When Controls are Out of Control
Graphical Theory Building for Politics

Zbigniew Truchlewski
London School of Economics

Akisato Suzuki
Dublin City University

Alexandru D. Moise
European University Institute

May 21, 2022

Abstract

A general problem of empirical research is that controlling for many variables is often a substitute for, rather than guided by, a theory on the causal structure among variables. This obfuscates causal assumptions and their connection to empirical modeling. To rectify this problem, we identify two under-appreciated benefits of Directed Acyclic Graphs (DAGs). First, DAGs compel us to theorize the entire causal structure and connect it with empirical modeling explicitly. This helps us define the treatment effect of our interest theoretically and interpret what statistical estimates mean causally. Second, DAGs increase the transparency of the theorized causal structure, thereby promoting productive discussions on the plausibility of the causal assumptions. This is a more rigorous and principled alternative to the question "Did you control for...?" that is not guided by a causal structure. As an example, we construct a DAG from a canonical study of civil wars (Fearon and Laitin, 2003). We reveal inconsistent causal assumptions, and find a substantial difference in the estimate of one of the key treatments.

Keywords: DAG, theory, causal inference, confounder, collider

1 Introduction

It is common to see regression controlling for many variables – classically known as "garbage-can regression" (Achen, 2005). It is often used as a substitute for, rather than guided by, a theory on how each variable is causally related to one another. The usual theory section of research discusses only the causal relationship between the treatment(s) and the outcome and not a larger causal structure such as what factors can bias this relationship. The latter matter is left to the research design section, which usually mentions, as criteria on what variables to include in a statistical model, (1) variables that are causally related to the outcome of interest, or (2) variables that are causally related to the treatment of interest and the outcome of interest. These criteria are imprecise to identify a set of variables that enables us to interpret a statistical estimate causally. Not all variables meeting either criterion are necessary for causal identification (e.g., pre-treatment covariates affecting the outcome only). Some are even harmful (e.g., post-treatment variables). To select "good" controls from "bad" controls (Cinelli, Forney and Pearl, 2020) and interpret statistical estimates causally, we must theorize the entire causal structure among variables. In short, there is oftentimes a disjunction between theory and empirics (Hall, 2003).

To rectify this problem, we identify two under-appreciated benefits of Directed Acyclic Graphs (DAGs) – a graphical representation of a causal model usually used for causal identification. First, we show DAGs compel us to theorize the entire causal structure and connect it with empirical modeling explicitly. In so doing, DAGs help us define the treatment effect of our interest theoretically (Lundberg, Johnson and Stewart, 2021) and interpret what statistical estimates mean causally (Keele, Stevenson and Elwert, 2020): e.g., given the DAG, is it a total or direct effect that we are after, and does a statistical model correctly identify that?

Second, we argue that DAGs increase the transparency of the theorized causal structure and help promote productive discussions on the plausibility of the causal assumptions, thereby avoiding the problematic question of "Did you control for...?" that is not guided by a causal structure. It is extremely difficult to see the assumed causal structure among several variables without an image such as a DAG. Transparency in empirical research is key for scientific progress. In political science, transparency has improved first via the almost mandatory publication of replication datasets and then the recommendation of preregistrations. We argue a next step is the use of DAGs so that an assumed causal structure also becomes transparent.

We use an actual political science example in our explanation, unlike most of the literature (e.g., Cinelli, Forney and Pearl, 2020; Pearce and Lawlor, 2016; Pearl and Mackenzie, 2018; Hernán and Robins, 2020; Morgan and Winship, 2015) – for an exception, see Keele, Stevenson and Elwert (2020). We select a canonical study of civil war, Fearon and Laitin (2003), not only because it is probably one of the most cited quantitative political science articles (Breton, 2015), but also because it has been used in many methodological studies as a test case (e.g., Acharya, Blackwell and Sen, 2016; Hug, 2010; Montgomery and Nyhan, 2010; Muchlinski et al., 2016). Therefore, many are familiar with the study, which makes it easy for the reader to focus on the methodological points.

Our aim is not to diminish the significant contributions Fearon and Laitin (2003) made; the study was published well before the causal inference literature became familiar to political scientists. Nor is it to make novel contributions to this particular literature on civil war.[1] We use Fearon and Laitin (2003) simply as a most convenient example (in that many know it already) to explain why the lack of DAGs deserves attention to advance empirical research further.

The problems outlined in this article are by no means unique to Fearon and Laitin (2003). We conducted a review of all selection-on-observables articles in the top three political science journals (the *American Political Science Review*, the *American Journal of Political Science*, and the *Journal of Politics*) for 2021. The issues raised in this article are most relevant to selection-on-observables studies (although DAGs are also useful for (quasi-)experiments; see section 4.3). We coded whether articles used a DAG or at least discussed the causal structure behind their empirical models (see Table 1). Many articles still used selection on observables. Of these, a rather small share discussed the causal structure.[2] Only two papers used DAGs. Overall, we find that the problems that we identify in this article apply to the majority of current empirical research in political science.

[1] There have already been many substantive studies to advance/critique Fearon and Laitin (2003) (e.g., Cederman and Girardin, 2007; Buhaug, Cederman and Rød, 2008).

[2] Our coding used a low threshold, a paper needing to discuss how at least one variable is causally related to the treatment and the outcome and if it is pre-treatment or post-treatment. In fact, many of the papers that we code as discussing their causal structure still fall pray to the issues discussed here, including inconsistencies in the assumptions and the unclear identification of types of effects. For details on our coding scheme, see the appendix.

Journal	N. Art.	N. Sel. Obs.	N. DAG	N. Causal Str.	% Sel. Obs.	% Causal Str.
APSR	82	42	2	10	51%	24%
AJPS	61	22	0	9	36%	41%
JOP	110	50	0	12	45%	24%

Table 1: Summary statistics for our coding of the APSR, AJPS, and JOP in 2021. N. Art. is the total number of articles; N. Sel. Obs. is the number of articles using a selection-on-observables approach; N. DAG is the number of articles using a DAG; N. Causal Str. is the number of articles that discuss the causal structure of their empirical model; % Sel. Obs is the share of selection-on-observables articles; % Causal Str. is the share of articles that discuss their causal structure.

