \mathcal{A}_5) \text{ OR } (\mathcal{A}_4 \text{ AND } \mathcal{A}_6)$, then we can rule out B_{con} . The causal effect is then identifiable so long as we do not condition on the LDV. Alternatively, if we are willing to assume $\mathcal{A}_{coll}^{\mathcal{I}} = \mathcal{A}_6 \text{ OR } \mathcal{A}_7$ then we can rule out B_{coll} . The causal effect is then identifiable so long as we do condition on the LDV. Finally, if and only if we assume $(\mathcal{A}_4 \text{ AND } \mathcal{A}_6)$ will both $\mathcal{A}_{con}^{\mathcal{I}}$ and $\mathcal{A}_{coll}^{\mathcal{I}}$ be true, allowing us to identify the causal effect with or without the LDV.³¹

A number of insights can be drawn from these results. (1) Most fundamentally, these results affirm the truth of the proposition, for this context, that

the data alone cannot tell us which estimator to use. Given additional causal assumptions, diagnostics can be constructed.

We can also simplify our problem if we can make, or rule out, certain assumptions. (4) For example, assume \mathcal{A}_4 no reverse causation. Our problem then simplifies to determining which is a more likely source of temporal dependence: outcome autocausation or unobserved persistent causes of the outcome. If the temporal dependence comes mostly from outcome autocausation, then we should condition on LDVs; if mostly from temporally persistent causes of the outcome, then we should not condition on LDVs. The logic for this rule is described in Online Appendix A.10. (5) Alternatively, for many areas of study it is not plausible that we could measure all persistent causes of the outcome. Thus, we would reject \mathcal{A}_7 . We then see that an LDV is only justified if we think there are no unobserved causes of W and Y' (\mathcal{A}_6), whereas no LDV is preferred if we think there is no reverse causation and no outcome autocausation (\mathcal{A}_4 and \mathcal{A}_5).

We can apply these lessons for nonparametric identification to the democratic peace. In my assessment: \mathcal{A}_4 is plausible, \mathcal{A}_5 is somewhat plausible, \mathcal{A}_6 is not plausible (though maybe with W'), and \mathcal{A}_7 is not plausible. Accordingly, this recommends not conditioning on the LDV (contrary almost all work on this subject), or, if there is still sufficient variation after conditioning on lagged democracy, conditioning on both the LDV and lagged democracy. For other questions, different conditioning strategies will be appropriate. Online Appendix A.9 considers the plausibility of these assumptions for the study of the resource curse (government rents from natural resources promote authoritarianism, e.g., Haber and Menaldo, 2011) and the economic determinants of vote share for a presidential incumbent (gross domestic product promotes incumbent vote share, e.g., Lewis-Beck and Stegmaier, 2000).

In my assessment of the resource curse, \mathcal{A}_4 is plausible, \mathcal{A}_5 is not plausible, \mathcal{A}_6 is somewhat plausible (and especially with W'), and \mathcal{A}_7 is not plausible. This recommends conditioning on the LDV, probably also with lagged treatment. For the economic determinants of incumbent vote share, \mathcal{A}_4 and \mathcal{A}_5 are plausible, \mathcal{A}_6 is probably not plausible, and \mathcal{A}_7 is not plausible. This recommends not conditioning on the LDV. Of course, in each of these applications there are reasonable alternative interpretations of the plausibility of the assumptions. What is important is that scholars articulate and justify the assumptions underlying their inference.

Temporal dependence is a sign that our causal estimates may be biased. There is no purely technical fix for these biases. Any fix depends on causal assumptions, and if the wrong fix is used, biases can be introduced where there were none to begin with. To provide more informed causal inferences, scholars

should articulate and defend their causal identification assumptions. This article outlined what these causal assumptions are for nonparametric identification under temporal dependence; parametric identification assumptions are often similar, though they obviously also typically depend on assumptions about functional form. Given that social scientists rarely have confident causal knowledge about the processes we study, we should continue to evaluate the robustness of our estimators to violations of their assumptions. By doing so, we will be better able to recognize when our inferences rely on certain causal assumptions, to direct research toward investigation of those crucial assumptions, and to accurately appraise the confidence of our inferences.

