How Can We Estimate the Effectiveness of Institutions? Solving the Post-Treatment versus Omitted Variable Bias Dilemma*

Michaël Aklin† Patrick Bayer‡

September 30, 2016

Abstract

How effective domestic and international institutions are is one of the central questions that international relations and comparative scholars ask. It however also happens to be one of the most difficult questions to answer. Post-treatment confounding variables create a dilemma for applied researchers who want to estimate the effectiveness of these institutions. Including these variables in regression models generates post-treatment bias. Excluding them generates omitted variable bias. Traditional estimation strategies can therefore not provide unbiased estimates in the presence of such variables. This has significant implications from a research perspective, because it also means that traditional regression methods offer biased estimates of the direct and the indirect effects of an intervention, even though these are important to test competing theories about what makes institutions effective or not. In this paper, we introduce an estimation algorithm that can recover unbiased estimates of the direct, indirect, and total effect of an intervention. This approach is particularly useful in contexts in which the treatment variable, such as IO membership, has a diffuse effect, as is often the case in international and comparative political economy. We show the usefulness of our algorithm by re-exploring the effect of democracy on FDI.

Keywords: institutions; IO effectiveness; post-treatment bias; omitted variable bias; direct effects; indirect effects.

*We are grateful to Emine Deniz, Jude Hays, Marc Meredith, Matto Mildenberger, and David Steinberg for useful comments on earlier versions of this paper. We also thank audience members at the 16th Annual Policy Conference at the University of Pittsburgh and at ISA for suggestions. All errors are ours.

†Department of Political Science. University of Pittsburgh. 4814 WWPH, 230 S. Bouquet Street, Pittsburgh, PA 15260, USA. Email: aklin@pitt.edu.

‡School of Social and Political Sciences. University of Glasgow. Adam Smith Building, Bute Gardens, Glasgow, G12 8RT, Scotland, UK. Email: patrick.bayer@glasgow.ac.uk

1 Introduction

As social scientists, we devote considerable efforts to estimate the effect of an intervention on an outcome of interest.[1] We could for example be interested in the question of whether the World Trade Organization (WTO) does indeed increase trade (Rose, 2004; Tomz, Goldstein, and Rivers, 2007; Goldstein, Rivers, and Tomz, 2007). To answer this question, we typically define a statistical model and estimate the parameter of the intervention, e.g., IO membership, with an appropriate method, such as least squares. Among many threats, we are particularly worried about omitted variable bias. As a result, the search for sound identification strategies, whether by relying on experiments or quasi-experiments, has instilled valuable creativity in political science (Samii, 2016). For many reasons, however, not all interesting research hypotheses can be tested through these means and we need to rely on observational data. In this case, because the intervention has not been randomly assigned, control variables are included to our statistical model to ensure that the main estimate does not accidentally capture confounding factors.

This approach often leads to misleading findings. The first issue, which we leave aside in this paper, is that knowing the correct model is not possible. The second, less well-known problem, is that even if we knew the true model, it may be impossible to recover unbiased estimates with traditional estimation methods because of the causal sequence of events. When a variable mediates our intervention but it also has an effect of its own, then the statistical model can neither omit it (since it is a confounder) nor can it include it (since it is post-treatment, that is, causally following our intervention). To clarify what we mean, consider the following example. Suppose that we wish to measure the effect of democracy on foreign direct investments (FDI). Democracy could increase the amount of FDI directly because democracies are less likely to expropriate investors. But democracies are also more likely to implement freer capital movement regulation. The statistical model will therefore contain FDI on the left-hand side, and democracy on the right-hand side. However, what to do about capital openness regulation? Should we control for it or not?

Herein lies the problem. If we do not control for it, our estimates are likely to suffer from omitted

[1] We use *interventions*, instead of alternatives such as *treatments*, to denote the independent variable whose effect we wish to study.

variable bias. But if we do, then the estimates will suffer from post-treatment bias. Post-treatment bias refers to the bias generated by including control variables causally affected by the intervention itself. The issue is ubiquitous in political science across subfields (Acharya, Blackwell, and Sen, forthcoming). When researchers ask whether the WTO (or GATT) affects trade, they generally control for income. But income is likely to be measured post-treatment. The problem is particularly salient in international relations, where the effects of international treaties and organizations tend to be very diffuse and operate through multiple channels. But the issue is also common at the micro level. How to deal with post-treatment confounders remains, in King's (2010) words, one of the "hard questions" in the social sciences.

This, in turn, sheds light on an even bigger issue: the need to distinguish between the *direct*, the *indirect*, and the *total effect* of an intervention (Imai et al., 2011; Acharya, Blackwell, and Sen, forthcoming). The direct effect of a intervention, or DEI, refers to the unmediated and unmodeled consequences of an intervention. These are often the estimates that are reported in empirical studies and they are typically associated with regression coefficients of the intervention indicator. But an intervention often operates through a (post-treatment) mediator (Gerber and Green, 2012; Imai et al., 2011). We call this the *indirect* effect of the intervention, or IEI. The *total* effect of an intervention (TEI), then, is the sum of the DEI and IEI. Distinguishing between these quantities is desirable because it helps us assess the validity of competing theories.

In summary, our problem is the following. We would like to be able to get reliable estimates of the direct, indirect, and total effects of an intervention. Unfortunately, even if our theory offers clear guidance as to what the (post-treatment confounding) mediator is, since it is measured post-treatment, we cannot get unbiased estimates of these quantities. We are trapped between a rock (post-treatment bias) and a hard place (omitted variable bias).

In this paper, we offer a simple, yet powerful strategy to recover estimates of direct, indirect, and total effects that performs remarkably well under relatively weak assumptions. In the spirit of Abadie, Diamond, and Hainmueller's (2010) synthetic control method, our estimates rely on untreated observations to recover unbiased estimates of post-treatment variables, and use these to reconstruct the unobserved counterfactual among treated observations. Based on these coun-

terfactuals, we can generate unbiased estimates of the intervention effect even in the presence of post-treatment confounders. We demonstrate the qualities of our approach analytically. We also use Monte Carlo simulations to show that this technique is superior to a variety of competing and often used approaches.

Our approach deviates from existing techniques in how it treats the indirect effect (Acharya, Blackwell, and Sen, forthcoming; Imai et al., 2011). Acharya, Blackwell, and Sen (forthcoming) recently offered a creative estimation strategy that allows the recovery of the direct effect. In contrast, we can derive estimates for *both* the DEI *and* the IEI, rather than treating the mediating variable as a nuisance. Acharya, Blackwell, and Sen (forthcoming) de-mediate the dependent variable to "net out" the effect of the mediator. While useful for ruling out some mediating channels, indirect effects themselves cannot be estimated. But indirect effects are important, because they are in essence the modeled (or theorized) channels through which an intervention affects an outcome. In their re-analysis of the classic Fearon and Laitin (2003) example, Acharya, Blackwell, and Sen (forthcoming) can re-assure us that ethnic fractionalization has its own effect on civil war, not mediated through political instability, but we cannot say how the effect of political instability compares to the one from ethnic fractionalization. This however is the problem we as social scientists often face when we seek to adjudicate the relative merit of competing theories.

