The Perils of the All Cause Model∗

Luke Keele† Felix Elwert‡ Randolph T. Stevenson§

July 22, 2015

Abstract

One of the most common identification strategies in political science is selection on observables. Under this strategy, analysts assume that they have observed enough covariates to make treatment status as-if random. Adjustments are then made for observed confounders through statistical methods such as regression or matching. Under adjustment methods such as matching or inverse probability weighting, coefficients for control variables are treated as nuisance parameters and are not directly estimated. This is in direct contrast to regression approaches where estimated parameters are observed for all covariates. Analysts often find it tempting to give a causal interpretation to all the parameters in such regression models, which is not possible under some methods like matching. In this paper, we ask when we can justify interpreting two or more coefficients in a regression model as causal parameters. Using Directed Acyclic Graphs, we show how even if some effects are identified in a regression model, many estimated parameters do not represent causal effects.

∗We would like to thank Judea Pearl and Josh Kertzer for helpful comments.
†Associate Professor, Department of Political Science, 211 Pond Lab, Penn State University, University Park, PA 19130 Email: ljk20@psu.edu.
‡Vilas Associate Professor of Sociology, Department of Sociology, University of Wisconsin-Madison, 4426 Sewell Social Sciences, Madison, WI 53706, Email: felwert@ssc.wisc.edu
§Professor, Department of Political Science, Rice University, P.O. Box 1892, MS-24, Houston, TX 77251, Email: stevenso@rice.edu
1 Introduction

Statistical models are pervasive in political science research. Sometimes these models are used to describe empirical phenomena, sometimes to predict political outcomes, and often to draw inferences about causal relationships. Political scientists’ focus on causal inference stems largely from the important role such inferences play in the process of developing, evaluating, and modifying causal theories. For example, the authors of a leading undergraduate textbook on political science research methods exhort students to remember:

...[the] first rule of the road to scientific knowledge about politics is “Make your theory causal.” If your answer to the question “Is your theory causal?” is anything other than “yes,” you need to go back to the drawing board until the answer is an emphatic “yes” (Kellstedt and Whitten 2013).

Further, a classic text on the “Craft of Political Research” reminds us that:

...the ultimate purpose of theories is to give us levers on reality, some basis for choosing how to act. If A and B coincide but A does not cause B, changing A will not change B. Coincidence without cause gives you no lever (Shively 1998).

Given the centrality of causal theory and inference to the discipline, it is no surprise that political scientists have long been concerned with understanding the conditions under which one can make causal inferences from data. Indeed, in recent years, a new generation of political methodologists have taken up the rigorous study of causal inference. Various labeled the “potential outcomes perspective,” the “causal modeling perspective,” or “design-based inference” this work has championed the Neyman-Rubin model of causality, promoted design-based inference, and advocated for greater use of randomized experiments. While this work has devoted much emphasis to research designs like randomized and natural experiments (Gerber, Green and Kaplan 2004, Dunning 2009, Sekhon 2009, Keele and Minozzi 2012), a considerable amount of work has also
explored how we can make causal inferences with non-experimental data, given that experimental approaches are simply not feasible in many areas of political science.

Despite all this activity, however, many of the methodological questions raised in this literature have so far failed to impact the practice of political science research outside of the few domains in which randomized or natural experiments are feasible. In this essay, we attempt to remedy that failure for one question that is of practical significance to almost all empirical researchers in political science: what causal interpretation can we give to coefficients in multivariate statistical models? Our concern with this question stems from a curious disconnect between the advice one finds in the statistical literature on causality and the way most political scientists specify, estimate, and interpret empirical models. Specifically, almost all expositions of modern theories of causality focus on the estimation of a single causal effect - and specify an empirical model to identify that single effect. Further, to the extent that other variables must be included in an empirical model to identify that effect, the associations between these variables and the outcome variable are either never estimated (e.g., in matching) or are unreported. Certainly, they are given no causal interpretation.\footnote{This is not to say that the literature on causal modeling rules out causal interpretations of multiple estimated parameters in the same model. As we shall see below, this is sometimes possible and a useful thing to do. However, it is fair to say that most of this work recommends specifying empirical models to identify one causal effect of interest.}

In contrast, the usual practice in political science is to specify a statistical model with multiple independent variables and then, in many cases, to interpret two or more of those variables causally. Sometimes this practice reflects the underlying goals of the study, which may pose one or more rival hypotheses and seek to use the estimates of different causal effects (from the same statistical model) to adjudicate between them.\footnote{Here, we group studies that pose a “critical test” between two or more competing hypotheses with those that simply seek to compare the size of different effects in a sort of “race of variables.” The point is that in many cases, the researcher includes several variables of interest in the same model precisely because the design requires a comparison of their relative causal effects.} Likewise researchers almost always include a set of “control” variables in the statistical model that are not of primary interest, but that are thought necessary to identify the causal effects of interest. While nothing about the design requires that the researcher give estimated associations for these control variables a causal interpretation, a review of recent
empirical works reveals that such interpretations are frequently forthcoming. A careful reading of many such interpretations makes it clear that researchers often feel compelled to provide a causal interpretation of the estimated effects of a control variable precisely because, in specifying the model in the first place, they expected the variable to impact the outcome in a specific, known way (perhaps because of previous empirical work). If these expectations are then confirmed, the researcher takes this result as both additional empirical evidence to add to the literature on the causal effect of that control variable on the outcome and as a type of positive specification check for the estimated model. If instead the estimated effect of the control variable is unexpected, this is seen as an anomaly to explain in the discussion section or, if sufficiently egregious, as a sign of model misspecification. The questions we explore here concern the conditions under which such causal interpretations of control variables, or of multiple variables of interest (as in the test of rival hypotheses set up) are justified. To do so, we must first understand the general conditions under which causal effects are identified in most empirical models and so begin with a discussion of a general method for thinking about causal identification—the directed acyclic graph (DAG). Using these tools, we can plainly characterize what is required to give causal interpretations to multiple estimated parameters in a model (and so answer more specific questions about the causal status of estimated effects for control variables, or how we can test rival hypotheses). Before introducing DAGs, however, it will be useful to briefly state what we mean by a causal effect, contrast that with the measures of association that we commonly get out of our statistical models, and explain how an identification strategy is needed to interpret a given estimated regression coefficient as a causal effect.

1.1 Associations, Identification, and Causal Effects

We say that there is an association between two variables D and Y if the distribution of Y varies across levels of D. We say there is a causal relationship between D and Y in a population if and only if there is at least one unit in that population for which intervening in the world to change D will change Y (Pearl 2009). Intuition tells us that associations are the result of causal

---

3If this does not hold, then D and Y are said to be independent.
relationships. That is, if D causes Y this will produce an association between D and Y. However, the challenge of moving in the other direction (from association to causation) is that there are forces other than causality, such as confounding and selection, which can induce an association between D and Y. Moreover, association is non-directional. An association might imply that Y causes D instead of D causing Y.

Therefore, an observed association between D and Y contains some unknown mix of causal and noncausal (spurious) components. Identifying a causal effect from an observed association requires an identification strategy: a research design that allows one to conclude that an observed association reflects causation. Thus, any research that makes a causal claim must adopt, at least implicitly, some identification strategy. That said, not all identification strategies are equally convincing. While some research designs are easily justified based on simple logic or previous empirical research, others are not. One of the principal responsibilities of researchers making causal claims is to articulate and justify the identification strategy that supports those claims.