A. APPENDIX FOR “NONPARAMETRIC IDENTIFICATION OF CAUSAL EFFECTS UNDER TEMPORAL DEPENDENCE.”

A.1. Peace-Year as a Transformed LDV

Consider the study of militarized conflict. The form of time most commonly employed is a kind of gap time: years since the last militarized conflict. Denote the vector of the outcome variable during the past k time periods as $LY_{i,t} = (Y_{i,t-1} \ Y_{i,t-2} \ \dots \ Y_{i,t-k})^T$. The “peace-year” is then a count of the number of consecutive years since the present during which the outcome hasn’t been a 1. Denote this $PY_{i,t} = \text{Count}(LY_{i,t})$, where $\text{Count}(\cdot)$ counts the number of consecutive 0s starting from the present. Scholars then model the probability of an event as a function of peace years: $g(PY)$. Beck, Katz and Tucker (1998) recommended cubic splines of the PY as the functional form for $g(PY)$. Carter and Signorino (2010) advocated cubic polynomials for $g(PY)$.

A.6. Non-Identification in Process \mathcal{D}

Proposition 1: *The effect of W on Y is not identifiable in Process \mathcal{D} .*

Proof. To prove that the effect of W on Y is not identifiable requires showing that backdoor adjustment is not possible, as well as showing that any other kind of non-parametric identification is not possible (such as front-door adjustment). A general means of proving this requires additional graphical tools and results. Specifically, from Theorem 4 in (Shpitser and Pearl, 2006), to show that an effect is not identifiable in graph \mathcal{D} we can show that there exists a *hedge* for the effect of W on Y . Denote the causal connection of Y' on W as $Y' \rightarrow W$. Denote the graph that removes this arrow from \mathcal{D} as $F = \mathcal{D} \setminus (Y' \rightarrow W)$. Then F is a Y' rooted C -forest (see Shpitser and Pearl,

2006) that contains W ($F \cap W \neq \emptyset$). Denote the graph that removes W from F as $F' = F \setminus W$. F' is then a Y rooted C -forest that does not contain W ($F \cap W = \emptyset$). Graphs F and F' are depicted below. This is sufficient to show that \mathcal{D} contains a hedge for the effect of W on Y .

A.7 Non-Identification in Process \mathcal{D}_G

Theorem 1: *The effect of W on Y is not nonparametrically identifiable in any causal process, \mathcal{D}_G , whose graph has Graph \mathcal{D} as a subgraph.³⁵*

Proof. \mathcal{D}_G also contains F and F' (with $F \subseteq F'$) as Y rooted C -forests, with $F \cap W \neq \emptyset$ and $F \cap W = \emptyset$. Then, by Theorem 4 and associated results in (Shpitser and Pearl, 2006), \mathcal{D}_G contains a hedge for the effect of W on Y .

Acknowledgment

For helpful comments, I am grateful to Chris Achen, Peter Aronow, Larry Bartels, Neal Beck, Scott Bennett, Henry Brady, Giacomo Chiozza, Thad Dunning, Robert Franzese, Kristian Gleditsch, Sophia Hatz, John Henderson, Greg Huber, Kosuke Imai, Xiaojun Li, Daniel Masterson, Will Moore, Betsy Ogburn, Judea Pearl, Paul Poast, Jonathan Renshon, Heiner Schulz, Ilya Shpitser, Jasjeet Sekhon, James Stimson, Laura Stoker, Michael Tomz, Guadalupe Tunon, Jiahua Yue, Baobao Zhang, Magnus Öberg, and especially David Freedman.