One well-known issue arising from the lack of the consideration of causal structure is documented by Keele, Stevenson and Elwert (2020) and Westreich and Greenland (2013). On one hand, we run regressions and interpret all coefficients as total average effects. This forces us to have an unrealistic theory of causal structure, where no independent variable causes, or is caused by, each other. On the other, if we want to have a realistic causal structure, we cannot keep reading all coefficients as total effects, as some variables can be mediators and make some coefficients represent direct rather than total effects. In short, we need a causal structure to interpret the causal meaning of each coefficient correctly. Our article is different from these previous works, in that we emphasize the importance of articulating a theory on the causal structure among variables from a perspective of theory building, transparency, and interpretability, regardless of the statistical method or causal identification strategy. While we use multiple regression to exemplify our argument, the argument equally applies to other statistical methods for observational studies (e.g., matching) and even to (quasi-)experiments (Montgomery, Nyhan and Torres, 2018).

The article proceeds as follows. First, we explain how DAGs connect a theorized causal structure and empirical modeling. We empirically show that using DAGs can make significant differences in substantive conclusions we draw from statistical models. Second, we illustrate how using a DAG increases the transparency of causal structure and helps identify inconsistent causal assumptions. We find the theoretical arguments in Fearon and Laitin (2003) imply conflicting DAGs that have different implications for which variables to include in an empirical model. Third, we discuss the implications of our article for the common points in causal inference: (1) DAGs are necessary even if all covariates are pre-treatment, (2) DAGs are useful even if one wants to avoid making explicit causal claims, (3) DAGs are also relevant for (quasi-)experiments, and (4) Multiple DAGs improve the credibility of one's causal assumptions. The final section is the conclusion.[3]

2 Connecting the Theorized Causal Structure and Empirics

This section illustrate how DAGs allow us to connect a theory on causal structure with empirical modeling. DAGs are a tool to visualize our assumptions on the causal relationships among variables. They are called directed because the graphs indicate the direction of causality between variables. They are acyclic because effects do not go back, i.e., there is no feedback loop to describe an endogenous relationship; such a relationship can instead be described by a temporal order from time at t_0 to time at t_1. Finally, they are graphs because they picture the causal relationships among variables with nodes and arrows. DAGs represent our assumptions on the causal structure among variables. They are nonparametric and applicable to any empirical modeling for causal inference.[4] The structure of DAGs guide us which variables we should or should not adjust for, to estimate the treatment effect of a certain variable. For more on the basics of DAGs and their use for causal identification, see Section 2 of the appendix.

2.1 An applied example

We construct one possible DAG illustrating the causal relationships among the variables in Fearon and Laitin (2003), partly based on what these authors suggest and partly based on other research and our own reasoning. Fearon and Laitin (2003) look at the determinants of the onset of civil war from 1945 to 1999, in a sample of 161 countries with at least half a million citizens in 1990. Their database contains 127 conflicts out of 6610 country years. They consider how civil war might be caused by prior civil war (denoted as `PrWar` in our DAGs), per capita income (`Income`), population (`Pop`), mountanous areas (`Mont`), noncontiguous states (`NonCont`), oil exporter (`Oil`), new states (`NwState`), political instability (`Instab`), democracy (`Dem`), ethnic

[3] The R codes that verify many of the implications of the DAGs in this article are available in the appendix.

[4] The DAG resembles the Structural Equation Model (SEM) on the surface, but they are qualitatively different. The SEM is an empirical strategy while the DAG is a theoretical model. In short, the DAG is necessary to interpret the SEM causally.

fractionalization (EthFrac), and religious fractionalization (RelFrac).

Fearon and Laitin (2003) argue against the prevailing wisdom at the time, that greater ethnic and religious fractionalization are more likely to lead to civil wars (the ethnicity argument). In their analysis they show that indicators for ethnic and religious fractionalization (as well as their outcomes such as ethnic grievances) are no longer predictive once they account for per capita income. Instead, the authors argue in favor of "conditions for insurgency" as the main causal factors producing civil wars (the opportunity argument). Insurgencies are a type of military conflict conducted by smaller bands of armed individuals, using guerrilla warfare. Fearon and Laitin (2003) argue that most civil wars are caused via insurgency by small groups. According to them, ethnic grievances are widespread across many countries, while few experience civil wars. Moreover, if civil wars are started by such small insurgent groups operating in weak states, then it is not necessary to have larger ethnic minorities that espouse grievances.

Our DAG is presented in Figure 1 (ignore the blue and red coloring for EthFrac and CvlWar at the moment). The theoretical rationale for this DAG is available in the appendix. The main point is that we distinguish between (almost) time-invariant variables and time-variant ones, as the former generally should not be explained by the latter. We made the DAG as realistic as possible given the current state of the literature. If the reader finds our DAG contains a problematic causal assumption, this is precisely the reason why we are advocating DAGs: the DAG clarifies the causal assumptions explicitly so that the reader can see a problematic one right away, while just listing controls hides these assumptions.

The DAG implies we need a different statistical model to estimate the total average treatment effect ("total effect" in short), depending on which variable we consider as the treatment (also often referred to as "exposure" in the DAG literature) of interest (see Table 2 for the summary). We focus on the total effect of each variable here, as so do Fearon and Laitin (2003). For simplicity, we also assume there is no measurement errors or missing data (which could complicate the selection of covariates further). To estimate the total effect, we generally need to (1) close all backdoor paths, i.e., the path of a confounding effect, and (2) leave all post-treatment variables unadjusted for so that, between the treatment and the outcome, no mediating path of