In practice, the value of the approach we recommend in this paper depends on the quality of our theories. To tease out the direct, indirect, and total effect of an intervention, we need to have a good understanding of the data generating process, which only well specified theory can provide us with. From a normative perspective, this is a desirable feature, as any empirical technique will only be as powerful as the social science theories that underlie it it. We thus join recent voices arguing in favor of better ties between theory and empirical work (Samii, 2016), and hope that our method will help applied researchers test new predictions from their work.

2 What is the Problem?

We discuss our new approach for estimating direct, indirect, and total effects in the simplest framework possible with time-series, cross-sectional data. Despite its simplicity, this basic setup

is sufficient to illustrate the key problems pertaining to the standard regression approaches. The setup is the following:

- $Y \in \mathbb{R}$ is the *outcome* (dependent) variable.
- $D \in \{0, 1\}$ is the *intervention*. "Zero" means the absence of the intervention, and "one" that it is present in a given unit. It is not randomly assigned.
- $X \in \mathbb{R}$ is a post-treatment confounding variable. Post-treatment means that part – but not all – of X's variation is caused by variation in D. We refer to this variable as the *mediator*.

To estimate the effect of D on Y, we could regress Y on D with least squares. The problem is that ignoring X generates a biased estimate of the true effect of D. This is a classical case of omitted variable bias (Greene, 2008). The bias is caused by the fact that X is not perfectly explained by D and varies on its own.

Could we instead regress Y on D and X? No. Because X is a post-treatment variable, including it creates a different kind of bias, typically referred to as post-treatment bias (Rosenbaum, 1984). The risk of post-treatment bias is well-known and makes the estimate smaller (most of the times) than the true effect of the intervention (King, 2010; Gerber and Green, 2012).

This dilemma has no clear solution. We show below that most of the standard approaches, be they simple OLS regressions or more advanced 2-step procedures like instrument variable estimation, produce either biased estimates or can only retrieve either the direct or total effect. At its core, approaches that suffer from bias overlook the fact that including post-intervention data makes identification of the quantity of interest impossible. This is simply for the reason that post-intervention variables confound their true effect (as evident in the data pre-intervention) with the effect of the intervention on these variables. On the other hand, approaches that produce unbiased estimates pay the price of only being able to retrieve one of the direct or total effect. Our approach, in contrast, gets around the problem of post-intervention bias by only using pre-intervention data to estimate post-intervention counterfactuals.

So far, we have framed the issue as a problem of estimation. But it also raises epistemological questions. We need to think carefully about the ways in which the intervention affects the outcome.

Often, analysts are interested in the *total* effect of an intervention. This is the effect obtained 'once the dust has settled.' But sometimes, researchers want to distinguish between various channels through which the intervention may operate. The intervention could have a direct effect. For instance, when a country becomes a member of the WTO, it generally exports more because its firms face lower tariffs. But WTO membership could also have an indirect effect. WTO membership could make a country wealthier, strengthen its exchange rate, and reduce its exports. Possibly, these two effects could cancel each other, leaving the analyst with the mistaken impression that the intervention has had no effect.

Thus, not only is post-treatment bias a thorny empirical problem, but it also impedes our ability to parse out the channels through which an intervention operates. This has serious consequences for applied research. For instance, many theories predict that democracy increases income. Democratic institutions promote economic liberties, stimulate investment, provide better public services, protect property rights, and may limit governmental interference with markets, all of which contribute to economic growth and higher income. However, how and why this link prevails is still unclear and the empirical literature produces inconsistent results. Depending on the set of control variables, the direct effect of democracy often disappears (Baum and Lake, 2003). Yet, many of these variables are likely to be measured post-treatment. So, in the absence of a way to overcome post-treatment bias, we are unable to provide a satisfactory answer to these questions.

3 The Quantities of Interest

In this section, we define the direct, indirect, and total effects of an intervention. Then, we examine various ways in which we could estimate these. We show that these estimates generally suffer from either post-treatment or from omitted variable bias.

Suppose that the true data generating process is the one shown in the directed graph in Figure 1 (Pearl, 2009). This setup is similar to Acharya, Blackwell, and Sen (forthcoming, Figure 3) and Imai et al. (2011, Figure 1(a)).

As above, D is our intervention indicator which affects Y both directly and indirectly through X. The mediator X is jointly determined by D and W, which are possibly correlated, inducing

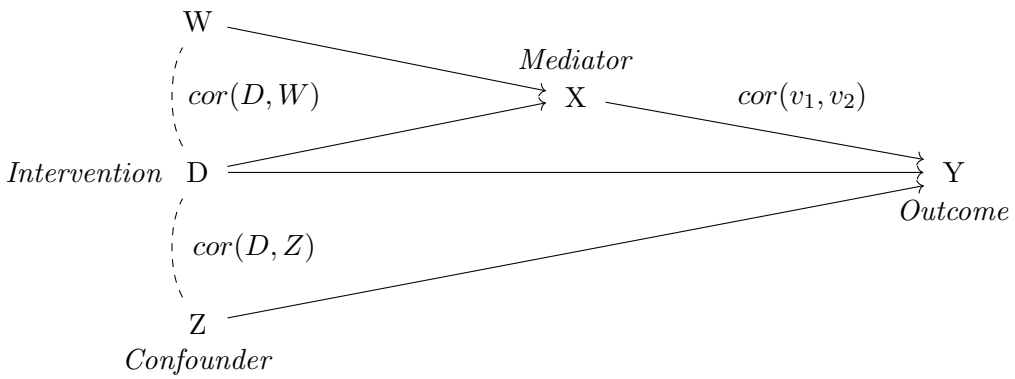


Figure 1: Directed acyclic graph showing the data generating process from our system of equations. Y is the outcome variable, D is our intervention of interest, W is a confounding variable, X is a post-treatment confounder, and Z is an exogenous control covariate. Arrows indicate causal relationships and dotted lines denote otherwise correlated variables.

correlation between D and X as well.

W can be interpreted in various ways. First, X and W could be separate variables and both are observable. Second X and W could be separate variables, but W is unobservable or unknown. For instance, W could be a latent process. Third, W could the source of X's own variation, independent of D, which thus exercises an individual effect on Y. In this case, W is distinct from X for analytical purposes, but does not have a life of its own. It is incorporated in X. As we shall see, whether W is known and observable has important implications for our own approach. If it is observable, then we are in a position to recover unbiased estimates of all quantities of interest: DEI, IEI, and TEI. In turn, this depends on theory (to identify what exactly W is) and on data (whether W can be measured). If W is unobservable, we can only obtain unbiased estimates of the DEI.

Finally, Z is a 'traditional' confounder: it is correlated with D and Y, but unrelated to any post-treatment variable. Its presence or absence is not material to our findings, but it makes the model more realistic. Overall, this setup is flexible to accommodate a range of situations commonly faced by applied researchers.