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

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Temporal dependence is thus a joint property of a data generating process and a particular estimation strategy.
2. These works have been cited, according to Google Scholar, respectively, 4,281, 2,023, and 1,329 times; data retrieved on February 3, 2015.
3. In this article, the term *causal effect* refers to the class of causal effects that can be identified from the marginal potential outcomes. This includes the population average treatment effect, which is the average effect of a causal factor for some target population, as well as the average treatment effect on the treated (ATT), the average treatment effect on the control (ATC), all other average treatment

effects for appropriate subgroups, as well as other population-level effects. It excludes other individual-level effects such as the effect on a specific individual, which cannot be nonparametrically identified. To simplify matters, the reader may substitute “average treatment effect” for “causal effect” throughout this article.

4. Where time is measured since the last event.
5. All participants to this conversation (Beck 2010; Carter and Signorino 2010) agree that temporal dependence should ideally be eliminated through the inclusion of appropriate theoretically meaningful independent variables. However, this task also depends on causal assumptions that have as yet not been articulated. This article shows how the tools introduced in this article can be used to know what kinds of variables one would have to measure to eliminate the bias associated with temporal dependence, for a particular causal system.
6. The choice of temporal specification has been shown in various settings to matter. For example, Beck (2003), in a conversation with Oneal and Russett (1999), reports that “it makes a great deal of difference” which temporal specification is used for estimating the effect of trade on peace.
7. This article will confine itself largely to identification via backdoor adjustment, which involves conditioning away confounding dependence. There are other sets of assumptions that could be made that would permit other identification strategies, such as by using knowledge of an exhaustive isolated mediator of the treatment effect to perform front-door adjustment (Pearl 2009b:theorem 3.3.4).
8. Others have similarly stated this maxim. For example, Morgan and Winship (2015: 355) write “... it is critical that researchers be explicit about the assumptions that they have made and be able to defend those assumptions.” My phrasing of this maxim is borrowed from Rubin (2005): “In the causal inference setting, assumptions are always needed, and it is imperative that they be explicated and justified.”
9. Implying independence of the disturbance term.
10. “Outcome autocausation” is a new term which I believe is necessary to precisely refer to the intended causal process. A close alternative term is Keele and Kelly’s (2006:193) “dynamic DGP”, which they use to refer to an effect of the outcome on future values of the outcome. I adopt outcome autocausation because it more precisely specifies the kind of dynamic process at work. The econometric literature often ambiguously uses the same terminology (“models containing lagged variables”) to describe both the regression model and the causal process generating the data, though this terminological ambiguity is less present in recent work.
11. As mentioned in note 3, this article refers to any causal effect identifiable from the marginal potential outcomes. For simplicity, however, the term “causal effect” can be read as referring to the average treatment effect.
12. Formally, a functional of the causal process, θ , is identifiable if $\theta(M) \neq \theta(M') \Rightarrow P(M) \neq P(M')$, where M and M' are specific causal models