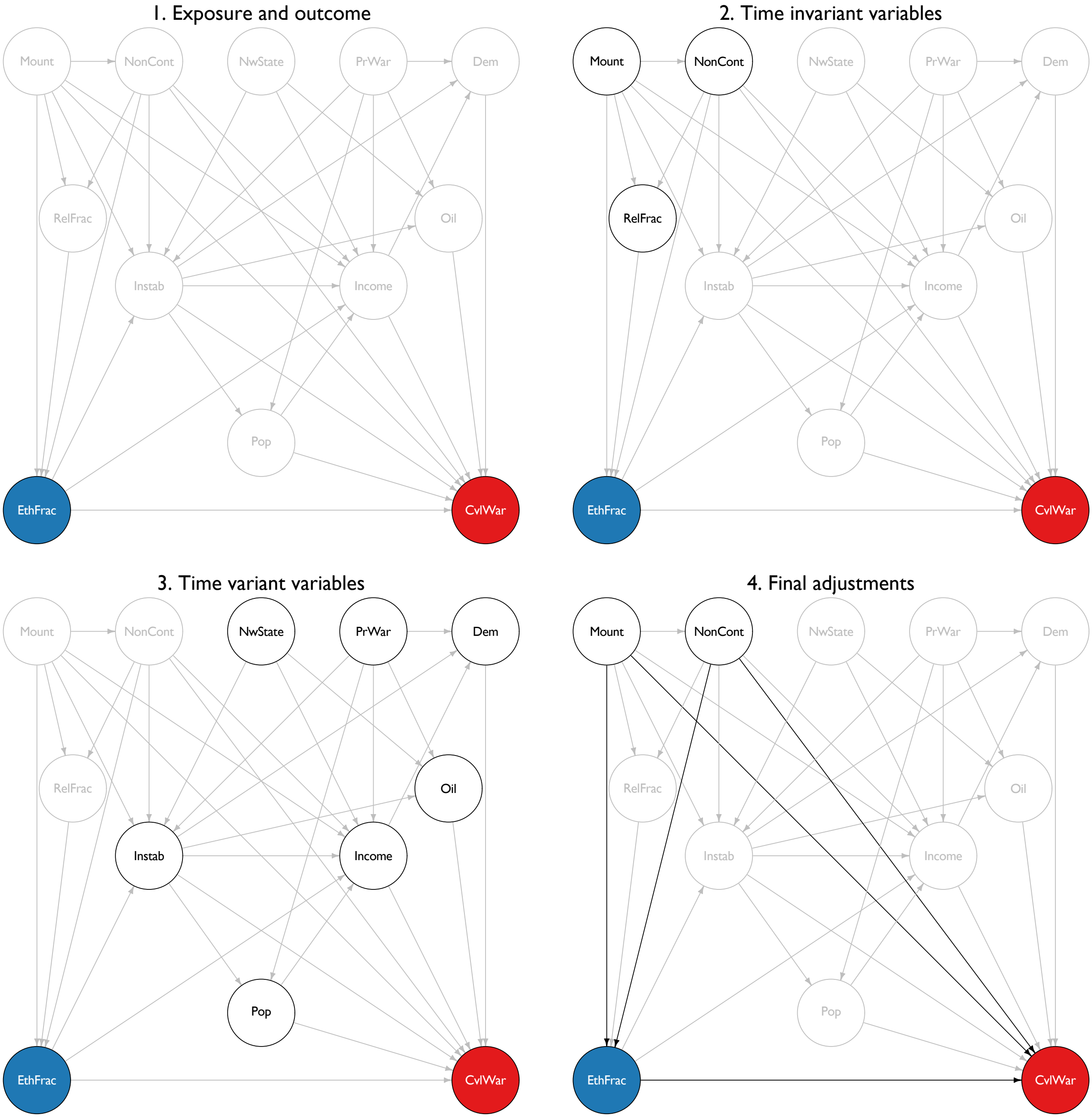

Figure 1: Full DAG representing model 1 in Fearon and Laitin (2003). Mount: Mountainous areas, NonCont: Noncontiguous states, RelFrac: Religious fractionalization, EthFrac: Ethnic fractionalization, Instab: Political instability, NwState: New states, Pop: Population size, Income: Per capita income, PrWar: Prior civil war, Oil: Oil exporters, Dem: Democracy.

its indirect effect is closed and no non-causal collider path is opened up.[5]

Those indicated by the ✓ mark in Table 2 are good controls and compose an adjustment set to identify the total effect of each independent variable. Some variables have more than one possible adjustment set; for simplicity, we indicate the minimal set, i.e., the one that uses the smallest number of covariates. For example, to estimate the total effect of ethnic fractionalization on the likelihood of civil war onsets (as indicated by the blue and red coloring in Figure 1), we control for mountainous areas and non-contiguous states as the minimal adjustment set (see the right bottom panel of Figure 1).

Those indicated by the (✓) mark are covariates that do not harm the identification of the causal effect (provided that they do not bring measurement issues) and can increase the precision of the point estimate (but with greater variance as known as the bias-variance trade-off). For example, prior war does not constitute a backdoor path for ethnic fractionalization, but is a pre-treatment covariate.

Where there are more than one adjustment set, the redundant covariates that otherwise would be good controls are indicated by the (✓) mark as well. For example, provided that democracy is the treatment variable of interest, the population constitutes the backdoor path for democracy ($CvlWar \leftarrow Pop \rightarrow Income \rightarrow Dem$), but the minimal adjustment set includes per capita income so that this backdoor path is blocked anyway and, therefore, population is unnecessary to control for.

Those indicated by the ✗ mark are post-treatment variables and, therefore, must not be controlled for to estimate the total effect. For example, to estimate the total effect of ethnic fractionalization, per capita income must not be adjusted for, as it lies on the mediating path, $EthFrac \rightarrow Income \rightarrow CvlWar$.

Finally, the one indicated by the (✗) mark is a special case of bad controls: an instrumental variable used as a control (and not as an instrument in an instrumental-variable model). Its adjustment is harmless unless there is a confouder that is left uncontrolled for, in which case the adjustment amplifies this confounding bias (Middleton et al., 2016).

Given the current DAG, the original model in Fearon and Laitin (2003) including all these

[5] For another, less oft-used criterion called the front-door criterion, see Pearl, Glymour and Jewell (2016, 66-69).

independent variables together, estimates only the total effect of democracy and oil without bias (i.e., the rows that have no ✗ in Table 2).

Treatment	Covariates										
of interest	PrWar	Income	Pop	Mount	NonCont	Oil	NwState	Instab	Dem	EthFrac	RelFrac
PrWar	-	✗	✗	(✓)	(✓)	✗	(✓)	✗	✗	(✓)	(✓)
Income	✓	-	✓	✓	✓	(✓)	✓	✓	✗	✓	(✓)
Pop	✓	✗	-	(✓)	(✓)	(✓)	(✓)	✓	✗	(✓)	(✓)
Mount	(✓)	✗	✗	-	✗	✗	(✓)	✗	✗	✗	✗
NonCont	(✓)	✗	✗	✓	-	✗	(✓)	✗	✗	✗	✗
Oil	✓	(✓)	(✓)	(✓)	(✓)	-	✓	✓	(✓)	(✓)	(✓)
NwState	(✓)	✗	✗	(✓)	(✓)	✗	-	✗	✗	(✓)	(✓)
Instab	✓	✗	✗	✓	✓	✗	✓	-	✗	✓	(✓)
Dem	✓	✓	(✓)	(✓)	(✓)	(✓)	(✓)	✓	-	(✓)	(✓)
EthFrac	(✓)	✗	✗	✓	✓	✗	(✓)	✗	✗	-	(✗)
RelFrac	(✓)	✗	✗	✓	✓	✗	(✓)	✗	✗	✗	-

Table 2: Covariates as good and bad controls for the estimation of the total effect of each independent variable. ✓ denotes good controls; a set of the covariates marked with it is the minimal adjustment set. (✓) denotes the pre-treatment covariates that are unnecessary to adjust for once all covariates marked with ✓ are controlled for, but whose adjustment does not bias the estimation of the total effect either. ✗ denotes bad controls. (✗) denotes instrumental variables, whose inclusion could amplify uncontrolled confounding bias.