There are a number of identification strategies that have been extensively studied (see Keele (2014) for a detailed overview). These include randomized experimental designs, regression discontinuity designs, and encouragement designs (i.e., instrumental variables). However, the most commonly used identification strategy in political science—and the one with which we will be concerned here—is selection on observables, commonly referred to as “no omitted variables,” “correct specification,” “controlling for confounders,” or “conditional ignorability.” A set of phrases that refer to the possibility that we can make treatment status as-if random if we condition on a set of observed covariates.

2 The DAG Approach to Causal Modeling

The key to identifying a causal effect of interest using the selection on observables strategy is to control for the right set of measured variables. Over the years, there has been a lot of

---

4This view of identification is one found in Pearl (2009) and is somewhat different from the view of identification found in econometrics. However, the two views can be reconciled, and we think this view is a useful way to think about identification.
confusion about what exactly this set of variables should be. With some political scientists advocating the need to control for all causes of the outcome (Berry and Feldman 1985) and others suggesting empirical models be limited to a small set of controls (Achen 2002b). Indeed, there have even been whole issues of political science journals devoted to debating this question (Kadera and Mitchell 2005). Recently, however, the new causal inference literature has brought considerable clarity to the question of which control variables must be included in an empirical model to identify a given causal effect. Importantly, this seldom includes all the causes of the outcome, but neither is it possible to apply some arbitrary cut off on the number of variables needed. Instead, it is now clear that one need control only for those variables that block all open “back door” paths between the outcome and the causal variable of interest (Pearl 1995); or, in the language of potential outcomes, render treatment assignment “ignorable” (Rosenbaum and Rubin 1983). Disconcertingly, it is possible to include some variables as controls that, while they are themselves causes of the outcome, actually compromise the identification of causal effects of interest.

Consequently, the specific set of control variables needed for causal identification will be different in any given empirical project. They may be different for different causal effects of interest in the same project, and will generally be different for different plausible assumptions about the underlying causal structure producing outcomes. It is therefore essential that empirical researchers make clear, in each and every application in which causal identification is through selection on observables, which sets of control variables (if any) are necessary to identify each causal effect of interest.

While this may seem like a daunting task, developments in one important vein of the causal inference literature have made the task much simpler. Specifically, Pearl (2009, 1995) has shown that one can use graphical methods to clarify the causal assumptions one is making when one estimates a given statistical model (and interprets one or more of its parameters causally) and to efficiently identify the minimal set of variables that it is necessary to measure (and control for) to identify any given causal effect. These graphical models are called directed acyclic graphs or
DAGs are graphical representations of nonparametric structural equation models (Pearl 2009) and are often useful for reasoning about causal structure, since they allow us to formalize identification assumptions for causal effects. From a given DAG, a researcher can derive which causal effects are identified for a given set of control variables and easily identify the theoretical assumptions that support that identification. In the sections that follow, we explain how political scientists can use DAGs to understand the assumptions needed to identify causal effects under a selection on observables identification strategy and give several illustrations of the method.

2.1 DAGs: From Theory to Diagrams

While there are many references available on the use of DAGs, there is little guidance on how to translate a theory into a DAG (Elwert 2013). Here, we feel that it is critical that the reader understand at least one way this might happen. To illustrate this process, we use the theory in Rogowski (2014), where the interest is in the causal effect of ideological divergence between political elites on political participation. More specifically he studies the effect of ideological divergence between U.S. Senate and House candidates on turnout in elections. He also explores multiple causal mechanisms to explain why ideological divergence has a causal effect on turnout. We translate his theory into the DAG in Figure 1, where ideological divergence is a cause of turnout.

DAGs are graphs that consist of labeled “nodes” (circles) connected by “edges” (arrows) that encode the qualitative causal structure of the data generating process. A node represents a variable or set of variables. We represent measured variables (on which we can condition) with a solid circle and unobserved variables with an open circle. For example, in Figure 1 ideological divergence between the candidates A is a measured variable, whereas the extent of individual alienation towards the candidates D is unmeasured. When one variable causes another, we draw an arrow between them. These arrows order variables in time. Since a cause must precede its effect, DAGs do not contain cycles. All variables directly or indirectly caused by a given variable

---

5DAGs do not allow “mutual causation,” but recognize this as a kind of shorthand for a more elaborate temporal sequencing of different effects that should be diagrammed out. See the appendix for an example.
Figure 1: One possible DAG for theory in [Rogowski (2014)]. The nodes in the DAG represent the following variables: A — ideological divergence, B — perceived ideological divergence, C — importance of voting, D — alienation from candidates, E — political sophistication, Y — turnout, F — Demographics, U — Unmeasured causes.

are called its descendants. Any two nodes may be connected by one or more “paths”, which just means that one get from one of the two nodes to the other by following a sequence of arrows. If, in doing so, the arrows along the path always point in the same direction the path is called a “causal path.”

The first step in using a DAG to justify an identification strategy is to translate one’s theory into a DAG. This process begins with a substantive description of the known variables (measured or unmeasured) in the theory and the causal relationships between them (i.e., which nodes should be included in the diagram and which should have arrows between them?). In our reading of [Rogowski (2014)] the primary covariates are as follows: A — ideological divergence, B — perceived ideological divergence, C — importance of voting, D — alienation from candidates, E — political sophistication, Y — turnout, F — Demographics, and U — unmeasured causes, which are a set of variables beyond demographics that cause both demographics and turnout.

Since the DAG represents the causal theory nonparametrically i.e. qualitatively, this translation need not include assumptions about functional forms (exactly how one causal variable impacts

---

6It is important to understand that DAGs are not simply heuristic diagrams of the kind many scholars use to illustrate their verbal theories, but are based on rigorous mathematical theory that allows us to derive causal identification results from these diagrams.
another) or interactions between different variables. Instead one simply indicates, for all relevant variables, which ones have any causal impact on which others—no matter how exactly that causal impact is felt. For example, the DAG in Figure 1 would be the same if the size of the causal effects of political sophistication (E) and ideological divergence (A) on turnout (Y) were dependent on each other (i.e., interactive) or if they were not.\footnote{More formally, a DAG is a visual representation of a system of non-parametric structural equations. The DAG only tells us which variables enter into which functions, not the particular way in which they do so. For example, the non-parametric structural equations corresponding to the DAG in Figure 1 are: $Y = f(C, D, U)$, $C = f(B, E)$, $D = f(B, E)$, $B = f(A, E)$, $E = f(F)$, $F = f(U)$, where $f$ in each case is unspecified and potentially different for each equation. For example, the equation for $Y$ tells us that the value of $Y$ can take a distinct, and unconstrained, value for each combination of the values of $C$, $D$, $F$, and $U$. Thus, the notation allows for any of the possible interactions or functions that would substantiate that mapping. All the diagram (and corresponding equations) tells us are the variables that are relevant to determining the value of $Y$. This is powerful since, because the diagram can also tell us about identification, it does so without needing to specify these details.} This also emphasizes the important point that the arrows in the diagram do not indicate the direction or size of effects—or that these effects may vary in direction or size for different units in the population, as in an interactive specification.\footnote{Recall that a causal relationship exists between two variables, D and Y, if there is at least one unit in the population for which intervening to change D will change Y. With an infinite amount of data, which this kind of identification analysis assumes, even the effect for this single unit could be empirically identified (as could the lack of effects for all other units).}

Indeed, the construction of the DAG, and the identification analysis that follows from it, precedes specific questions of how to practically estimate the causal effects that appear in the DAG. The DAG is about identification rather than estimation and so asks: with an infinite amount of data what causal relationships can we identify? This also makes it clear why the DAG is nonparametric. With an infinite amount of data, we need never rely on parametric assumptions due to data limitations. To put it another way, with an infinite amount of data, there will always be a nonparametric estimator that can describe the response of Y to any and all combinations of the values of its directly antecedent causal variables no matter their specific functional relationships to it or each other (including, for example, complex interactions).