- in the set \mathcal{M} of possible causal models, $\theta(M)$ is a functional of M , and $P(M)$ is a probability distribution on observable variables induced by M . θ can be nonparametrically identified if it can be identified without restricting \mathcal{M} to a finitely parameterized family (this definition is from Shalizi and Thomas 2011:footnote 2; see also Manski 2007).
13. There are other desirable properties of an estimator, notably the variance of its estimates; evaluating these is beyond the scope of this article.
 14. And with the node-splitting innovation of the single-world intervention graph (SWIG; Richardson and Robins 2013), any statement about potential outcomes can be expressed and analyzed using the tools of causal graphs.
 15. Expressed in potential outcomes notation: $W \rightarrow Y \Leftrightarrow \exists w_1, w_2, i, \mathbf{m}$ s.t. $Y_i(W = w_1, \mathbf{M} = \mathbf{m}) \neq Y_i(W = w_2, \mathbf{M} = \mathbf{m})$ where \mathbf{M} includes all observed variables except W and Y .
 16. I use the term *process* to refer to the underlying causal process that generates the data and reserve the term *model* to refer to an analyst's representation of a causal process.
 17. I write "typically" because two causal effects could exactly cancel. Assuming that this "knife edge" canceling never happens, for a particular causal model is called the assumption of stability (Pearl 2009b:§2.4) or faithfulness (Spirtes et al. 2001:§2.3.3). The assumption of stability is powerful because it allows the observation of independence (the failure to observe dependence) to be used to infer d -separation. Stability is probably reasonable for most causal processes, though for a dissenting view see (Freedman 1997). Stability is not needed for the results in this article.
 18. W and Y are also d -connected through the direct effect ($W \rightarrow Y$).
 19. To see this " M " structure in Figure 1, remove the arrow between W and Y and rotate the figure 90 degrees clockwise. Other terms for bias from conditioning on a collider are collider bias (Greenland 2003), endogenous selection bias (because it involves conditioning on an endogenous variable, Elwert and Winship 2014), and selection bias.
 20. Implicit to \mathcal{A}_2 , maintained through this article, is the assumption that there are no exhaustive isolated mediators of the causal effect. If there were, then nonparametric identification through "front-door adjustment" would be possible (Pearl 2009b:theorem 3.3.4).
 21. Glynn and Quinn (2013) also employ this notation.
 22. Note that we would have similar problems if we made Y' posttreatment, contrary \mathcal{A}_2 . B_{con} would consist of B_{AC} . B_{coll} would contain the above backdoor path, as well as $W \rightarrow \overline{Y'} \leftarrow e' \rightarrow Y$ and would now also block the indirect causal effect $W \rightarrow Y' \rightarrow Y$.
 23. If we observe exhaustive mediators, then it may be possible to block B_{con} and/or B_{coll} with these mediators. Alternatively, if we find an exhaustive (unconfounded)

- mechanism for the treatment effect $W \rightarrow Y$, then we could employ front-door adjustment (permitting condition 4 of Pearl's Theorem 4.3.2 to be true).
24. S is a subgraph of G if the set of nodes and directed edges in S is a subset of the set of nodes and edges in G .
 25. A terminological point. For the purposes of this article, a *condition* refers to some feature of a causal process. An assumption is an assumed condition.
 26. It may seem inconsistent that we earlier assumed \mathcal{A}_3 no unobserved causes of W and Y but are now scrutinizing the similar condition \mathcal{A}_6 of no unobserved causes of W and Y' . This article assumed \mathcal{A}_3 because it was necessary for identification, which is the focus of this article. Even if \mathcal{A}_3 is false, however, B_{coll} and B_{con} can still be present and induce bias, therefore examining them will help us understand and calibrate the bias arising from them.
 27. Note that it is not helpful to condition on exhaustive isolated mechanisms to try to make \mathcal{A}_4 or \mathcal{A}_5 true because of the same collider bias that prohibits us from conditioning on Y' . Any such mechanism would be a descendant of Y' , and if Y' is a collider between c' and e' then this mechanism would also be a collider.
 28. In addition, if W is slowly changing, as it is in the democratic peace, then there may not be sufficient information in the data remaining after conditioning on W .
 29. These assumptions are minimal with respect to structural causal models. The SWIG framework (Richardson and Robins 2013) would permit the above assumptions about causality or absence of causality to be stated solely in terms of marginal potential outcomes, rather than joint potential outcomes.
 30. This means that it will not be possible to add an edge to these graphs without losing identification.
 31. It is not possible to make both $\mathcal{A}_{con}^{\mathcal{I}}$ and $\mathcal{A}_{coll}^{\mathcal{I}}$ true by assuming $(\mathcal{A}_4 \text{ AND } \mathcal{A}_5 \text{ AND } \mathcal{A}_7)$ because the observation of temporal dependence implies $\neg(\mathcal{A}_5 \text{ AND } \mathcal{A}_7)$.

Supplementary Material

The Online Appendices are available at <http://smr.sagepub.com/supplemental> and <http://www.allandafoe.com>

References

- Acemoglu, Daron, Simon Johnson, James A. Robinson, and Pierre Yared. 2008. "Income and Democracy." *The American Economic Review* 98:808-42.
- Achen, Christopher. 2000. "Why Lagged Dependent Variables Can Suppress the Explanatory Power of Other Independent Variables." Annual Meeting of the Political Methodology Section of the American Political Science Association, UCLA. (<http://www.polmeth.wustl.edu/media/Paper/achen00.pdf>).