In a probabilistic context, we can rewrite this data-generating process with the following set of

two equations:

$$Y = \alpha_1 + \beta_1 D + \gamma_1 X + \delta_1 Z + v_1 \tag{1}$$

$$X = \alpha_2 + \beta_2 D + \gamma_2 W + v_2. \tag{2}$$

We include two error terms v_1 and v_2, which may or may not be correlated. The direct effect of D is β_1. But as we noted, the direct effect is not the same as the total effect. Indeed, D also affects Y through X. What, then, is the total effect of the intervention? To find this out, we can substitute X from equation (2) into equation (1), which—after rearranging terms—yields:

$$Y = [\alpha_1 + \gamma_1\alpha_2] + D[\beta_1 + \gamma_1\beta_2] + \gamma_1\gamma_2 W + \delta_1 Z + [v_1 + \gamma_1 v_2]$$

We can then compute the total effect of the intervention (TEI) as

$$\text{TEI} \equiv \frac{\partial Y}{\partial D} = \beta_1 + \gamma_1\beta_2,$$

which can be split into the *direct effect* of the intervention (DEI) and the *indirect effect* of the intervention (IEI):

$$\text{TEI} \equiv \text{DEI} + \text{IEI}, \quad \text{where DEI} \equiv \beta_1 \text{ and IEI} \equiv \gamma_1\beta_2.$$

To avoid trivial cases, we assume that β_1, β_2, and γ_1 all differ from zero. In other words, D has both a direct and an indirect effect on Y. Note that this does not preclude the total effect to be zero if $-\beta_1/\beta_2 = \gamma_1$.

4 Possible Estimands

How can we now recover unbiased estimates of the direct, indirect, and total effects? In Table 1, we report a summary of the most common approaches that we are aware of which try to estimate these quantities. The first approach, which we call the *common approach*, simply ignores the issue

of post-treatment bias, controls for X (and Z), reports the least square estimate $\hat{\beta}_1$, and interprets it as the DEI or the TEI. We call it 'common,' because in our experience, many model-based papers ignore the risk of post-treatment altogether. A second strategy is to exclude post-treatment variables altogether, an approach that we label *reduced form approach*. Implicitly or explicitly, the hope here is that $\hat{\beta}_1$ estimates the TEI by absorbing the indirect effects through X. Such an approach is used, for instance, by Pepinsky (2014, 444).

These two strategies generate a single quantity of interest, captured by $\hat{\beta}_1$. Whether it is an estimate of the total effect or only of the direct effect depends on the assumptions made by the researcher. The two will be identical as long as there are no indirect effects. If D does have indirect effects—and this is what we assume throughout in this paper—then we show below that $\hat{\beta}_1$ can at best only recover the DEI (common approach) or the TEI (reduced form approach), but not both, and each single estimate is virtually always biased.

The third and fourth approaches seek to overcome this limitation by estimating two models in two separate steps and recovering the three parameters β_1, γ_1, and β_2 which jointly allow to calculate the total effect. The two strategies differ based on whether they do control for W in the second step. We name the approach that ignores W *reduced form 2-step* and the one that includes it *comprehensive 2-step*. The former is implicitly estimated when W is unknown or unobserved.

The fifth approach relies on an instrumental variable strategy. The mediator X is instrumented with W. From the outset, we know that D cannot serve as an instrument because the exclusion restriction assumption is not met. The implication is that we cannot estimate β_2, and therefore cannot recover the TEI. As long as W truly is exogenous and that its own exclusion restriction assumption is met, we can recover the DEI. We refer to this last approach as the *instrument variable approach*

5 Downside of Existing Approaches

To assess how well these five approaches perform, we ran Monte Carlo simulations with randomly generated data. We use the data generating process from Figure 1, which in a regression framework can be characterised by the two equations (1) and (2). We set coefficients β_1 and δ_1 to 2, and

Existing Strategies

#	Model name	Specification	Estimand	Limitations
(1)	Common approach	$Y = \alpha_1 + \beta_1 D + \gamma_1 X + \delta_1 Z + v_1$	DEI ($\hat{\beta}_1$)	PTB; no TEI, IEI
(2)	Reduced form	$Y = \alpha_1 + \beta_1 D + \delta_1 Z + v_1$	TEI ($\hat{\beta}_1$)	OVB; no DEI, IEI
(3)	Reduced form 2-step approach	$Y = \alpha_1 + \beta_1 D + \gamma_1 X + \delta_1 Z + v_1$ $X = \alpha_2 + \beta_2 D + v_2$	TEI ($\hat{\beta}_1 + \hat{\gamma}_1\hat{\beta}_2$), DEI ($\hat{\beta}_1$) IEI ($\hat{\gamma}_1\hat{\beta}_2$)	PTB for DEI; PTB for TEI
(4)	Comprehensive 2-step approach	$Y = \alpha_1 + \beta_1 D + \gamma_1 X + \delta_1 Z + v_1$ $X = \alpha_2 + \beta_2 D + \gamma_2 W + v_2$	TEI ($\hat{\beta}_1 + \hat{\gamma}_1\hat{\beta}_2$), DEI ($\hat{\beta}_1$) IEI ($\hat{\gamma}_1\hat{\beta}_2$)	PTB for DEI; observable W
(5)	Instrument variable approach (2SLS)	$Y = \alpha_1 + \beta_1 D + \gamma_1 \hat{X} + \delta_1 Z + v_1$ $X = \alpha_2 + \gamma_2 W + v_2$ where $\hat{X} = \hat{\alpha}_2 + \hat{\gamma}_2 W$	DEI ($\hat{\beta}_1$)	No TEI; observable W
(6)	Average controlled direct effect (ACDE)	$\tilde{Y} = \alpha_1 + \beta_1 D + \delta_1 Z + \epsilon_1 W + v_1$ $Y = \alpha_2 + \beta_2 D + \gamma_2 X + \delta_2 Z + \epsilon_2 W + v_2$ where $\tilde{Y} = Y - \hat{Y}$ with $\hat{Y} = \hat{\gamma}_2 X$	DEI ($\hat{\beta}_1$)	PTB; no TEI
(7)	Total Effect Decomposition	See text	DEI, IEI, TEI	observable W

Table 1: Existing strategies to estimate the direct or total effects of the intervention. The true model is assumed to be the one shown in Figure 1. These approaches reflect estimated models in the existing literature. The last column indicates why they can be misleading. PTB=post-treatment bias; OVB=omitted variable bias; TEI=total effect of the intervention; DEI=direct effect of the intervention.

all other coefficients equal to 1. This means that TEI $\equiv \beta_1 + \gamma_1\beta_2 = 3$, DEI $\equiv \beta_1 = 2$, and IEI $\equiv= \gamma_1\beta_2 = 1$.

We impose a correlation structure that allows us to induce post-treatment bias and omitted variable bias independently from each other. For the former, we draw error terms v_1 and v_2 jointly from a bivariate standard normal distribution and set the post-treatment correlation coefficients

$cor(v_1, v_2)$ to four discrete values, as shown in the figures below.[2] For the latter, we control the correlation between D and X indirectly by fixing the correlation coefficients between D and W. As mentioned above, this allows X to have a life of its own, which cannot be fully explained by D. Additionally, we arbitrarily set the correlation of the traditional confounder Z with the intervention D to $cov(D, Z) = 0.3$ for all simulations, which is inconsequential for any of our results.

We then simulate 1,000 data sets for various correlation coefficients.[3] Our data represents a reasonably sized and perfectly balanced time-series, cross-sectional data set with 100 units (e.g., countries) over 20 time periods (e.g., years), for a total of 2,000 observations. The intervention occurs in time period $t = 10$, so that our intervention indicator D is zero for all units before this point in time and equal to 1 after. For each run, we then simply record the point estimates of the total, direct, and indirect effect estimates (if possible) for each approach and graph out their distributions as shown in the plots below.