2.2 Empirical illustration

The results of our DAG-based empirical models are presented in Figure 2, together with the estimates according to Model 1 in Fearon and Laitin (2003) for comparison. Note that Fearon and Laitin's estimates are from their single Model 1, while our DAG-based estimates are from separate statistical models for each treatment variable, as the DAG dictates which variables to be included or excluded as controls to identify the total effect of the variable of interest (recall Table 2). We use logistic regression and compare the average marginal effect (AME). In our case, the AME represents the total effect for all variables as we do not control for post-treatment variables. In the case of the original model in Fearon and Laitin (2003), the AME represents either the total effect or the direct effect[6] depending on which variable rests on where in the DAG, as their model includes all independent variables into a single regression model.

In Figure 2, most of the variables exhibit similar results. One noticeable difference is that in the original model, prior war has a statistically significant negative effect, while in the DAG-based model, the direction of its effect is uncertain. In other words, if we do not remove its indirect effects via the mediators of per capita income, the level of democracy, the population, and political instability, the total effect of prior war is ambivalent. This seems to make sense, as war can leave both kinds of legacy, one that can increase the probability of another civil war onset (e.g., by weakening the state capacities further) and the other that can decrease it (e.g., by exhausting the resources for fighting).

More importantly, unlike the original conclusion reached in Fearon and Laitin (2003), ethnic fractionalization is statistically significant. It also has as large an effect as mountainous areas, if we standardize their AMEs by computing the difference between the AME times the first quantile value and the AME times the third quantile value. Meanwhile, the treatment effect estimated based on the original model in Fearon and Laitin (2003) is, indeed as these authors argued, is statistically and substantively insignificant compared to that of mountainous areas. Our DAG suggests that the discrepancy between our result and their result is due to their model including mediators such as per capita income, thus making the model estimating only the direct

[6] More specifically, it is the average controlled direct effect that is, rather than computed at specific values of the mediators (Acharya, Blackwell and Sen, 2016), averaged over the empirical joint distribution of the mediators. For the different types of direct effect, see Pearl, Glymour and Jewell (2016, 121-124).

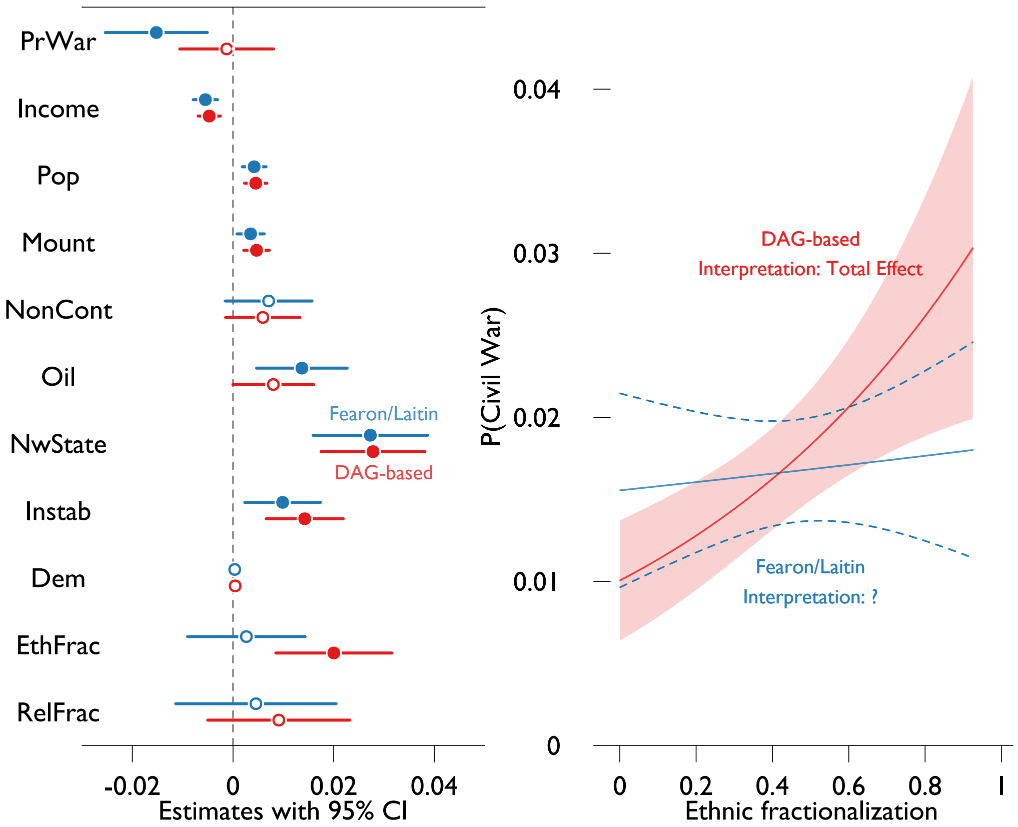

Figure 2: Results

effect of ethnic fractionalization rather than its total effect.[7] Although the direct effect might be the quantity of interest in some cases (see Acharya, Blackwell and Sen, 2016), the insignificance of the direct effect does not mean ethnic fractionalization has no significant effect *in general*. Indeed, the DAG-based model indicate it has as significant a total effect as mountainous areas, the factor Fearon and Laitin (2003) argue is important. Without the DAG, it remains unclear exactly what effect a statistical model is speaking of, thus making discussion over the "significance" of the effect muddled. In short, DAGs can make significant differences in substantive conclusions we draw from statistical models because they allow us to interpret our estimates.