Next, we provide a brief summary of the causal identification concepts in DAGs. We cover these same concepts in greater detail in the next section of the paper. All causal effects in a DAG are transmitted along one or more paths. For example in Figure 1, causal effects flow from A to Y along the direct path $A \rightarrow Y$ but also indirectly along paths such as $A \rightarrow B \rightarrow C \rightarrow Y$. The total
causal effect of A on Y is comprised of all the causal paths from A to Y. However, not all paths carry causal effects, some carry spurious association. All sources of spurious association between two nodes, D and Y, is comprised of all the non-causal paths between D and Y. In Figure 1 a non-causal path between E and Y is $E \rightarrow B \leftarrow A \rightarrow Y$. The concept of D-separation determines which paths carry causal effects and which carry spurious association. This determination often depends on whether a path is said to be “blocked.” A path is blocked if it contains a node on which the analysts has decided to condition on. For example if we control for F in a regression model that would block the noncausal path from E to Y through U. A path can also be blocked if the path contains a collider, and neither the collider nor any of its descendants are conditioned on. In Figure 1 B is a collider since there are two arrows into this node. Thus B blocks some paths in this DAG. This leads to the adjustment criterion which stipulates that identification of a causal effect can occur when one blocks all non-causal paths between T and Y, and leaves all causal paths between T and Y open.

A DAG displays three types of causal effects: direct effects, total effects, and indirect effects. It is important to understand the differences between these effects because the assumptions for identification differ depending on the type of effect of interest. Consider two nodes, labeled D and Y in Figure 2. The direct effect of D on Y is indicated by a single arrow from D to Y and can be considered a summary of all of those causal paths between D and Y for which the details of the casual mechanism are unknown (or at least unarticulated). As we noted above, the total effect of D on Y is indicated by all the causal paths that begin at D and end at Y (perhaps passing through other variables).

There are other, more fine, distinctions one can make – like natural direct effects and controlled direct effects, but one need not appreciate these subtleties to understand the points we want to make.
through other variables). Finally, indirect effects of D on Y are indicated by those directed paths that flow through a third variable usually referred to as a mediator.

In Figure 1, the direct effect of ideological divergence A on turnout Y is indicated by the single arrow between them. The total effect of ideological divergence A on turnout Y, however, comprises that causal path plus the two additional indirect effects that start at ideological divergence A, flow through perceived ideological divergence B and then, through perceived alienation D and perceived importance C, respectively, to turnout Y.  

There are certainly situations in which social scientists are interested in estimating all three kinds of effects. That said, direct effects may be of lesser interest, since they are essentially a theoretical “residual” category. Consider, for example, the direct effect of ideological divergence A on turnout Y in Figure 1 and contrast it with the indirect effects that both run through perceptions of ideological divergence B. These two indirect effects capture two proposed theoretical mechanisms, while the direct effect is essentially any causal connection between ideological divergence A and turnout Y for which the author has no other explanation. While there are, presumably, some actual mediators along this direct pathway, these are unknown and so go unspecified in the DAG.

Often, researchers are interested in indirect effects and total effects. Thus, in a typical application, a researcher will lay out, in the theoretical discussion, one or more, more or less elaborate mechanisms connecting D to Y (though a series of mediators). Then, based on these explanations, state one or more hypotheses about whether, under each hypothesis, D should have a positive/negative causal effect on Y. Rogoski’s (2014) paper (from which we developed

---

10In contrast, the path from ideological divergence A to turnout Y that passes through both perceived divergence B and political sophistication E is not part of the total effect of ideological divergence on turnout because it is not a causal path—part of it flows in the “wrong” direction.

11Which is not to say that they are of no interest. One might, for example, want to estimate whether any causal effect of D on Y remains after having accounted for some known mechanism.

12Alternatively, these mediators might be known but occur at a level of explanation that has is sufficiently detailed that disciplinary expectations do not require further theoretical elaboration of such mechanisms (as often happens for example, when an a political science explanation becomes sufficiently detailed that the only remaining mediators along a path would involve biological concepts). At this point, most researchers, while not denying the existence of such causal connections, are happy to invoke a “black box” in the theory – in other words, a direct effect. See Morgan and Winship (2014) for a discussion of the importance of “levels of explanation” and disciplinary expectations in causal explanations.
Figure 1) is a case in point. In that paper, he reviews two explanations for how the extent of ideological divergence A among candidates in a given race might impact turnout Y in that race. The first suggests that greater ideological divergence among candidates, if perceived by a potential voter, is a signal that the stakes (or importance) of the race (in terms of policy consequences) are likely to be high. Thus divergence should, in this view, cause higher turnout through the importance mechanism. Likewise, greater divergence, again if perceived by a potential voter, could produce a negative reaction in which the potential voter (assumed to be ideologically moderate) is more alienated from both candidates (either because the candidate’s positions are farther from this potential voter’s own policy positions or perhaps because of a negative reaction to perceived divergence in general). Clearly, then these two mechanisms imply a differentially signed relationship between ideological divergence and turnout.

Rogowski would presumably like to estimate the indirect effects of each mechanism and compare them — that is, to isolate the contribution to the total effect of each mechanism separately. However, to do so would require that he be able to measure individual feelings of alienation, and the importance of the race and then to manipulate those independently of divergence. Given this, Rogowski instead does what the majority of political scientists facing the same situation would do given that he suspects there are multiple unmeasured mechanisms connecting a causal variable of interest (in this case A) to an outcome Y: he focuses on estimating the total effect of that variable (ideological divergence A) on the outcome (turnout Y) through whatever mechanisms (as well as through any direct effect) and to then interpret that result as evidence for one mechanism over the other. If the total effect is positive, this is taken as evidence for the dominance of the importance mechanism, and if it is negative, for the dominance of the alienation mechanism.\footnote{This interpretation depends on whether there is a significant direct effect that does operate through either of Rogowsi’s proposed mechanisms (and, as with all direct effects, is interpreted as a summary of all other — ostensibly unknown or unarticulated causal paths between D and Y). If so, then an estimated total effect of a given sign is clearly not unambiguous evidence for one mechanism over the other (or that the effect of one mechanism is larger than the other), since much of that total effect could be due to the direct effect, which could be of either sign. Further, nothing in this set up justifies the conclusion, based, for example, on finding a negative total effect that the mechanism for which a positive effect is predicted is not at all operative. At best (absent any direct effect) one might conclude that the one mechanism dominates the other.} In the rest of our paper then, we will focus on the conditions under which...
which total effects can be identified using a selection on observables strategy, though we will
note when direct effects are identified and possibly mistaken for total effects. One critical aspect
of drawing a DAG is that the arrows that are not drawn between variables are as important, if not
more important, than the arrows that are drawn. If there is no arrow between two variables that
means we are willing to assume that there is no causal relationship between two nodes, which
corresponds to an exclusion restriction. Importantly, this includes situations in which one assumes
(by not including them in the model) that there is no causal relationship between variables that are
in the model and those that are left out of the model. For example in Figure 1 we drew the DAG
such that ideological divergence A is not affected by any factors. This is an exclusion restriction
that is encoded in this DAG. Exclusion restrictions are what let us identify causal effects. In a
DAG with no exclusion restrictions, no causal effect can ever be identified. Of course, it is not
the case that an exclusion restriction between two variables implies no association between them
(since they may well share a common cause that creates such an association). The restriction
only implies that neither variable causes the other.