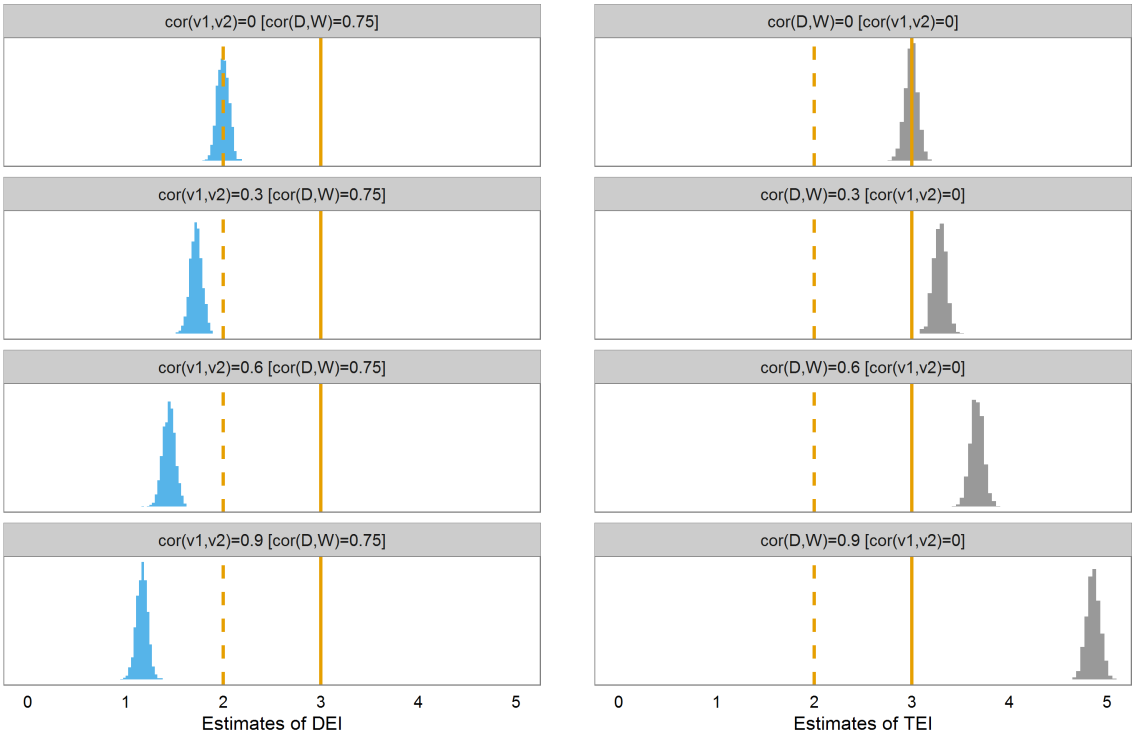
The results are reported in Figures 2-4. The thick yellow line represents the TEI and the dashed yellow line shows the estimate of the DEI. Remember that the common, 2SLS IV, and ACDE approaches provide estimates of the direct effect, $\hat{\beta}_1$, the reduced form approach retrieves an estimate of the total effect, $\hat{\beta}_1 + \hat{\gamma}_1 + \hat{\beta}_2$, and the 2-step approaches decompose the total effect into the direct and indirect effects.

Figure 2(a) reports the estimate of the common approach (Table 1, #1). As we can see, $\hat{\beta}_1$ is an unbiased estimate of the direct effect (dashed line) of the intervention as long as the errors v_1 and v_2 are uncorrelated. However, this approach never recovers the true TEI (solid line), and studies that rely on this strategy can only report the unmediated effect of the intervention. And worse yet, the estimate of the direct effect is also likely to be biased: as the error correlation between v_1 and v_2 increases, post-treatment bias becomes more severe and biases the estimate increasingly downwards.[4]

[2] Specifically, $(v_1, v_2) \sim \mathcal{BN}(\mathbf{0}, \mathbf{\Sigma})$, where the 2x2 variance-covariance matrix $\Sigma_{2x2} = (1, cor(v_1, v_2), cor(v_1, v_2), 1)$. We set correlations to $cor(v_1, v_2) = 0$, $cor(v_1, v_2) = 0.3$, $cor(v_1, v_2) = 0.6$, and $cor(v_1, v_2) = 0.9$.

[3] When we change the error correlations $cor(v_1, v_2)$, we assume throughout that $cor(D, W) = 0.75$; for the unlikely case in which D and W were uncorrelated, the comprehensive 2-step approach would be able to recover the TEI without bias even though the DEI estimate would still be biased. When we manipulate $cor(D, W)$ (i.e., impose omitted variable bias), we set $cor(v_1, v_2) = 0$ to isolate the effect of this bias alone.

[4] Note that the observed downward bias results from positive error correlation of v_1 and v_2, which according to Gerber and Green (2012, 325) is not uncommon.



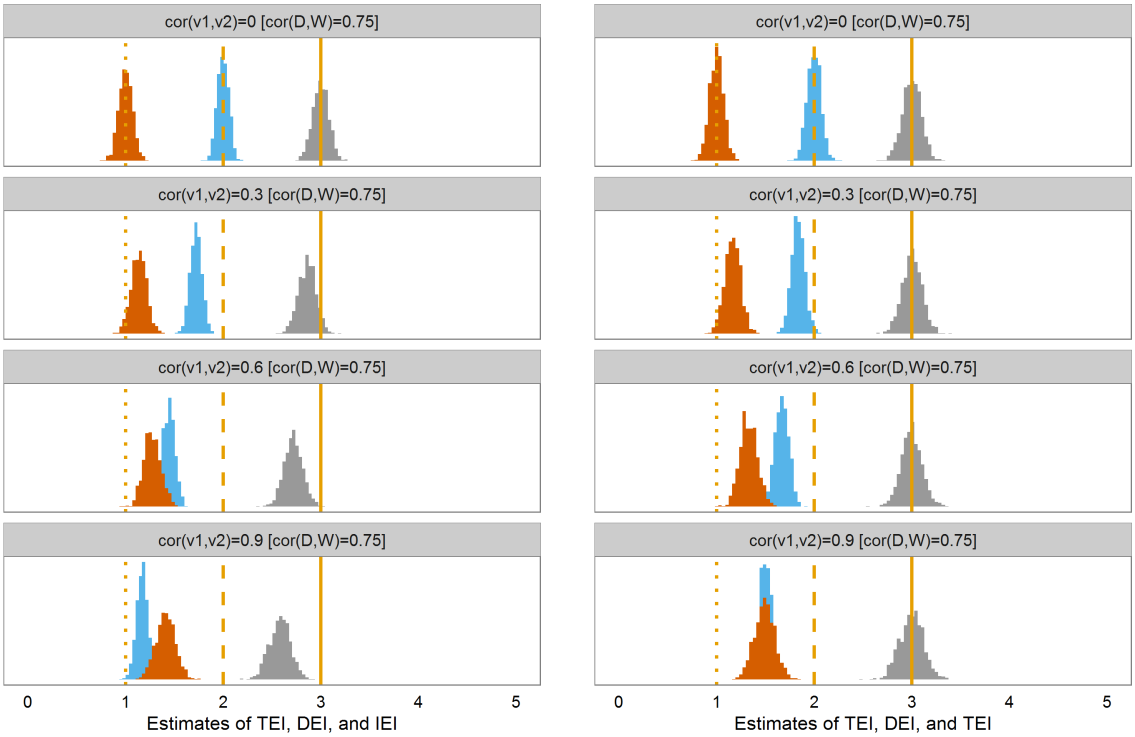
(a) Common approach.