- Allison, Paul D. 1990. "Change Scores as Dependent Variables in Regression Analysis." *Sociological Methodology* 20:93-114.
- Beck, Nathaniel. 2003. "Modeling Dynamics in the Study of Conflict: A Comment on Oneal and Russett." Pp. 165-78 in *Globalization and Armed Conflict*, edited by Gerald Schneider, Katherine Barbieri, and Nils Petter Gleditsch. Oxford, UK: Rowman and Littlefield
- Beck, Nathaniel. 2010. "Time Is Not a Theoretical Variable." *Political Analysis* 18: 293-94.
- Beck, Nathaniel and Jonathan N. Katz. 1995. "What To Do (and Not To Do) with Time-series Cross-section Data." *American Political Science Review* 89:634-47.
- Beck, Nathaniel and Jonathan N. Katz. 1996. "Nuisance vs. Substance: Specifying and Estimating Time-series-cross-section Models." *Political Analysis* 6:1-36.
- Beck, Nathaniel and Jonathan N. Katz. 2011. "Modeling Dynamics in Time-series-cross-section Political Economy Data." *Annual Review of Political Science* 14:331-52.
- Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. "Taking Time Seriously: Time-series-cross-section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42:1260-88.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2002. "How Much Should We Trust Differences-in-differences Estimates?" *Quarterly Journal of Economics* 119:249-75.
- Blackwell, Matthew. 2013. "A Framework for Dynamic Causal Inference in Political Science." *American Journal of Political Science* 57:504-19.
- Blackwell, Matthew. 2014. "A Selection Bias Approach to Sensitivity Analysis for Causal Effects." *Political Analysis* 22:169-82.
- Blackwell, Matthew and Adam Glynn. 2014. "How to Make Causal Inferences with Time-series Cross-sectional Data." Working Paper, Harvard University, Cambridge, MA.
- Box-Steffensmeier, Jan and Bradford Jones. 2004. *Event History Modeling*. Cambridge, UK: Cambridge University Press.
- Box-Steffensmeier, Janet M., Suzanna De Boef, and Kyle A. Joyce. 2007. "Event Dependence and Heterogeneity in Duration Models: The Conditional Frailty Model." *Political Analysis* 15:237-56.
- Carter, David B. and Curtis S. Signorino. 2010. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis* 18:271-92.
- Cochrane, Donald and Guy H. Orcutt. 1949. "Application of Least Squares Regression to Relationships Containing Auto-correlated Error Terms." *Journal of the American Statistical Association* 44:32-61.
- Collier, David, Jasjeet S. Sekhon, and Philip B. Stark. 2010. "Editors' Introduction: Inference and Shoe Leather." Pp. 12-15, in *Statistical Models and Causal Inference: A Dialogue with the Social Sciences*, edited by David Collier,