Identification depends on exclusion restrictions. Indeed, in the special case of randomized
experiments, the exclusion restrictions are justified by the design, i.e., that there is no causal
relationship between any measured or unmeasured variable, the treatment, and the outcome.
Thus, experiments, in which the treatment is properly randomized, are sufficient to make most
of that theoretical structure of the total effect ignorable (and so unnecessary to write down to
assess identification). For other kinds of designs, however, it will almost always be necessary to
include at least part of the assumed theoretical structure in the DAG to access the identification
of a given causal effect.

It will often be the case, of course, that some parts of the DAG, which are necessary to access
identification of an effect of interest, are not clearly specified by relevant theory (e.g., should
one impose an exclusion restriction between two variables? Do they share one or more common
unmeasured causes? Does one of the variables cause the other? Should a given variable even
appear in the DAG at all?). In these cases, the relevant part of the DAG must be specified either
from previous empirical work, informed speculation, or by assumption. What the DAG will make
very clear, however, is the extent to which the identification of key causal inferences critically
depends on these theoretically underspecified parts of the DAG. If so, then these inferences should
be correspondingly tentative. Indeed, Pearl (2009) envisions a process whereby researchers in a
given substantive area develop a set of widely accepted DAGs that instantiate the key theoretical
debates in the literature and so give researchers both a common language to debate them and
an immediate link between these theoretical questions and their empirical implications. That is,
DAGs can serve as a means of identifying which potentially debatable causal arrows and exclusion
restrictions drive causal identification in the literature.

In the examples that follow, we will highlight the different ways in which identification de-
pends on the theoretical structure instantiated in a given DAG (including assumptions about the
existence of measured and unmeasured covariates, their causal connections, and exclusion restric-
tions). It is important to remember, however, that use of a DAG doesn’t imply effects are actually
identified. Identification results assume the DAG is correct, which we don’t actually know. The
DAG just makes the structure of the theory and its implications for causal identification clear.

Once the DAG is completed, we can derive which of the specified causal relationships are non-
parametrically identified, which measured variables, or sets of variables, must be controlled for to
achieve identification, and which pairs of measured variables should be conditionally independent
(giving us potential tests of the causal structure). Next, we provide a brief outline of graphical
rules for identification through conditioning on observed covariates.

2.2 Graphical Identification

Deriving identification results from a DAG depends on the concept of d-separation (Verma and
Pearl 2013; Pearl 1988). Paths in a DAG may transmit either an association or a causal effect.

The concept of d-separation provides a set of rules for whether a path transmits an association

While we present the rule intuitively using the language of backdoor paths that can be readily seen in a DAG, it is important to keep in mind that the justification for this intuition is provided in the various mathematical proofs found in (Verma and Pearl 2013; Pearl 1988). Indeed, Galles and Pearl (1998) and Richardson and Robins (2013) show that a theorem under DAGs is a theorem under potential outcomes as the two frameworks are equivalent mathematically.
or causal effect.

Following Elwert’s (2013) exposition, the d-separation rule posits that all empirical associations between two variables are transmitted along paths that connect them—though not all paths between variables transmit association. Again, paths are sequences of edges, which can be traversed with or against the direction of their arrows, as long as no node is touched more than once on a given path. Intuitively, d-separation tell us that if one can connect an outcome variable Y and the treatment D along some path in a DAG, then, in the absence of a "collider variable" along the path, or if there is no attempt to block the path via a conditioning variable, there will be an association between Y and D that is transmitted along that path. Below, we briefly explain collider variables and then discuss the identification strategy we focus on in this paper: conditioning on observables to block all “back-door” paths.

All causal effects are transmitted along so-called causal paths, in which the arrows point strictly away from the treatment and toward the outcome. All other paths between treatment and outcome, i.e., paths in which at least one edge is traced against the direction of the arrow, are called non-causal paths and represent potential sources of spurious, i.e. non-causal, association. These are often referred to as “backdoor paths” which are are non-causal paths that start with an arrow into the treatment. All non-causal paths transmit spurious association, i.e. are “open” unless they contain a collider variable that hasnâĂŹt been conditioned on or contain a non-collider that has been conditioned on. Recall that a collider is a variable into which two arrows point (e.g., A → B ← C). Colliders block spurious associations from flowing along that specific path. Further, it is important to identify collider variables since measuring and controlling for a collider variable can unblock a non-causal path.

For example, in Figure 1, there are 12 paths that connect ideological divergence A and turnout Y. Three of these are causal paths (i.e., A → Y, A → B → C → Y, and A → B → D → Y) and nine are noncausal paths (A → B → C ← E ← D → Y, A → B → C ← E ← F → Y, A → B → C ← E ← F → Y, A → B → D ← E ← F → Y, A → B → D ← E ← F → Y, A → B → C ← E ← F → Y, A → B → D ← E ← F → Y, A → B → D ← E ← F → Y, A → B → D ← E ← F → Y, A → B → D ← E ← F → Y, A → B → D ← E ← F → Y).
Further, all of these non-causal paths contain collider variables (B, C, and D, respectively) that block any spurious association along the respective path. Indeed, if we measured B, C, or D and controlled for them, doing so would create spurious associations between A and Y, where none otherwise existed. Furthermore, controlling for B, C, or D would control away part or all of the causal effect of interest by blocking causal paths.

We can identify causal effects if we can use observed covariates to block all open backdoor paths through which spurious associations flow. This is an informal statement of a more general strategy—called the "backdoor criterion"—for identifying a set of adjustment or conditioning variables that are sufficient to identify a causal effect of interest. The backdoor criterion simply tells us how to use the DAG to identify the adjustment set: the set of variables for which one must adjust to identify the causal effect of interest. Specifically, a valid adjustment set blocks all open backdoor paths without blocking any causal paths between treatment and outcome.

More formally, the backdoor criterion (Pearl 1993, 2009) states that a set of observed variables Z (which may be empty) satisfies the backdoor criterion relative to the total causal effect of D on Y if

1. No element of Z is a descendent of D; and

2. Z blocks all backdoor paths from D to Y.

To use the backdoor criterion, one writes down all the backdoor paths from D to Y. Then one searches for sets of variables that will block any open backdoor paths. For a set of conditioning variables that block all back-door paths, one next must verify that none of these variables is descendent of D. For a specific DAG, there may be more than one adjustment set which identifies the total effect. If any of these sets contains only measured variables, then identification of the effect of D on Y by conditioning on these variables is possible. If no such set exists (i.e., each conditioning set contains unmeasured variables), then the message of the DAG is clear: additional
effort to measure variables along the relevant noncausal paths is necessary.\footnote{Shpitser, VanderWeele and Robins (2010) developed the adjustment criterion which is more general than the backdoor criterion but has alternative uses. See Morgan and Winship (2014) and Elwert (2013) for a discussion of the relative differences between these two graphical identification rules.}

With this set of graphical identification rules defined, we now use them to understand whether the causal effect of ideological divergence $A$ on turnout $Y$ is identified. First, the effect for ideological divergence on turnout is identified, and this identification result depends on the exclusion restrictions in the DAG: we know the causal effect of ideological divergence is identified, since we drew the DAG so that there are no arrows flowing into it—there are no backdoor paths. This exclusion restriction implies that ideological divergence shares no common causes with turnout and thus is as-if randomly assigned.