(b) Reduced form approach.

Figure 2: Distribution of coefficient estimates for the common approach and the reduced form approach. The *true* value of the DEI is 2 (dashed yellow line) and the *true* value of the TEI is 3 (solid yellow line). The common approach (**left panel**) estimates the direct effect $\hat{\beta}_1$. The true TEI is never correctly recovered and the true estimate for $\beta_1 = 2$ is only correctly estimated for zero error correlation. Higher error correlations bias $\hat{\beta}_1$ downwards due to post-treatment bias. The reduced form approach (**right panel**) finds the correct estimate of the TEI, but only when there is no correlation of D and X through W. Gray and blue histograms show the simulated coefficient distributions for the point estimates of the TEI and DEI.

The reduced form approach (Table 1, #2), which omits X altogether, almost exactly mirrors the previous findings. This strategy recovers the total effect. However, the estimate from this approach is only unbiased under the assumption that D and X are uncorrelated.[5] In our case, we induce correlation between the intervention D and the omitted variable X indirectly through controlling the correlation structure between D and W. If these two are positively correlated, as in our Monte Carlo simulations, the omitted variable produces upward bias.

[5] With our data generating process, omitted variable bias can analytically be characterized as $\hat{\beta}_1 = \beta_1 + \gamma_1 \frac{cov(D,X)}{var(D)}$ (Greene, 2008, 134).



(a) Reduced form 2-step approach.

(b) Comprehensive 2-step approach.

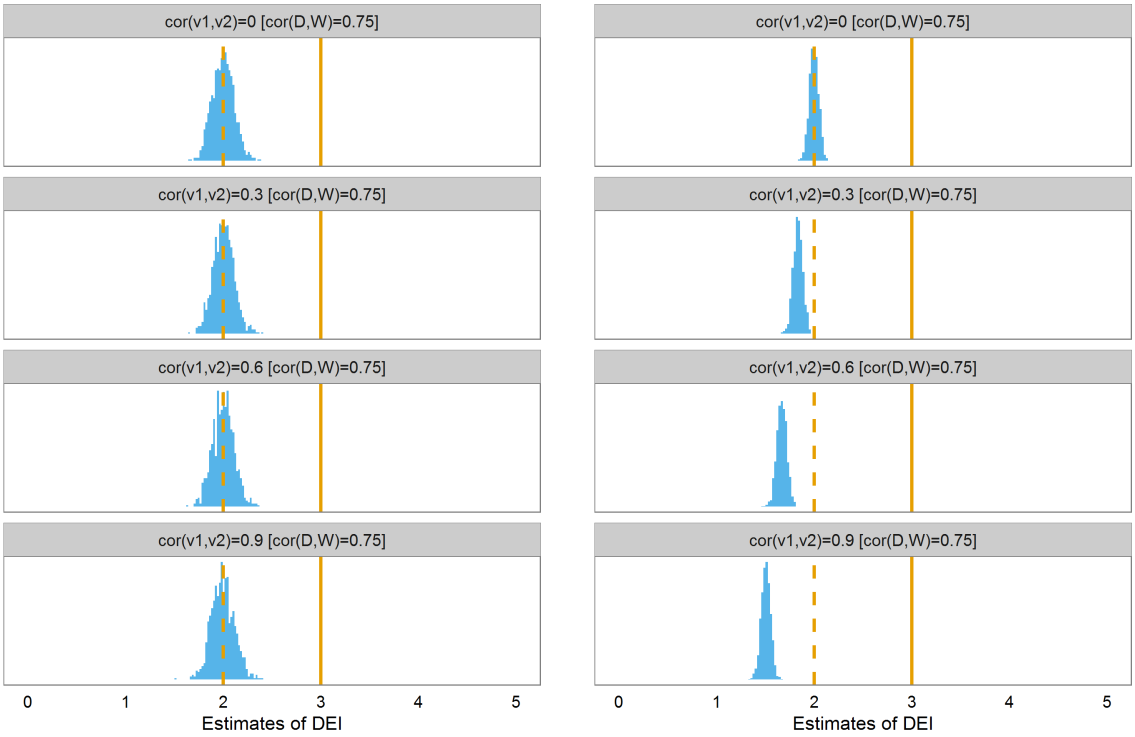
Figure 3: Distribution of coefficient estimates for 2-step approaches. The *true* value of the DEI is 2 (dashed yellow line) and the *true* value of TEI is 3 (solid yellow line). The reduced form 2-step approach (**left panel**) recovers the true direct effect for $\hat{\beta}_1 = 2$ and the true TEI when there is no post-treatment bias; increasing levels of error correlation $cor(v_1, v_2)$ bias both DEI and TEI downwards. The comprehensive 2-step approach (**right panel**) correctly retrieves the TEI, but fails to recover the direct effect DEI ($\hat{\beta}_1$) without bias, at least for non-zero error correlations; as long as either the DEI or TEI is biased, so must be the IEI estimates, $\hat{\gamma}_1\hat{\beta}_2 = 1$. Gray, blue, and red histograms show the simulated coefficient distributions for the point estimates of the TEI, DEI and IEI.

Next, Figure 3 shows the 2-step estimates for a full set of effects; these approaches can separate the total effect into its direct and indirect effect components. We distinguish between a reduced form (Table 1, #3) and a comprehensive 2-step variant (Table 1, #4) depending on whether W is included in the first OLS regression. TEI estimates are colored in gray, DEI estimates in blue, and IEI estimates are shown in red.

Once again, we find that estimates perform well when the error terms v_1 and v_2 are uncorrelated. As the correlation between the two increases, the bias of DEI estimates increases across both estimation procedures. When the error correlation is positive, as for most applications (Gerber and Green, 2012), then the bias is negative and estimates tend towards zero. In these cases, both 2-step approaches systematically underestimate the direct effect of the intervention. We may thus not find effects of international organizations, for example, even though these effects do exist. The same is also true for TEI estimates as long as we estimate reduced form regression. For a more comprehensive specification, however, which includes W in the first regression, the true total effect can be recovered without bias as shown by the gray histograms in Figure 3(b). For completeness, we show that estimates for indirect effects necessarily have to be biased as long as either the DEI or TEI are biased.

Finally, we consider an instrumental variables (IV) approach (Table 1, #5), which is estimated with 2-stage least squares (2SLS), and the average controlled direct effect (ACDE) approach, recently introduced by Acharya, Blackwell, and Sen (forthcoming). In the IV estimations, we use W as a valid instrument for X and assume the exclusion restriction to be met by definition. This is sufficient for the IV approach to produce an unbiased estimate of the DEI as long as W is measured pre-intervention. If this assumption is however not met, post-treatment bias creeps in and biases the estimate of the DEI. The total effect cannot be retrieved as β_2 is not properly defined as long as D cannot be used as an instrument in the first-stage regression.

In summary, we find that all most popular strategies to deal with post-treatment bias suffer from a number of severe problems. Some cannot distinguish between the direct and total effects, and others are vulnerable to bias in at least DEI or TEI estimates if (i) the error terms v_1 and v_2 are correlated and if (ii) the intervention is not randomly assigned, but correlated with the mediator



(a) IV approach.

(b) ACDE approach.