- Jasjeet S. Sekhon, and Philip B. Stark. Cambridge, UK: Cambridge University Press.
- Dafoe, Allan. 2011. "Statistical Critiques of the Democratic Peace: Caveat Emptor." *American Journal of Political Science* 55:247-62.
- Dafoe, Allan, John R. Oneal, and Bruce Russett. 2013. "The Democratic Peace: Weighing the Evidence and Cautious Inference." *International Studies Quarterly* 57:201-14.
- De Boef, Suzanna and Luke Keele. 2008. "Taking Time Seriously." *American Journal of Political Science* 52:184-200.
- Ding, Peng and Luke 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.
- Dunning, Thad. 2008. "Model Specification in Instrumental-variables Regression." *Political Analysis* 16:290-302.
- Elwert, Felix and Christopher Winship. 2014. "Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable." *Annual Review of Sociology* 40:31-53.
- Franzese, Robert J., Jr., and Jude C. Hays. 2007. "Spatial Econometric Models of Cross-sectional Interdependence in Political Science Panel and Time-series-cross-section Data." *Political Analysis* 15:140-64.
- Freedman, David A. 1997. "From Association to Causation via Regression." Pp. 113-61, in *Causality in Crisis?*, edited by V. R. McKim and S. P. Turner. Notre Dame, IN: University of Notre Dame Press.
- Freedman, David A. 2005. *Statistical Models: Theory and Practice*. New York: Cambridge University Press.
- Freedman, David A. 2006. "On the So-called "Huber Sandwich Estimator" and "Robust Standard Errors."" *American Statistician* 60:299-302.
- Gelman, Andrew and Jennifer Hill. 2006. *Data Analysis Using Regression and Hierarchical/Multilevel Models*. New York: Cambridge University Press.
- Glynn, Adam N. and John Gerring. 2013. "Strategies of Research Design with Confounding: A Graphical Description." Manuscript for the 2011 IQMR at Syracuse University. (<http://blogs.bu.edu/jgerring/files/2013/06/structureSMR.pdf>).
- Glynn, Adam and Konstantin Kashin. 2013. "Front-door Versus Back-door Adjustment with Unmeasured Confounding: Bias Formulas for Front-door and Hybrid Adjustments." Manuscript for the 2013 Meeting of the Midwest Political Science Association. (<http://scholar.harvard.edu/files/aglynn/files/glynnkashin-frontdoor.pdf>).
- Glynn, Adam N. and Kevin M. Quinn. 2013. "Structural Causal Models and the Specification of Time-series-cross-section Models." Working paper. (<http://scholar.harvard.edu/aglynn/files/tscs-scm.pdf>).
- Greene, William. 2012. *Econometric Analysis*. 7th ed. Upper Saddle River, NJ: Prentice Hall.

- Greenland, Sander. 2003. "Quantifying Biases in Causal Models: Classical Confounding vs Collider-Stratification Bias." *Epidemiology* 14:300-306.
- Haber, Stephen and Victor Menaldo. 2011. "Do Natural Resources Fuel Authoritarianism? A Reappraisal of the Resource Curse." *American Political Science Review* 105:1-26.
- Hamilton, James Douglas. 1994. *Time Series Analysis*. Vol. 2. Princeton, NJ: Princeton University Press.
- Hayes, Jarrod. 2012. "The Democratic Peace and the New Evolution of an Old Idea." *The European Journal of International Relations* 18:767-91.
- Hernan, Miguel A. and James M. Robins. 2015. *Causal Inference*. Chapman & Hall/CRC. Retrieved November 9, 2015. (<http://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>).
- Keele, Luke. 2010. "Proportionally Difficult: Testing for Nonproportional Hazards in Cox Models." *Political Analysis* 18:189-205.
- Keele, Luke and Nathan J. Kelly. 2006. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analysis* 14:186-205.
- King, Gary. 1998. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Ann Arbor: Techniques in Political Analysis, University of Michigan Press.
- King, Gary and Margaret Roberts. 2015. "How Robust Standard Errors Expose Methodological Problems They Do Not Fix." *Political Analysis* 23:159-79.
- Lewis-Beck, Michael S. and Mary Stegmaier. 2000. "Economic Determinants of Electoral Outcomes." *Annual Review of Political Science* 3:183-219.
- Manski, Charles. 2007. *Identification for Prediction and Decision*. Cambridge, MA: Harvard University Press.
- Morgan, Stephen L. and Christopher Winship. 2015. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. 2nd ed. Cambridge, UK: Cambridge University Press.
- Mousseau, Michael and Yuhang Shi. 1999. "A Test for Reverse Causality in the Democratic Peace Relationship." *Journal of Peace Research* 36:639-63.
- Newey, Whitney K. and Kenneth D. West. 1987. "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55:703-8.
- Oneal, John R. and Bruce Russett. 1999. "The Kantian Peace: The Pacific Benefits of Democracy, Interdependence, and International Organizations, 1885-1992." *World Politics* 52:1-37.
- Pearl, Judea. 2009a. "Causal Inference in Statistics: An Overview." *Statistics Surveys* 3:96-146.