2.3 Implications

We can draw three implications here. First, as we add more variables, a causal structure quickly becomes more complex. A DAG enables us to keep track of how variables influence one another and to control our controls to causally identity our treatment.

[7] Note that Fearon and Laitin (2003) did not show the estimate of ethnic fractionalization as the direct effect. Thus, we indicate this ambiguity as "?" on the right panel of Figure 2.

Second, adding new variables in a statistical model can transform the meaning of statistical estimates. We thus need to be careful with the *causal interpretation* of statistical estimates. In the example above, political instability and per capita income are on the path from ethnic fractionalization to civil war. If we do not control for the mediators (per capita income, population, oil, political instability, and democracy), we are estimating the total effect. If we control for them, we are estimating the direct effect. Relatedly, while robustness checks often include running more models with more covariates, the analyst must ensure that including more covariates does not change the type of the effect of the treatment.

Third, a DAG tells us that not all variables in a single statistical model can be interpreted as total effects (Westreich and Greenland, 2013; Keele, Stevenson and Elwert, 2020). For instance, if we control for a counfounder for ethnic fractionalization such as mountainous areas, we cannot say that the regression coefficient for mountainous areas represents its total effect because the effect of the confounder, by definition, goes through the treatment (i.e., ethnic fractionalization is a mediator in terms of mountainous areas). If we wanted to estimate the total causal effect of mountainous areas on the probability of civil war onsets, we would need a different statistical model (i.e., excluding all post-treatment variables from the perspective of mountainous areas being the treatment) and, possibly, a different DAG as well (if the confounders for mountainous areas had not been contemplated when one was theorizing the causal structure mainly from the perspective of ethnic fractionalization).

3 Promoting the Transparency of Causal Assumptions

In this section, we show how using DAGs increases the transparency of the assumptions on causal structure and helps identify inconsistent causal assumptions if any. We keep using Fearon and Laitin (2003) and focusing mainly on ethnic fractionalization as the treatment of interest for exposition. They discuss very little how the eleven factors (prior civil war, per capita income, population, mountainous areas, noncontiguous states, oil exporters, new states, political instability, democracy, ethnic fractionalization, and religious fractionalization) might be causally related to one another, i.e., the entire causal structure among the variables. We develop total four DAGs to reflect the possible causal structure that Fearon and Laitin (2003) assume. The following subsections explain the rationale for each DAG in turn.

3.1 First implicit DAG

We draw a first implicit DAG following their hypotheses, where each independent variables has the effect on the outcome only (DAG 1). It is unlikely to be what the authors had in mind when conducting their analysis. We discuss it here for exposition. DAG 1 implies complete independence among all causal factors. For example, it assumes that ethnically diverse and ethnically homogeneous countries are identical in all other respects that might influence the onset of civil war. The implication for analysis would then be that one would only need to look at the share of civil wars in ethnically diverse and ethnically homogeneous countries, and count the difference without controlling for other factors. Clearly, this is not a reasonable assumption; Fearon and Laitin (2003) would agree, since they control for several variables. Yet, it is unclear why Fearon and Laitin (2003) control for these variables, in absence of a discussion on their relationships not only with the outcome but also with the explanatory variables. In absence of this discussion we cannot know whether Fearon and Laitin (2003) consider these factors to be confounders (and if so, why), or are controlling for them to gain statistical precision.

3.2 Second implicit DAG

While there is no explicit discussion of the causal relationship between independent variables, Fearon and Laitin (2003) do make several notes about how some independent variables might impact others. We therefore build a second DAG encoding these assumptions.

The authors make the case that per capita income, once “controlled for,” takes away the effect of ethnic fractionalization (among others). No explicit argument is made for how per capita income might be causally related to it. This is problematic since per capita income should be controlled for only if it is a confounder; it should not if it is a mediator (unless the effect of interest is the direct effect, which is not what Fearon and Laitin (2003) suggest) or a collider (i.e., the common outcome of a treatment and either an outcome or a covariate that also affects the outcome). Simply showing a coefficient changes when adding a variable is insufficient to