Figure 4: Distribution of coefficient estimates for the IV and ACDE approach. The *true* value of the DEI is 2 (dashed yellow line) and the *true* value of TEI is 3 (solid yellow line). The 2SLS IV approach (**left panel**) produces unbiased estimates of the DEI, while the TEI cannot be calculated as D does not meet the exclusion restriction. The average controlled direct effect (ACDE) approach (**right panel**) provides unbiased estimates of the DEI were the sequential unconfoundedness assumption met, but cannot recover unbiased DEI estimates in our setup. Gray and blue histograms show the simulated coefficient distributions for the point estimates of the TEI and DEI.

X. Next, we suggest a competing approach that yields unbiased estimates of both the direct and total effects.

6 A Simple Solution: Total Effect Decomposition

6.1 The Basic Logic

We suggest a simple, yet powerful solution to the post-treatment bias problem. We call it the *total effect decomposition* approach. This solution can be implemented as long as the analyst has access

to (sufficiently many) data of observations before the intervention was assigned (pre-intervention). This requirement is fulfilled if the data is in time-series, cross-section format with within-unit variation on the intervention.

Before we explain the procedure, a note on the terminology used here: post-*treatment* variables are *causally* following an intervention. Post- or pre-treatment refers to a specific set of variables in a data generating process. The definition of this set is a matter of theory. Post-*intervention* observations are observed while or after the intervention occurred. For instance, if the intervention in question is democracy, then all observations in Chile after 1990 are post-intervention. The pre- or post-intervention status is attached to an observation and is therefore an empirical matter. A variable can be measured post-intervention without being post-treatment.

The intuition behind our approach is straightforward. We use pre-intervention data to recover estimates of the effects of X (the mediator), W, and Z (the true confounders) on Y. Then, we simulate the expected value of Y for post-intervention observations, ignoring the intervention D in our regressions altogether. These simulated values are an estimate of the *counterfactual* expected values of Y. Depending on whether we use estimated coefficients from regressions of Y on X (recovers DEI) or Y on W (recovers TEI) to calculate expected values $\hat{Y}$, the difference between the actual, observed values of Y and the counterfactual $\hat{Y}$ gives us the desired estimates of the direct and total effect.

Let us first explain how our method works in practice to provide the full intuition behind it. We then derive its properties more rigorously. In what follows, let us subscript pre-intervention observations with $D = 0$ and post-intervention observations with $D = 1$.[6] Further, assume that the dataset is structured in cross-sections and time series, with N units observed T times.

[6]For simplicity, we assume that all units receive the intervention at the same point in time, so that indexing by units i and time t is not necessary. In our application in section 7 we show that our algorithm easily generalizes to setups where units receive the intervention at different points in time.

If we only know X

Suppose we only know and observe X, while W is unobserved. In this situation, we can only recover the DEI. To do so, we first estimate:

$$Y_{D=0} = \alpha^* + \gamma^* X_{D=0} + \delta^* Z_{D=0} + v^*$$

with least squares. We only use data in the absence of the intervention. Call $\hat{\cdot}$ the OLS estimates of each parameter. Based on this, we can calculate:

$$\hat{Y}_{D=1} = \hat{\alpha}^* + \hat{\gamma}^* X_{D=1} + \hat{\delta}^* Z_{D=1}$$

using *post*-intervention data but *pre*-intervention coefficients from the first step. The DEI can then be easily computed as the average of the difference between the observed value of Y ($Y_{D=1}$) and the counterfactual ($\hat{Y}_{D=1}$) over all post-intervention observations. That is:

$$\widehat{\text{DEI}} = E\Big[Y_{D=1} - \hat{Y}_{D=1}\Big] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{D=1} - \hat{Y}_{D=1}).$$

If we do not know or cannot observe W, then this is where the story ends. In such a situation, we cannot recover more than an unbiased estimate of the DEI. If, however, we know and can observe W, then we can go further and recover both the IEI and the TEI as well.

If W is known and observable

Suppose now that W is known and observable. First, we would follow the procedure above to obtain $\widehat{\text{DEI}}$. Then, to obtain $\widehat{\text{TEI}}$, begin by estimating:

$$Y_{D=0} = \alpha^* + \gamma^* W_{D=0} + \delta^* Z_{D=0} + v^*$$

with least squares. We use $\tilde{\cdot}$ to distinguish these estimates from the ones from the DEI. Based

on these estimates, compute

$$\tilde{Y}_{D=1} = \tilde{\alpha}^* + \tilde{\gamma}^* W_{D=1} + \tilde{\delta}^* Z_{D=1},$$

which then allows computing the TEI as

$$\widehat{\text{TEI}} = E\Big[Y_{D=1} - \tilde{Y}_{D=1}\Big] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{D=1} - \tilde{Y}_{D=1}).$$

Finally, we can then recover an unbiased estimate of the indirect effect $\widehat{\text{IEI}}$ by subtracting the direct effect estimate $\widehat{\text{DEI}}$ form the total effect estimate $\widehat{\text{TEI}}$.

6.2 Why It Works

To show more formally why the above 'cookbook procedure' works, consider the full data generating process as described in equations (1) and (2), and let us again index *pre*-intervention variables with $D = 0$ and *post*-intervention variables with $D = 1$. Then, when we regress observed outcome data Y on X in pre-intervention periods, such that

$$Y_{D=0} = \alpha^* + \gamma^* X_{D=0} + \delta^* Z_{D=0} + v^*,$$

we can construct post-intervention expected values $\hat{Y}_{D=1}$, as our counterfactual, using post-intervention data $X_{D=1}$. This yields:

$$\hat{Y}_{D=1} = \hat{\alpha}^* + \hat{\gamma}^* X_{D=1} + \hat{\delta}^* Z.$$

Treating Y as the dependent variable from the data generating process in pre-intervention periods, we know that $Y_{D=0}$ is given as

$$Y_{D=0} = \alpha_1 + \gamma_1 X_{D=0} + \delta_1 Z_{D=0} + v_1.$$

This demonstrates that the first stage regression of our procedure essentially estimates $\hat{\alpha}^*$ as α_1, $\hat{\gamma}^*$ as γ_1, and $\hat{\delta}^*$ as δ_1, which are all unbiased estimates of the *true* coefficients. This follows

immediately from the OLS standard properties from the Gauss-Markov theorem, and from the fact that using pre-intervention data ensures error terms to be uncorrelated. As post-treatment bias is a function of $cov(v_1, v_2)$, no bias creeps in for $cov(v_1, v_2) = 0$, which is the case for any pre-intervention variables (Gerber and Green, 2012).

With these estimates, the expected value of the first difference of actually observed, post-intervention $Y_{D=1}$—as obtained from the data generating process—and $\hat{Y}$ gives

$$\begin{aligned} E\Big[Y_{D=1} - \hat{Y}_{D=1}\Big] &= E\Big[(\alpha_1 + \beta_1 + \gamma_1 X_{D=1} + \delta_1 Z_{D=1} + v_1) - (\hat{\alpha}^* + \hat{\gamma}_1^* X_{D=1} + \hat{\delta}^* Z_{D=1})\Big] \\ &= \beta_1 + E[v_1] \\ &= \beta_1 \equiv \text{DEI}. \end{aligned}$$

This shows that as long as $E[v_1] = 0$, which is a standard assumption, our procedure can recover the true direct effect without any bias. Importantly, using *pre*-intervention data only when regressing Y on X in the first step of our algorithm ensures that our estimates of $\hat{\alpha}^*$, $\hat{\gamma}^*$, and $\hat{\delta}$ are indeed unbiased and recover the true values of α_1, γ_1, and δ_1. By using only pre-intervention variables in how we construct the counterfactual, we avoid the problem of post-treatment bias altogether.