Thus, in the DAG, as drawn, the impact of ideological divergence $A$ on turnout $Y$ is identified without the need to condition on any variables. Of course, that conclusion is completely a function of the decisions we made in constructing this DAG. In particular, following the theory in Rogowski, we assumed that there are no arrows into ideological divergence, which implies that ideological divergence is as-if randomly assigned. This highlights one strength of using DAGs to derive identification results. One can transparently see what is assumed for identification. One could argue this is an unrealistic assumption—that is a debate for this particular literature, at least the assumption is transparent.

In short, the backdoor criterion simplifies matters by stating that one only needs to worry about noncausal paths that start with an arrow into the treatment of interest. In this case, if we want to identify the total causal effect of $A$, we need only worry about noncausal paths with arrows into $A$, $\rightarrow A$. All other noncausal paths descend from $A$, and therefore we should not condition on any covariates along those paths.

Different decisions about the structure of the DAG would lead to different identification results. Suppose, for example, that we thought there were joint causes of individual political sophistication $E$ and candidate's ideological divergence $A$ – as might be the case, for example, if individuals tended to pay more/less attention to politics (and so become more politically sophisticated) in
districts in which candidates had previously been more divergent in which case previous ideological
divergence may cause current divergence and political sophistication. This case is depicted in
Figure 3, where we add an unobserved variable, W, that causes A and E. This addition creates
new backdoor paths between A and Y. One of these paths A ← W → E ← F → Y, contains a
collider variable E, which blocks that path without the need to condition on anything. However,
the presence of W creates other backdoor paths from A to Y (such as A ← W → E → D → Y)
that do not include any colliders. As such, if W remains unobserved the total effect of ideological
divergence is no longer identified without conditioning.

Alleviating this unfortunate situation would require that we measure and condition on W to
identify the effect of A on Y. Conditioning on W would block all the backdoor paths and identify
the effect of A on Y. Alternatively, we could condition on both E and F and that would close
all the backdoor paths and thus identify the effect of A on Y. We would argue that the DAG
in Figure 3 most closely fits the identification strategy in Rogowski (2014). He assumes that
conditioning on E and F, is sufficient to close any backdoor paths between ideological divergence
and turnout. One could debate whether that is true or not, but the DAG makes the assumption
clear.

While we have so far focused on the total causal effect of ideological divergence on turnout,
there is no reason not to ask, given this causal structure, whether the causal effects of the other
measured variable, political sophistication E, is identified? For this exercise, we use the DAG in
Figure 3. We begin by listing all the non-causal backdoor paths between the causal variable of
interest (political sophistication E and turnout Y. There are five: (1) E ← F → Y, (2) E ← F
← U → Y, (3) E ← W → A → Y, (4) E ← W → A → B → C → Y, and (5) E ← W → A →
B → D → Y. The next question, then, is whether there is a set of measured variables along
the remaining backdoor path for which we can control for to achieve identification? The answer
is yes. If we condition on A and F, the total effect of E on Y is identified, since that will block

\[16\] There are also four causal paths between E and Y (E → C → Y; E → D → Y; E → B → C → Y; and E → B → D → Y) and we might consider adding a fifth by allowing for the possibility of a direct effect between political sophistication and turnout.
all of the open backdoor paths between E and Y.

Beyond deriving identifiability results, we can also derive testable implications about the structure of the DAG. For example, the DAG in Figure 1 implies that ideological divergence and political sophistication should be independent. This serves as a testable implication of this DAG. In the data, we should find that measures of sophistication and ideological divergence are independent of each other. If they are not independent, that would serve as empirical evidence that the causal model represented by this DAG is wrong.

The example in Figure 1 also contains several causal mechanisms. For example, ideological divergence, A, has a direct effect on turnout and also two indirect effects through alienation, D, and importance, C. Importantly, unlike total effects, the identifiability of direct and indirect effects cannot be determined using the backdoor criterion. Identification of these effects requires additional assumptions, and must be deduced from graphs using alternative graphical criteria (Pearl 2014). Indeed, we will often be able to identify the total effect but not separately identify the indirect or direct effects that make up that total effect. That is the case in Figure 1. While

---

17Because of the algorithmic nature of the analysis of DAGs it has been possible to build computer programs to assist in the analysis of DAGs and the identification of such independence implications. A particularly useful, and free to use, example with a convenient web interface is at http://www.dagitty.net/.
the total effect of A can be identified, the separate direct effect of ideological divergence on turnout cannot be identified via conditioning and neither can any of the indirect effects since B, C, and D are unmeasured. Identification of these indirect and direct effects would be possible if we could measure the mediating variables and invoked additional assumptions. For example, sometimes analysts invoke an assumption known as sequential ignorability to identify direct and indirect effects. This assumption rules out confounding between mediating variables and the outcome (Imai et al. 2011). See VanderWeele (2015) for a complete introduction to indirect and direct effects.

Once a theory has been articulated (at least in part) as a DAG and the backdoor criterion has been applied to derive identification results, the analyst can then proceed to estimation. A DAG implies that nonparametric estimation is possible, but given data constraints, additional modeling assumptions may be necessary for estimation. The DAG, however, is silent on this point, since it is only concerned with nonparametric identification. Estimation concerns are subject to issues of finite samples while DAGs tell us what could be learned if the sample size were infinite.

To summarize, the procedure for using a DAG to guide identification follows a set of steps. First, based on theoretical and scientific knowledge, the analyst writes down a DAG that reflects the causal structure he or she thinks is true. This includes specification of the causal relationships between both measured and unmeasured variables and a set of exclusion restrictions. Ideally, this structure is based on theoretical knowledge, previous empirical work, and subject matter expertise; though the use of a DAG might well expose parts of this structure that are on less than firm ground (and so, if critical to identifying important effects, should be a locus of debate and further inquiry). Next, based on the DAG, the researcher can use the backdoor criterion to determine what total causal effects are identified, identify the set of covariates that one needs to condition on to identify effects, and derive a set of testable independencies, which can help validate the causal structure in the DAG. Importantly, all stated identification results are conditional on the validity of the DAG (Pearl 1995). The idea is for researchers to use the DAG to clearly convey what he or she thinks is the causal structure of the data generating process as derived from theory.
The DAG can then be critiqued by other scholars and defended or revised by the researcher.

Finally, we think DAGs usefully illustrate how phrases like "no omitted variables" are not useful for thinking about causal identification. It is unnecessary to observe all the causes of Y to identify the total causal effect of a given treatment. Instead, we need to observe the set of covariates that can block all backdoor paths between D and Y. It is this set of covariates that make D as-if randomly assigned. In the DAG, often we are not trying to identify all the effects but to determine what we need to observe to identify one specific effect. Thus DAGs provide absolute clarity about the role of control variables. They are used to block open backdoor paths, so we can identify specific causal effects—and there is no guarantee that the same set of variables (e.g., those included on the right hand side of a regression) will be sufficient to simultaneously identify the causal effect of all variables in the set.

In the rest of the paper, we use DAGs in two different empirical contexts to demonstrate (1) how they can guide empirical analysis in more complex situations, (2) how they can bring clarity about the substantive debates that are most consequential for the interpretation of empirical work, and (3) how they can be used to provide a principled answer to the question of the causal status of "control" variables in regression models. Before turning to these examples, however, we briefly contrast two quite different general approaches to conditioning. Doing so makes it clear that the question of identifiably is quite different from that of how one practically estimates causal effects and, we hope, alleviates some confusion that has existed among practitioners about the necessity of methods such as matching when attempting a causal analysis.