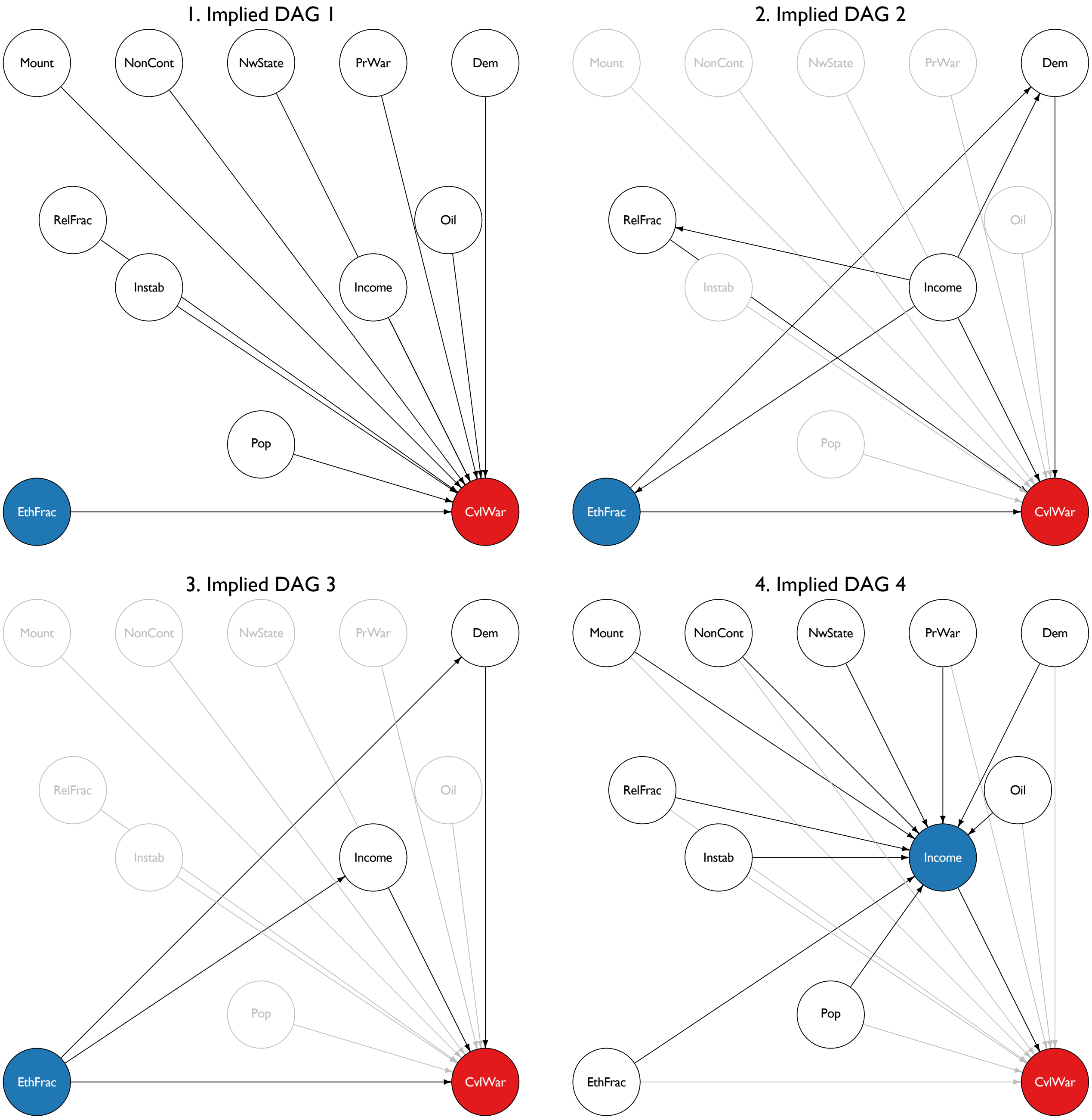

Figure 3: Four Possible DAGs Implied by Fearon and Laitin (2003)

categorize that variable as a confounder. It might be because of post-treatment bias.

The post-treatment bias could come in two forms in the current context. First, per capita income might be a collider. For example, in the fourth implied DAG we discuss later, per capita income is affected by both democracy and ethnic fractionalization, who are themselves not causally related. Controlling for income while attempting to estimate the effect of ethnic fractionalization then introduces bias from democracy by opening up the backdoor path (which could be closed again if democracy were also controlled for, which Fearon and Laitin (2003) do). Second, per capita income might be a mediator on the causal path from ethnic fractionalization to civil war. Adjusting for per capita income would therefore incorrectly estimate the size of the total effect of ethnic fractionalization, as can be seen in Implied DAG 3. In fact, since ethnic composition is (almost) time-invariant and income is time-variant, DAGs 3 and 4 are more plausible. We turn back to DAGs 3 and 4 later. For now, in building DAG 2, we assume, as the authors do implicitly, that per capita income is a pre-treatment factor and that it is a confounder.

The authors also expect that opportunities for insurgency, measured by per capita income, mountainous areas, the population size, newly independent states, political instability, regime types, non-contiguous states, and oil exporters, are *independent* of cultural differences (79–81). We can incorporate this assumption into our second DAG by omitting any edges from these factors to ethnic and religious fractionalization. However, we have already noted that per capita income should be causally related to ethnic fractionalization since the authors imply this by pointing out that the effect of ethnic fractionalization disappears once per capita income is controlled for. We are therefore faced with a contradiction of specifying relationships between these variables. Having a DAG that encodes all of these causal assumptions can prevent these kinds of theoretical contradiction. For now, we leave the edge going from per capita income to ethnic fractionalization.

Finally, Fearon and Laitin (2003) imply ethnic grievances are a mechanism of ethnic fractionalization. The effect of grievances is proxied through democracy. Therefore, we include it on the pathway from ethnic fractionalization to civil war onsets.

DAG 2 shows all of these assumptions explicitly. To estimate the effect of ethnic fractionalization, since we assume per capita income is a confounder we need to control for it to close the backdoor path. However, notice that we do not need to control for other variables, unlike Fearon and Laitin (2003) do. To estimate the total effect of ethnic fractionalization, controlling for democracy would actually be a **bad control** since democracy, according to Fearon and Laitin (2003), is on the causal path from ethnic fractionalization to civil war. Controlling for it would produce not the total effect but the direct effect of ethnic fractionalization.

As a side note, DAG 2 also shows if we wanted to estimate the effect of per capita income on civil war, we would need to control for nothing. As the DAG shows us, this effect of per capita income would be a total effect. If instead we control for democracy, ethnic fractionalization, and religious fractionalization, which are all mediators for per capita income, we would rather obtain a direct effect. DAGs can therefore show us not only what we need to control for, but also what types of effects we are estimating.

3.3 Third implicit DAG

Fearon and Laitin (2003) also state that ethnic fractionalization could cause civil war indirectly by affecting per capita income and the strength of the state (82). This would imply per capita income is a mediator for the effect of ethnic fractionalization. This contradicts the two assumptions mentioned in the previous subsection about per capita income, as a confounder and as an unrelated factor, to ethnic fractionalization. More importantly, it has profound implications for the analysis, namely that per capita income should not be adjusted for to obtain an unbiased estimate of the total effect of ethnic fractionalization.

DAG 3 shows the (contradictory) assumption of per capita income as a mediator for ethnic fractionalization. Note the different implications of the three DAGs. In DAG 1, we need not adjust for per capita income to obtain an unbiased estimate for ethnic fractionalization, while adjusting for it does not introduce bias either. In DAG 2, we need to adjust for per capita income since it is a confounder. In DAG 3, we should not adjust for per capita income since it lies on the causal path from ethnic fractionalization to the outcome.

3.4 Fourth implicit DAG

We have so far drawn the assumptions of the causal structure in Fearon and Laitin (2003) by looking at their theoretical arguments. Another way is to look at their statistical model and consider its implicit assumptions.

In Model 1 of Table 1 (84), they include all variables. There are many ways to interpret this choice. The authors do not clearly specify what their causal factor of interest is for that model. Instead, they seem to interpret the regression coefficient of each independent variable as the total effect. This is problematic, since as we have seen, each independent variable, when to be interpreted as a causal factor, can have its own associated confounders, mediators, and colliders. Which variable(s) are confounders, mediators, colliders, or none of these, differ for each independent variable. Therefore, separate statistical models are necessary to interpret each independent variable as a causal factor (as done in the section "Empirical illustration"). Fearon and Laitin (2003) discuss how all independent variables might cause civil war, but not how they causally relate to each other. In other words, they do not have an explicit causal structure among the variables used to guide the causal interpretation of each variable.