Following the very same logic, but regressing Y on pre-intervention values of W (instead of X), yields the following simple regression model:

$$Y_{D=0} = \alpha^* + \gamma^* W_{D=0} + \delta^* Z_{D=0} + v^*$$

As before, generating model predictions (and denoting estimated coefficients with $\tilde{}$ to distinguish them from the ones above) with post-intervention data for W results in counterfactual observations for Y, such that

$$\tilde{Y}_{D=1} = \tilde{\alpha}^* + \tilde{\gamma}^* W_{D=1} + \tilde{\delta}^* Z_{D=1}.$$

Recognizing that Y is a function of W through the data generating process and can, in pre-

intervention periods with $D = 0$, be written as[7]

$$Y_{D=0} = (\alpha_1 + \gamma_1\alpha_2) + \gamma_1\gamma_2 W_{D=0} + \delta_1 Z_{D=0} + (v_1 + \gamma_1 v_2),$$

regressing $Y_{D=0}$ on $W_{D=0}$ in the first stage effectively estimates $\tilde{\alpha}^*$ as $\alpha_1 + \gamma_1\alpha_2$, $\tilde{\gamma}^*$ as $\gamma_1\gamma_2$, and $\tilde{\delta}^*$ as δ_1. Therefore, the expected value of the first difference of $Y_{D=1}$ and $\tilde{Y}_{D=1}$ is equal to

$$E\Big[Y_{D=1} - \tilde{Y}_{D=1}\Big] = E\Big[\Big((\alpha_1 + \gamma_1\alpha_2) + (\beta_1 + \gamma_1\beta_2) + \gamma_1\gamma_2 W_{D=1} + \delta_1 Z_{D=1} + (v_1 + \gamma_1 v_2)\Big) - \Big(\tilde{\alpha}^* + \tilde{\gamma}^* W_{D=1} + \tilde{\delta}^* Z_{D=1}\Big)\Big].$$

With the above identities for $\tilde{\alpha}^* = \alpha_1 + \gamma_1\alpha_2$, $\tilde{\gamma}^* = \gamma_1\gamma_2$, and $\tilde{\delta}^* = \delta_1$, this simplifies to

$$\begin{aligned} E\Big[Y_{D=1} - \tilde{Y}_{D=1}\Big] &= \beta_1 + \gamma_1\beta_2 + E[v_1 + \gamma_1 v_2] \\ &= \beta_1 + \gamma_1\beta_2 \equiv \text{TEI}, \end{aligned}$$

proving that our algorithm can retrieve an unbiased estimate of the TEI if W is observable and if the following is true: $E[v_1] = 0$ and $E[\gamma_1 v_2] = 0$ need to hold, which follows immediately from errors v_1 and v_2 coming from the standard bivariate normal distribution. As before, the key analytical insight as to why our procedure works is that $\tilde{\alpha}^*$, $\tilde{\gamma}^*$, and $\tilde{\delta}^*$ can be estimated without bias because we only use *pre*-intervention data in the first step of our algorithm.

What assumptions need to be met for this procedure to be valid? Technically, we need cases observed in the absence of the intervention. For instance, if we study the effect of democratic institutions, we need to observe autocracies. Moreover, the other variables (W, X, and Z) need to vary within these cases. Otherwise we could not estimate the key parameters. More data pre-intervention helps generate more precise estimates of the counterfactual and is therefore desirable. Note that this technical assumption also has implications for the nature of our intervention. The intervention needs to be a discrete variable where the value zero denotes the absence of the intervention altogether.[8]

[7] Setting $D = 0$ in equation (3) and only using pre-intervention data, we obtain this expression.

[8] We focused so far on dichotomous interventions, but the procedure can be generalized to multiple categories.

We also make substantive assumptions. Of course, the usual assumptions, such as having a well-specified model, still must be met. But this is always true with observational data in non-experimental settings. Another assumption is one we label *parameter stability*. The effect of the confounding (W, Z) and mediating variables (X) must remain constant regardless of the status of the intervention. Another way to formulate this assumption is that the parameters need to stay the same once the intervention has been triggered. This assumption will not be met if the intervention has structural consequences and transforms the entire data generating process. Whether this is the case is not testable. And in fact, a large literature triggered by the Lucas critique of macroeconomic policy interventions suggests that there are important cases in which the stability of parameters cannot be taken for granted (Lucas, 1976). This being said, we note that analysts doing non-structural quantitative work implicitly make this assumption, regardless of the issue of post-treatment bias.

6.3 Monte Carlo Simulations

We now use the same randomly generated data for a Monte Carlo simulation to benchmark our method against the other approaches discussed above. Using time-series, cross-sectional data, we consider 100 units over 20 time periods for a total of 2,000 observations. As before, the intervention is implemented in all units simultaneously starting in the tenth time period. Thus, for all units, $D = 0$ if $t < 10$ and $D = 1$ after $t \geq 10$. All variables are generated from the same data generating process.

To recover estimates of the direct and total effects, we follow our own 'cookbook' outlined above. To obtain the DEI, we first regress Y on X (and Z) with only pre-intervention data. We then use the estimated coefficients, for which we know they are unbiased *by design*, to predict expected values for post-intervention periods. Finally, we calculate the difference between observed post-intervention data and simulated counterfactuals and average these individual differences across all post-intervention observations. For estimates of the total effect, we implement exactly the same routine, with the only difference that we replace X with W in the first two steps. For both the

However, it cannot work with continuous independent variables.

DEI and TEI, we repeat this procedure 1,000 times and plot results in Figure 5.

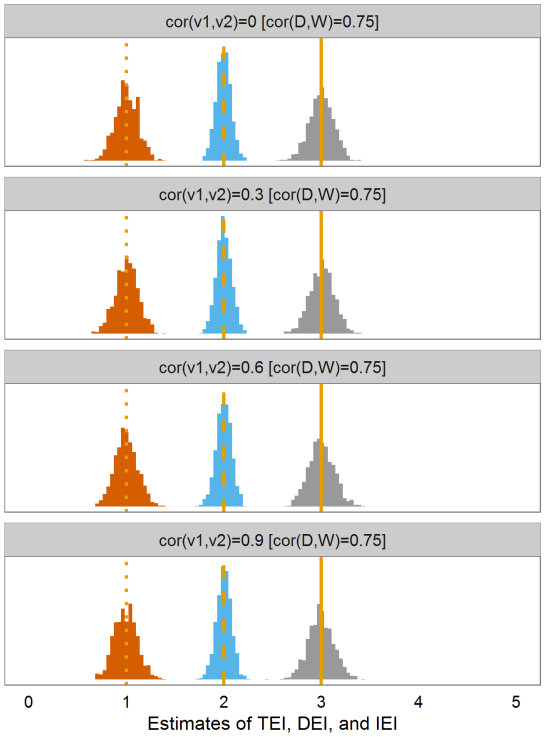


Figure 5: Distribution of coefficient estimates for our alternative approach. The *true* values for the DEI (dashed yellow line), TEI (solid yellow line), and IEI are 2,3, and 1. Our new proposed approach retrieves all three quantities of interest correctly even for high levels of error correlation. Gray, blue, and red histograms show the simulated coefficient distributions for the point estimates of the TEI, DEI, IEI.