3 Control Variables in Causal Models

We begin by contrasting the regression approach to causal modeling with statistical techniques that designate a single variable as the "treatment" and all others covariates as controls.

3.1 The Nuisance Parameter Approach

Under the selection on observables identification strategy, the analyst needs to apply a statistical method to adjust for observed confounders. By far the most common statistical technique applied
to data across the social sciences is that of regression. Here, we use the word regression to describe
a wide range of statistical techniques that impose very different functional form assumptions. In
this essay, we group under the term regression not just linear regression but also generalized
linear models which includes various types of logit and probit models as well as count models
like negative binomial regression. We include nonparametric regression models as well. We
group together any statistical method that produces interpretable output of some type for all
right-hand side covariates. For example, generalized additive models allow the analyst to specify
that continuous right hand side predictors be modeled nonparametrically to allow for arbitrary
nonlinear partial associations (Hastie and Tibshirani [1990]; Keele [2008]). For any right hand side
covariate with a nonparametric fit, one can produce a plot that allows for a basic interpretation of
the partial association between the covariate and the outcome. The method of kernel regularized
least squares (KRLS), which relaxes the additivity assumption from generalized additive models,
also allows analysts to back out interpretable partial associations for all right hand side covariates
(Hainmueller and Hazlett [2013]).

There are two primary alternatives to regression models for statistical adjustment: weighting and matching. We label both of these methods nuisance parameter approaches to control
variables. Regression methods stand in contrast to nuisance parameter approaches in which the
association between control variables and the outcome are never actually estimated.

Matching is better known than weighting methods in political science but weighting sees
widespread use in fields such as epidemiology, where it is often referred to as inverse probability
weighting (IPW). Weighting methods include both marginal structural models and a class of
models known as “doubly robust.” See Glynn and Quinn (2010) for an introduction to such
methods in political science. Typically weighting methods are considered to be semi-parametric,
and interpretable coefficients are not produced for any of the covariates included as controls. The
only coefficient estimated is that of the treatment effect.\footnote{There are, of course, exceptions. Ko et al. (2007) propose a double robust estimator through matching and regression. Under this approach, coefficients are produced for all right hand side variables.}

Usually, the contrast drawn between methods such as matching and weighting and regression
is that the former are nonparametric or semiparametric methods of adjustment while regression
is parametric. That is true when one compares matching to regression estimated with least
squares, but methods like KRLS are also fully nonparametric forms of adjustment. The other
difference, which we highlight here, is that under the nuisance parameter approach the coefficients
for covariates that are included as controls are not directly estimated. For example, in a standard
matching analysis, while one might match on many covariates, only a single coefficient for the
treatment is produced. Here, we group together statistical methods that only produce a treatment
effect estimate as the nuisance parameter methods.

Is the nuisance parameter approach to causal effect estimation just an accident of these
methods or actually a feature? As we will argue below, it is a feature. Under the nuisance
parameter approach, a single covariate is designated as the treatment. Since no parameters for
control variables are estimated, the analyst only gives that single estimate a causal interpretation.
In contrast, methods that produce an estimated parameter for every covariate in the model tempt
analysts to interpret all these estimates as causal effects. As we have seen above, however, such
interpretations are not justified in general. Instead, each case in which one seeks to make a causal
claim must be separately considered relative to the causal structure that underlies the analysis.\footnote{It is worth emphasizing again that such a structure exists whether it is made explicit or not and whether or not a DAG is used to describe it.}

### 3.2 The Single Coefficient Approach

To illustrate both the dangers of interpreting all covariates in a regression model causally, as
well as how DAGs can clarify what can be so interpreted in a given model, we provide two
additional empirical illustration drawn from different substantive fields in political science. For
the first empirical application, we examine an important question drawn from the literature on the
democratic peace. In international relations, a great deal of research has focused on identifying
the factors that promote or inhibit militarized conflict; and, much of this has focused on two
specific questions: are democracies less likely to engage in conflict with other democracies and
are trading partners more or less likely to engage in conflict with each other? Many researchers
have theorized about these relationships and estimated empirical models intended to provide
evidence about them (Maoz and Russett 1992, 1993; Oneal et al. 1996; Russett 1990, 1993;
Oneal and Russett 1997; Russett and O’neal 2001). Our example draws on a number of these
works and so we do not pretend that our DAG faithfully represents any one scholar’s theory or
captures all the important debates in this literature. Instead, the point of the illustration is to
bring out the way scholars in this area might fruitfully use DAGs to better justify their casual
claims and in some cases understand why they should not be made.

Figure 4 contains one DAG that represents a possible causal structure for interstate conflict.

We assume that the researcher is interested in estimating a single total causal effect (specifically,
the effect of joint democracy on conflict) from a regression model that includes as controls
the other measured variables in the DAG (we consider a case in which a researcher wants to
compare multiple causal effects in the next example). The question we ask then is what causal
interpretation, if any, is justified for these control variables?

Before answering that question, however, first notice that, given the DAG in Figure 4, the
main causal effect of interest, the total effect of joint democracy on conflict, is identified. While
there are many backdoor paths between joint democracy and conflict, each of these run through
at least one of the observable non-collider variables (allies, relative growth, trade, or contiguity)
and so can be blocked by including those variables as controls in a model. Likewise, there are no
variables that mediate between joint democracy and conflict and no collider variables introduced
by conditioning on any of the observables. Thus, the total effect of joint democracy on conflict
can be nonparametrically identified by conditioning on the four other observable variables in the
model. Suppose now that we believe that the DAG represents a linear model without interactions.
If so, what interpretation can we give to the estimated associations between the control variables
and conflict in this model?

More concretely, suppose that the actual regression model that an analyst estimated were the
following:

\[ Conflict = \beta_0 + \beta_1 Contiguity + \beta_2 JointDemocracy + \beta_3 Allies + \beta_4 Trade + \beta_5 Rel.\Delta GDP. \]  

Now what interpretation can we give to each of the estimated coefficients? Below is a summary of
how we can interpret each of the regression coefficients assuming the causal structure in Figure 4
is true:

- \( \beta_1 \) represents the direct effect of contiguity, and thus does not account for the indirect
effect through joint democracy.
- \( \beta_2 \) represents the total effect of joint democracy since this total causal effect is identified.
- \( \beta_3 \) does not represent a causal effect since the effect of allies is unidentified.
- \( \beta_4 \) does not represent a causal effect since the effect of trade is unidentified.
- \( \beta_5 \) does not represent a causal effect since the effect of relative economic growth is uniden-
tified.

Thus, only one of the five coefficients in Equation 1 represents a total causal effect, that
of joint democracy. In contrast, the total effect of continuity cannot be estimated from this
regression model, given the causal structure in the DAG. The regression coefficient on the measure of contiguity only represents a direct effect. Further, no effects (either direct or total) of trade, relative economic growth, or allies are identified. Therefore, in these cases, the estimated regression coefficients do not represent a causal quantity of any kind. While measures of trade, relative economic growth, and allies are necessary to identify the effect of joint democracy on conflict, their casual effects are not identified under this DAG. This demonstrates the hazards of attempting to interpret all the coefficients in a regression model. Unless one understands the causal structure, regression coefficients can represent any number of quantities.