- Pearl, Judea. 2009b. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge, UK: Cambridge University Press.
- Pearl, Judea and James Robins. 1995. "Probabilistic Evaluation of Sequential Plans from Causal Models with Hidden Variables." Pp. 444-53 in *Uncertainty in Artificial Intelligence 11*, edited by P. Besnard and S. Hanks. San Francisco, CA: Morgan Kaufmann.
- Prais, Sigbert J. and Christopher B. Winsten. 1954. "Trend Estimators and Serial Correlation." Cowles Commission Discussion Paper No. 383, Chicago..
- Richardson, Thomas S. and James M. Robins. 2013. "Single World Intervention Graphs (SWIGs): A Unification of the Counterfactual and Graphical Approaches to Causality." Working Paper Number 128, Center for Statistics and the Social Sciences, University of Washington, Seattle. Retrieved November 20, 2015 (<https://www.csss.washington.edu/Papers/wp128.pdf>).
- Robins, James M., Miguel Angel Hernan, and Babette Brumback. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." *Epidemiology* 11:550-60.
- Rosenbaum, Paul R. 2002. *Observational Studies*. 2nd ed. New York: Springer.
- Rosenbaum, Paul R. and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70:41-55.
- Rubin, Donald B. 2005. "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions." *Journal of the American Statistical Association* 100:322-31.
- Russett, Bruce. 1993. *Grasping the Democratic Peace: Principles for a Post-cold War World*. Princeton, NJ: Princeton University Press.
- Sekhon, Jasjeet S. 2009. "Opiates for the Matches: Matching Methods for Causal Inference." *Annual Review of Political Science* 12:487-508.
- Shalizi, Cosma Rohilla. In Press. *Advanced Data Analysis from an Elementary Point of View*. Cambridge University Press.
- Shalizi, Cosma Rohilla and Andrew C. Thomas. 2011. "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies." *Sociological Methods & Research* 40:211-39.
- Shpitser, Ilya and Judea Pearl. 2006. "Identification of Joint Interventional Distributions in Recursive Semi-Markovian Causal Models." P. 1219 in *Proceedings of the National Conference on Artificial Intelligence*, edited by Ford, Kenneth, Ken Forbus, Pat Hayes, Janet Kolodner, George Luger, Robert Morris, Alain Rappaport, and Brian Williams. Vol. 21. Menlo Park, CA; Cambridge, MA; London, UK; AAAI Press; MIT Press.
- Spirtes, Peter, Clark N. Glymour, and Richard Scheines. 2001. *Causation, Prediction, and Search*. Vol. 81 Cambridge, MA: MIT Press.
- Stone, Richard 1993. "The Assumptions on Which Causal Inferences Rest." *Journal of the Royal Statistical Society. Series B (Methodological)* 5:455-66. Retrieved November 9, 2015 (<http://www.jstor.org/stable/10.2307/2346206>).

- Thompson, William R. 1996. "Democracy and Peace: Putting the Cart Before the Horse." *International Organization* 50:141-74.
- Thoemmes, Felix. 2015. "M-bias, Butterfly Bias, and Butterfly Bias with Correlated Causes – A Comment on Ding and Miratrix." *Journal of Causal Inference* 3(2): 253-258.
- Verma, Thomas and Judea Pearl. 1990. "Equivalence and Synthesis of Causal Models." Pp. 220-27 in *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*, edited by Bonissone, Piero, Max Henrion, Laveen Kanal, and John Lemmer. Cambridge, MA: AUAI.
- Wilson, Sven E. and Daniel M. Butler. 2007. "A Lot More To Do: The Sensitivity of Time-series Cross-section Analyses to Simple Alternative Specifications." *Political Analysis* 15:101-23.
- Wooldridge, Jeffrey M. 2009. *Introductory Econometrics: A Modern Approach*. Manson, OH: Cengage Learning.
- Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

Author Biography

Allan Dafoe is an assistant professor of political science at Yale University. His research investigates the causes of war and social science methodology. His research can be found at www.allandafoe.com.