We therefore cannot draw a single coherent DAG based on Model 1, in absence of additional causal assumptions. We have already noted there are three conflicting assumptions about per capita income and ethnic fractionalization – that per capita income is a confounder for ethnic fractionalization, that per capita income is unrelated to ethnic fractionalization, and that per capita income is on the causal pathway between ethnic fractionalization and civil war onsets. For example, if we took the third assumption, the assumed causal structure could be something like DAG 4, assuming that all other factors are confounders for per capita income since they are adjusted for in the empirical model.[8] However, since the authors do not state which of the three conflicting assumptions should be used to interpret Model 1, it remains unclear whether DAG 4 is indeed what the authors had in their mind.

[8] It might be the case that Fearon and Laitin (2003) added these factors even if they did not believe them to be confounders, since they could have in principle increased precision under the causal assumption such as DAG 1. But as we have discussed, the causal structure such as DAG 1 is highly unrealistic.

3.5 Implications

We draw three implications from the discussion in this section. First, without an explicit causal structure, it is unclear why some factors are controlled for while other are not. DAGs make these explicit and make the process of deciding which factors to adjust for more transparent. Additionally, DAGs encourage authors to consider relationships between their variables they might otherwise not consider.

Second, it becomes difficult for readers to identify the causal assumptions authors are making. We have shown that several mutually exclusive causal structures can be implied by the theory and statistical models of Fearon and Laitin (2003). We argue that particularly when presenting results, DAGs allow audiences to immediately see which assumptions the researchers are making. In return, they can also facilitate discussion since criticisms around models using observational data often revolve around which factors the authors did and did not control for. Using a DAG can allow authors to better justify and explain their selection of control variables.

Third, without formalizing assumptions, it is easy to run into contradiction, without knowing it and allowing others to correct you. From this point of view, DAGs improve self-correction mechanisms in (political) science by formalizing and visualizing a theory, allowing it to be criticized by audiences (Ioannidis, 2012; Alberts et al., 2015).

4 Discussion

4.1 DAGs are necessary even if all covariates are pre-treatment

In theory, if covariates are pre-treatment (but not instrument variables), it should be safe to include all of them (Imbens, 2020). However, in practice, this is not always the case. First, if only a subset of pre-treatment covariates are observable, then blindly including all of them might amplify rather than reduce the bias, depending on the direction of the bias induced by each unobservable, nonadjustable confounder (Steiner and Kim, 2016). A DAG can encourage us to contemplate how each confounder, observed or unobserved, biases the treatment effect and what selection of pre-treatment covariates is most likely to reduce the bias overall, by the visualization of the entire causal structure. Second, not all (pre-treatment) covariates are measured with the same degree of quality. Some may suffer from missing values, which is often difficult to address and can introduce bias (Lall, 2016; Shpitser, Mohan and Pearl, 2015; Montgomery, Nyhan and Torres, 2018). Some may be measured with greater errors, which also biases the estimate (Kuroki and Pearl, 2014). Both missing data and measurement errors can be expressed via a DAG as a function of some (unobserved) variable (Montgomery, Nyhan and Torres, 2018; Kuroki and Pearl, 2014). Our point is that the visualization of the causal structure via DAGs helps make these issues explicit and transparent, to both the analyst and the reader.

4.2 DAGs are useful even if one wants to avoid making explicit causal claims

Empirical studies, especially those using regression, often avoid making an explicit causal argument. Instead, a regression coefficient is interpreted as an association. However, they also control for several covariates. Then, it means that they look at the predictive power of a certain independent variable while the data are stratified by these covariates. It is left unclear why such stratification is meaningful to examine the predictive power of one independent variable, rather than simply looking at the bivariate correlation between the predictor of interest and the outcome, or computing the predictive power of the overall model only (such as the ROC curve). If a study controls for covariates, avoids making a causal interpretation, and does not explain why the stratification is meaningful, such empirical research is neither predictive nor causal.

It might be argued that one controls for covariates to see which hypotheses *win* or address a spurious correlation. But such an argument inherently assumes some sort of causal structure. The predictive power of one independent variable may be reduced when another is controlled for, because they are causally related (e.g., one being a mediator or confounder of the other). A "spurious" correlation is actually used to mean spurious causation, i.e., interpreting a non-causal correlation as causation. A correlation is a statistical measure, while whether a correlation can be interpreted as causation requires causal assumptions (Pearl and Mackenzie, 2018).

4.3 DAGs are also relevant for (quasi-)experiments

DAGs are important even for experiments and quasi-experiments for two reasons. First, the random assignment of treatments cannot remedy bias caused by the non-random selection of units into a sample, when the non-random selection is a function of both treatments (e.g., participants more likely to drop out from the experiment after receiving one type of treatment than whey they receive the other types) and some covariate that also affects the outcome (Montgomery, Nyhan and Torres, 2018). In other words, just because the random assignment of treatments is a credible assumption in itself, it does not guarantee that the average treatment effect is unbiased, unless the assumption is made that the selection into a sample is independent of the treatments. This assumption is empirically unverifiable as so is selection on observations. Even if the rate of attrition or missing outcome values is the same across all treatment statuses, it does not guarantee that the same kind of participants selected into the sample; it is possible that different kinds of participants (i.e., having different distributions of the covariate) *happen to* have the same propensity to select into the sample. DAGs can visualize these assumptions and make them transparent. Indeed, Montgomery, Nyhan and Torres (2018, 763–764n5) point out in developing the explanation of post-treatment bias in experiments that the DAG "is often especially helpful in clarifying which research designs can accurately recover causal estimates."

The other reason why DAGs are useful even for (quasi-)experiments is that it compels us to consider whether the balance on observed covariates is enough to claim that randomization has worked to produce an estimate close to the true parameter value (i.e., the matter of precision rather than unbiasedness). If one could think of an unobserved confounder, claiming that the balance indicates successful randomization in the above respect would equal assuming that the observed covariates are a proxy for (i.e., cause or are caused by) the unobserved confounder. If the unobserved confounder were causally independent from any of the observed covariates, the balance on the latter would say nothing about the balance on the former (Deaton and Cartwright, 2018). DAGs are useful to clarify what assumption the experimenter makes with respect to the causal relationship between the unobserved confounder and the observed covariates.