It will strike some readers as absurd for us to say that $\beta_4$ represents nothing since the effect of trade on conflict is unidentified. One might say: aren’t regression coefficients always interpretable even when they don’t represent causal effects? In fact, that is true. What we are saying is that $\beta_4$ represents an association but not a causal effect given the DAG in Figure 4. Therefore, while one might choose to interpret the coefficient as an association, it cannot be given a causal interpretation if one believes the structure of the DAG in Figure 4. Thus while trade and conflict may be (negatively) correlated in this regression, one should not take this as evidence that trade acts as lever, so that one should expect an increase in trade to lower levels of conflict.

In general, this leads to a broader point about estimated associations (as opposed to causal effects). In a given empirical literature, there are almost always some relationships for which estimated associations in different empirical models are unstable (i.e., when a given covariate, D, is included in different specifications for Y, the estimated association between D and Y is sometimes statistically significant and sometimes not and/or its sign changes frequently). It is easy to accept the conclusion that the estimated coefficients for these kinds of relationships are just associations that are likely spurious. It is more challenging, however, to accept this judgment for more stable associations—that is, a relationship between D and Y, that across many different empirical models, reliably produces statistically significant coefficients with the

$^{20}$The identification of this direct effect arises from the structure of this graph, but cannot be derived using the backdoor criterion. However, it can be derived from the adjustment criterion for direct effects which is related to the backdoor criterion but is more general.
Figure 5: DAG structures that lead to regression coefficients that cannot be interpreted as causal effects. In panel (5a) $L$ has no effect on $Y$ and in panel (5b) the confounder, $L$, has an effect on $Y$. In both cases, one must condition on $L$ to identify the effect of $D$ on $Y$, but the effect of $L$ on $Y$ is not identified.

Whether one can interpret such coefficients as causal depends completely on what causal structures one believes to be true and there is no reason to assume that the stability of the estimated coefficients indicates, by itself, a causal effect. It is entirely possible that there are set of unobserved confounders that are themselves stable – thus producing a reliable estimated association in many different models and specifications.

Returning to the conflict example, if one really believed that the DAG in Figure 4 was correct, but also knew that in many previous empirical models trade produced a stable association with conflict, then one should conclude that the question of whether trade has a causal effect on conflict remains open and the empirical agenda of the literature should surely include an effort to identify the causal effect of trade in some way (e.g., finding observable factors could block the backdoor paths in the DAG for trade).

Finally, we outline two specific structures in a DAG that lead to this lack of causal interpretability for a regression coefficient. Figure 5 contains two DAGs with the general structure that leads to this phenomenon. In both cases, $L$ blocks a backdoor path between $D$ and $Y$, therefore we must condition on $L$ to identify the effect of $D$ on $Y$. However, in both cases, the effect of $L$ is not identified, therefore it would be problematic to provide any interpretation for a regression coefficient for $L$ when one is trying to estimate the effect of $D$ on $Y$. Therefore, if one suspects these structure is in operation, one must take care in the interpretation of coefficients
as casual effects.

### 3.3 The Multiple Coefficient Approach

In the last example we considered the case in which an analyst specified a model with the goal of identifying a particular causal effect—the impact of democracy on conflict—and so chose a set of control variables considered sufficient to block all open back door paths to identify that effect. As we saw, it is not always possible to give the estimated associations between these control variables and conflict a causal interpretation. Another common situation, however, occurs when a researcher has several different causal hypotheses that he or she wishes to simultaneously examine (and adjudicate among) and so includes measured variables that capture these different hypotheses in a single regression model. Clearly, given the structure of such studies, the estimated coefficient on each variable requires a casual interpretation.

As an example, consider a researcher who has collected survey data on a set of variables capturing the major generally accepted drivers of vote choice and wants to estimate the relative importance of these drivers for the election at hand. A stylized, but eminently recognizable, selection of such vote drivers might include party identification, perceptions of economic performance, relative issue stances (perhaps aggregated into general left-right positions), and perceived candidate qualities, characteristics, or images. Each of these drivers comes from a well-developed theoretical literature and we suspect that most electoral researchers would agree that these concepts are thought to have causal effects on vote choice—recognizing, of course, that these effects may or may not be apparent in any given election. Clearly, the hypotheses that our electoral researcher wishes to test are causal and so the question is whether all four causal effects (party id, economic performance, candidate quality, and relative issue position) can be tested in the kind of single equation empirical model usually used in such situations? Again, the DAG answers this question.

In short, the DAG approach is useful if one can show in the DAG that each of the causal effects of interest are identified. For example, the DAG in Figure 6 provides one plausible causal
structure for vote choice with these variables and others as causes. Using the backdoor criterion, we can determine that in this DAG three of the main causal effects are identified. Thus, we can identify the effects of candidate qualities, economic perceptions, and issue positions. It is worth repeating that these causal inferences are conditional on the theory that created this DAG that represents the data generating process. If the analyst articulated a different data generating process, they would draw different conclusions about what causal inferences are possible.

This pattern of identification stems from two sources. First are the arrows that do not appear in the diagram (the exclusion restrictions). Specifically the DAG assumes that there are no unmeasured confounders linking the various nodes in the graph in ways that create backdoor paths between any of the variables of interest and the outcome (for example, there are no unmeasured variables that impact a respondent’s issue positions and her views of the economy). In a mature literature like this one, this is perhaps a reasonable position to take—and is, at least implicitly, the position taken by many electoral researchers that estimate and interpret vote
choice models. Note, however that the effect of party identification is not identified, since there
is an unblocked backdoor path—running through \( U \)—between it and vote choice. Here, \( U \) might
represent early parental influences that condition both party id and vote choice.

Further, what is very clear from this DAG—and perhaps not otherwise—is that this happy
situation of simultaneous identification of the causal effects for each causal hypotheses of interest
(other than party id) depends almost entirely on the fact that in this theoretical specification,
party id is a sort of “super-blocker” of backdoor paths. By controlling for party id, one manages,
in one stroke, to simultaneously block the backdoor paths that would otherwise confound each
of the three (non-party id) variables of interest.

Next we assume that this data generating process is linear without any interactions. Therefore,
we estimate the following regression model:

\[
Vote = \beta_0 + \beta_1 PartyId + \beta_2 Demographics + \beta_3 Cand.Qualities + \beta_4 Issues + \\
\beta_5 Econ.Perceptions + \beta_6 ActualEconomicPerformance
\]  

(2)

In this regression model, vote choice is function of all the observed covariates in the DAG.
Under standard advice about complete specifications, one would hesitate to leave any of these
covariates out of the right hand side of the model. Following the discussion above, the coefficients
for economic perceptions, candidate qualities and issues can be interpreted as estimates of the
identified total effects, while the coefficient for party identification cannot be given a causal
interpretation as any type of effect. In this regression model, party identification serves to block
open back door paths, but its coefficient cannot be given a causal interpretation since it causes
the other covariates and shares an unmeasured set of causes with vote choice. Indeed, even if
party id did not share common causes with vote choice, the fact that the regression specification
conditions on the decedents of party id means this estimated association is not a total causal

\[^{21}\text{It is however, a direct effect, in the case of no common cause with vote choice. Likewise, under these conditions, the total effect of party id could be estimated in a bivariate regression of party id on vote choice.}\]
effect of the real economy on vote choice (which is identified by conditioning on party id and
perceptions, but should theoretically be zero given and there is no such path in the causal diagram
— the entire effect runs through economic perceptions). Finally, the coefficient on demographics
cannot be interpreted as a total causal effect in this model (since the model conditions on its
descendants) and neither does it represent a direct effect. However, its total effect can easily be
estimated (given the DAG) with a bivariate regression of demographics on vote choice.

What we hope this DAG makes plain is that if one takes the theoretical position that party id
is the “unmoved-mover” — as many students of American electoral behavior do (e.g., Campbell
et al. (1966); Miller and Shanks (1996); Green, Palmquist and Schickler (2004); Green and
Palmquist (1994)), the DAG corresponding to this assumption justifies exactly the kind of single-
model rival hypotheses testing that has been the usual practice of electoral researchers for decades.
Indeed, this conclusion holds even if party id shares unobserved common causes with vote choice.

Further, if this kind of DAG were widely accepted in the literature (and again, we think that
some DAG similar to this one is accepted among many in this area of research), it makes it clear
that a researcher interested in testing some novel causal hypothesis should first and foremost think
carefully about how the variable(s) capturing that hypothesis fit into the DAG. If the variable is
downstream of party id, then testing within a single model is likely feasible. In contrast, if the
variable is a precursor to party id, bypasses party id in the causal diagram or if these competing
variables effect each other, the identifiably of the causal effect of the new variable is likely to be
more complex.

Of course, one may reject the characterization of party id as the unmoved mover - as would,
for example, many electoral researchers in democracies outside the United States (Butler and
Stokes 1969; Thomassen 1993; Kaase 1976). Even in the American case, a large literature
recognizes that short-term factors like issue positions (Carsey and Layman 2006; Highton and
Kam 2011), candidate characteristics (Page and Jones 1979) and economic perceptions (Fiorina
1981; MacKuen, Erikson and Stimson 1989) can, at least under some conditions, move party

\footnote{We recognize that it may not be obvious why this holds, given that the rules for identifying direct effects versus total effects are beyond the scope of this article.}
id. Of course, few of these studies would go so far as to say that party id is fully caused by, rather than a cause of, these short-term perceptions and attitudes. Instead, most seem to have in mind some, more or less explicit, notion of mutual causation (Page and Jones 1979; Markus and Converse 1979; Achen 2002a). Those who reject party id as an unmoved mover would be more likely to agree with a DAG like that in Figure 7. Now party id shares an unmeasured cause with issue positions. In this case $U$ could represent socialization, media framing, or that party id and issues positions affect each other (a situation that, as we show in the appendix can usefully represent what most scholars really mean by mutual causation). As a result, there is an unblocked backdoor path between issue positions and vote running through $U$, and so the total effect of issue positions is now not identified. If one were even more pessimistic and thought that economic perceptions, party id, issue positions, and candidate qualities also shared common unobserved causes, then none of these total effects would be identified either.

This makes it plain that the status of party id as a cause or consequence of issues positions, economic perceptions, and the like should rightly be at the heart of the theoretical debate in electoral research—since one’s theoretical position on this issue must have important implications for how one can identify casual estimates of the effects of these drivers of vote choice.

More generally, this example makes it plain that it is possible, under certain not entirely implausible assumptions, to conduct a test of rival hypotheses in a single regression model, where the total effects of interest are all identified. Further, our example illustrates that theoretical questions are critical to the justification of such a test (and to the specification and interpretation of empirical models more generally). In our example, the interpretation of coefficients from a regression model of vote choice largely depends on what one believes about the role of party identification. Different theories of party id lead to very different identification results. By instantiating these theories in competing DAGs, the empirical implications of such assumptions are made strikingly explicit.

---

23 One exception is Jackson 1975 who concludes “party IDs are highly influenced by people’s evaluations of what policies each party advocates relative to their own preferences, party affiliations have little or no direct influence on the voting decision...”

24 Of course, we do not mean to imply that careful empirical researchers have not up until now been aware of
4 Conclusion

In this paper we attempt to explain how careful theoretical thinking about causal structures is necessary to answer a set of questions that have long troubled political scientists: which control variables should be included in an empirical model and what interpretations of those parameters are justified? Further, in examining these questions, we have also been able to speak to related issues like the conditions under which multiple causal hypotheses can be examined in the same empirical model (for example, studies framed as a race of variables). In addition, we argue that relatively new graphical tools, rooted in rigorous mathematical formulations equivalent to better visualized and understood in a DAG than in traditional verbal and/or mathematical presentations.
known (among political scientists) formulations in the literature on counter-factual inference, can greatly facilitate the ease with which applied empirical researchers can apply the lessons of the causal modelling literature to their own empirical work.

Using these tools, it is relatively straightforward to provide examples in which some or all of the coefficients on control variables in a regression model will have no causal interpretation or only a partial interpretation (e.g., as direct effects). That said, these same tools can certainly be used to make plain the sets of assumptions that allow such interpretations and so can be empowering to empirical researchers ready to defend those assumptions.
References


Butler, David and Donald Stokes. 1969. “Political change in Britain: Forces shaping electoral choice.”.


Fiorina, Morris P. 1981. “Retrospective voting in American national elections.”.


Richardson, Thomas S and James M Robins. 2013. “Single World Intervention Graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality.”


A Design Based Inference and the use of DAGs

We briefly turn to an important aside that highlights contexts in which DAGs may be of less practical use even though the essential logic of the DAG is unchanged. In the renewed emphasis on causal inference in political science, many authors have emphasized the need for randomized experiments and design-based inference, where design based inference only includes natural experiments such as regression discontinuity designs (Gerber and Green 2012; Dunning 2012). In this approach to causal inference, DAGs are seldom used. Why is that the case? In this approach to causal inference, the goal is to rely on an identification strategy that produces a DAG like the one in Figure 8. This DAG implies that \( D \) causes \( Y \), but it also encodes the assumption that there are no other covariates that are common causes of both \( D \) and \( Y \), which clearly implies that the causal effect of \( D \) on \( Y \) is identified. A DAG of this type could be used to represent the causal effect in a randomized experiment or in an ideal natural experiment since randomization implies that \( D \) and \( Y \) have no common causes.

\[
\begin{align*}
D \rightarrow Y
\end{align*}
\]

Figure 8: DAG with no common causes for \( D \) and \( Y \).

The great advantage to randomized experiments is that the researcher creates a DAG like the one in Figure 8 by design. That is, controlled randomization can make this DAG true by construction. When this is the goal, DAGs add little to the research process. The design-based approach is quite prominent in both political science and economics and is partially responsible for why DAGs see little use. However, the most common identification strategy in political science is that of selection on observables. Here, we think DAGs add considerable value, as the examples which follow illustrate. Moreover, in many more complex natural experiments DAGs may be useful for clarifying identification conditions see O’Malley et al. (2014) for an example.
B Mutual Causation in a DAG: The Case of Party Id

Here, we discuss how mutual causation can be represented in a DAG using party id and issue positions as an example. Since DAGs are causal diagrams and temporal ordering is fundamental to such diagrams, the informal idea of reciprocal causation (with accompanying double headed arrows) is not allowed (and, at a more fundamental level, is insensible). Instead, we can capture the idea that, for example, issue positions are a partial cause of party id which in turn is a cause of issue positions simply by drawing an arrow from party id to issue positions and giving both nodes a common unmeasured cause.

More specifically, authors who invoke ideas of reciprocal causation usually have in mind some temporal sequence like that captured in the following DAG (where each node occurs as a different instant in time):

![Figure 9: A dynamic structure for party id and issue positions.](image)

However, if we assume that the first two nodes in the sequence are unobserved (because for example our data come from a cross-sectional survey that ask a question about current party id and then a few minutes later a question about current issue positions), we can replace them both by one node for an unmeasured cause that impacts both party id and issues as measured in our survey.
Figure 10: Mutual causation as a shared unobserved cause.