4.4 Multiple DAGs improve the credibility of one's causal assumptions

An attempt to identify the causal structure usually results in the situation where one cannot be sure whether one variable causes another or not. In such cases, we can construct multiple DAGs to express our uncertainty over the exact form of the causal structure. We can then develop empirical models each of which reflects each of the plausible DAGs. The results from all models can then be averaged with weights if necessary, as done in the Bayesian model averaging (Montgomery and Nyhan, 2010).

5 Conclusion

This article proposed we use DAGs to articulate the connection between a theory and empirics (Hall, 2003) and increase the transparency of the causal assumptions. Throughout the application to Fearon and Laitin (2003), we substantiated these proposals. For instance, the replication of Fearon and Laitin (2003) showed how a DAG can make a substantive difference in statistical estimates for causal inference. Unlike in the original article, the effect of ethnic fractionalization turned out significant, if it was estimated by a DAG-based statistical model. When we tried to construct a DAG from Fearon and Laitin (2003), it appeared that, while the study suggested per capita income has three possible causal relationships, it did not settle on one and, therefore, included the inconsistent causal assumptions.

Of course, DAGs are no panacea. It is a way of representing our assumptions of the causal structure among variables and deducing a viable empirical strategy for causal identification. If the DAG is incorrect, the estimates are not reliable. But this applies to any kind of causal models or assumptions, regardless of using a DAG or not. In this sense, DAGs are more useful to clarify our assumptions and make them explicit and criticizable. After all, the true causal structure is never really known. Yet, DAGs offer clarity on how to think in a theoretically driven way and have a empirical strategy consistent with the theorized causal structure. In short, DAGs help strengthen, and make transparent, our theories on causal structure and the connection between these theories and empirical models, thereby bringing controls under control. We recommend that empirical studies include DAGs as the explicit expression of their causal assumptions.

References

Acharya, Avidit, Matthew Blackwell and Maya Sen. 2016. "Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects." *American Political Science Review* 110(3):512–529.

Achen, Christopher H. 2005. "Let's Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong." *Conflict Management and Peace Science* 22(4):327–339.

Alberts, Bruce, Ralph J. Cicerone, Stephen E. Fienberg, Alexander Kamb, Marcia McNutt, Robert M. Nerem, Randy Schekman, Richard Shiffrin, Victoria Stodden, Subra Suresh, Maria T. Zuber, Barbara Kline Pope and Kathleen Hall Jamieson. 2015. "Self-correction in science at work." *Science* 348(6242):1420–1422.

Breton, Charles. 2015. "Most cited articles in political science by decade.".
URL: *http://charlesbreton.ca/most-cited-ps/*

Buhaug, Halvard, Lars-Erik Cederman and Jan Ketil Rød. 2008. "Disaggregating Ethno-Nationalist Civil Wars: A Dyadic Test of Exclusion Theory." 62(03):531–551.

Cederman, Lars-Erik and Luc Girardin. 2007. "Beyond Fractionalization: Mapping Ethnicity onto Nationalist Insurgencies." 101(01):173–85.

Cinelli, Carlos, Andrew Forney and Judea Pearl. 2020. "A Crash Course in Good and Bad Controls." *SSRN Electronic Journal* .

Deaton, Angus and Nancy Cartwright. 2018. "Understanding and misunderstanding randomized controlled trials." *Social Science & Medicine* 210(August):2–21.

Fearon, James D. and David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97(1):75–90.

Hall, Peter A. 2003. Aligning ontology and methodology in comparative politics. In *Comparative Historical Analysis in the Social Sciences*, ed. James Mahoney and Dietrich Rueschemeyer. New York: Cambridge University Press pp. 373–406.

Hernán, Miguel A. and James M. Robins. 2020. *Causal Inference: What If*. Boca Raton, FL: CRC Press.

Hug, Simon. 2010. "The Effect of Misclassifications in Probit Models: Monte Carlo Simulations and Applications." 18(1):78–102.

Imbens, Guido W. 2020. "Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics." *Journal of Economic Literature* 58(4):1129–1179.

Ioannidis, John P. A. 2012. "Why Science Is Not Necessarily Self-Correcting." *Perspectives on Psychological Science* 7(6):645–654.

Keele, Luke, Randolph T. Stevenson and Felix Elwert. 2020. "The causal interpretation of estimated associations in regression models." *Political Science Research and Methods* 8(1):1–13.

Kuroki, Manabu and Judea Pearl. 2014. "Measurement bias and effect restoration in causal inference." *Biometrika* 101(2):423–437.

Lall, Ranjit. 2016. "How Multiple Imputation Makes a Difference." *Political Analysis* 24(4):414–433.

Lundberg, Ian, Rebecca Johnson and Brandon M. Stewart. 2021. "What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory." *American Sociological Review* 86(3):532–565.

Middleton, Joel A., Marc A. Scott, Ronli Diakow and Jennifer L. Hill. 2016. "Bias Amplification and Bias Unmasking." *Political Analysis* 24(3):307–323.

Montgomery, Jacob M. and Brendan Nyhan. 2010. "Bayesian Model Averaging: Theoretical Developments and Practical Applications." 18(2):245–270.

Montgomery, Jacob M., Brendan Nyhan and Michelle Torres. 2018. "How Conditioning on Post-treatment Variables Can Ruin Your Experiment and What to Do about It." *American Journal of Political Science* 62(3):760–775.

Morgan, Stephen L. and Christopher Winship. 2015. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.

Muchlinski, David, David Siroky, Jingrui He and Matthew Kocher. 2016. "Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data." *Political Analysis* 24(1):87–103.

Pearce, Neil and Debbie A Lawlor. 2016. "Causal inference—so much more than statistics." *International Journal of Epidemiology* 45(6):1895–1903.

Pearl, Judea and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books. 2018.

Pearl, Judea, Madelyn Glymour and Nicholas P. Jewell. 2016. *Causal Inference in Statistics: A Primer*. Chichester, West Sussex: John Wiley & Sons. 2016.

Shpitser, Ilya, Karthika Mohan and Judea Pearl. 2015. Missing Data as a Causal and Probabilistic Problem. In *UAI'15: Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, ed. Marina Meila, Tom Heskes. AUAI Press pp. 802–811.

Steiner, Peter M. and Yongnam Kim. 2016. "The Mechanics of Omitted Variable Bias: Bias Amplification and Cancellation of Offsetting Biases." *Journal of Causal Inference* 4(2):1–22.

Westreich, Daniel and Sander Greenland. 2013. "The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients." *American Journal of Epidemiology* 177(4):292–298.