Showing distributions of point estimates for the DEI in blue and for the TEI in gray, we can see that our approach is remarkably reliable, *regardless* of the correlation structure of error terms. Both the DEI and TEI are unbiased for *any* correlation level. The three-step procedure which we propose here is the only one (relative to the other approaches discussed above) that is robust to correlation both in errors but also between D and W; the estimates get neither worse nor better as correlations increase.

7 Application: The Effect of Democracy on FDI

We next demonstrate a concrete application of our method. Foreign direct investments (FDI) are believed to be a key driver of economic growth (de Mello, 1997; de Mello, 1999; Chowdhury and Mavrotas, 2006). FDI refers to cross-national private investments generally assumed to follow a long time horizon. These investments increase the amount of capital in a given country, which generates growth and employment. Given these attractive features, governments often seek to attract FDI. In his seminal work, Jensen (2003) argues that FDI is particularly likely to be targeted toward democracies. This is because democracies reduce political risks faced by international corporations. Once an investment is made, governments may be tempted to expropriate firms—a classical time inconsistency problem (Kobrin, 1987). Jensen (2003, 594) offers two reasons why democracies are less likely to do so. First, democracies have more veto players, reducing the ability of the government of expropriating firms. Second, democratic governments could suffer from audience cost if they were deemed unreliable by voters.[9]

Empirically, Jensen (2003, 605) finds strong empirical support for his main claim: democracies receive more FDI. We focus on his cross-section time series results (Table 4), especially Model 10 and 11. The regression equation for the latter is:

$$\text{FDI (\% of GDP)}_{i,t} = \alpha_i + \tau_t + \beta \text{Democracy}_{i,t} + \gamma \text{Capital Controls}_{i,t} + \psi' \mathbf{W}_{i,t} + \theta' \mathbf{Z}_{i,t} + \varepsilon_{i,t}$$

where α are country fixed effects and τ are decade dummies. We discuss the content of the two vectors $\mathbf{W}$ and $\mathbf{Z}$ below.

Drawing on Polity data, an increase of the democracy index by one standard deviation (7.78) increases FDI per GDP by 0.16 percentage points. We are able to replicate perfectly these results using his useful replication package (Table 2).[10] Column (1) and (3) replicate Jensen's Table 4, Models 10 and 11, respectively. Since our approach does not work well with continuous interventions, we also verify that the results are robust when using the democracy indicator from Alvarez

[9] At the same time, Jensen acknowledges that autocracies may have advantages too, though he believes these to be overstated. Autocracies can offer better deals to firms, a cheaper labor force, a weaker regulations.

[10] We made one modification to Jensen's original dataset: the capital control variable he used is not publicly available. Instead, we use Karcher and Steinberg (2013).

et al. (1996), which we label *Democracy (ACLP)*. Again, Jensen's results are very robust (columns 2 and 4). Switching from autocracy to democracy increases FDI by 0.38 percentage points.

Replication of Jensen (2003), Table 4

	Model 10	Model 10 (alt.)	Model 11	Model 11 (alt.)
	(1)	(2)	(3)	(4)
FDI (t-1)	0.364***	0.309***	0.321***	0.271***
	(0.069)	(0.066)	(0.066)	(0.063)
Market Size	−0.554	−0.614	−0.193	−0.103
	(0.462)	(0.453)	(0.466)	(0.458)
Development	0.834*	0.947**	0.492	0.376
	(0.466)	(0.445)	(0.503)	(0.487)
Growth	0.024***	0.023***	0.021**	0.018**
	(0.008)	(0.008)	(0.008)	(0.008)
Trade	0.006	0.007	0.007	0.009*
	(0.004)	(0.005)	(0.005)	(0.005)
Budget Deficit	−0.023**	−0.024**	−0.018*	−0.020*
	(0.011)	(0.011)	(0.011)	(0.011)
Gov Consumption	−0.039**	−0.045**	−0.031*	−0.036**
	(0.017)	(0.018)	(0.017)	(0.018)
Democracy (Polity)	0.021**		0.021**	
	(0.008)		(0.009)	
Democracy (ACLP)		0.380***		0.384***
		(0.101)		(0.098)
Capital Openness (K&S)			0.167***	0.208***
			(0.043)	(0.045)
Observations	1630	1584	1532	1522
Country Fixed Effects	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes

Note: Panel corrected standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table 2: Replication of Jensen (2003, Table 4). Unit of analysis is country-year.

In our view, one weakness in this analysis is the crucial role played by capital controls. Capital controls refer to any policy that facilitates or prevents capital flows across countries (King and Levine, 1993). The free movement of capital across borders may strengthen the willingness of investors to send their capital abroad, since it should mean that they will not be penalized if they ever want to repatriate their investments. Since capital controls are a policy, they are under the (constrained) control of policymakers. The problem is that a vast literature posits that democracies are more likely to implement liberal economic policies (Milner and Kubota, 2005; Frye and Mansfield, 2004), and this is also true of capital controls (Brooks and Kurtz, 2007). It is therefore highly likely that capital openness is a post-treatment variable. Therefore, the results of Table 2, columns (1) and (3) are likely to suffer from omitted variable bias (capital controls have an effect

on FDI), while columns (2) and (4) are likely to suffer from post-treatment bias.

The question then is: Do democracies receive more FDI because (a) they are more credible (audience cost), or (b) because they adopt more liberal policies (capital controls)? Audience cost is notoriously difficult to measure (Chaudoin, 2014). However, we can use our measurement strategy to at least assess the strength and relevance of capital controls.

We adapt Jensen's (2003) model to our framework (Figure 6). In our previous notation, democracy is the intervention D and capital controls are the mediating variable X. What, then, are $\mathbf{Z}$ and $\mathbf{W}$? Variables that are causally related to capital controls should be included in the vector $\mathbf{W}$ and those that are unrelated should be attached to $\mathbf{Z}$. We draw on the literature on the determinants of capital controls to partition the confounding variables in either category. Specifically, we rely on Brooks and Kurtz (2007, 710), who analyze the economic and political sources of capital controls. We cross-check their explanatory variables and the ones used by Jensen (2003). Those present in both papers are incorporated in $\mathbf{W}$ while those that are not are included in the $\mathbf{Z}$ vector. That way we remain as close as possible to Jensen's original specification, which makes the results more comparable.[11] This enables us to define the two vectors as follows:

- $\mathbf{W}$: development level (i.e., GDP per capita), growth, trade;
- $\mathbf{Z}$: market size, budget deficit, government consumption, lagged FDI, and fixed effects.

We are then in a position to estimate all quantities of interest: DEI, IEI, and TEI. We follow the procedure described in Section 6. In order to construct confidence intervals for the total, direct, and indirect effects, we use non-parametric bootstrapping.[12] The results are reported in Table 3.

We find that democracy increases FDI when we consider the direct effect only (net of capital controls). The estimate of DEI is 0.45. The indirect effect, which is the effect of democracy through capital controls, is positive (0.03), but it is very small in comparison to the total effect. To further contrast the difference between the DEI and IEI, we plot the point estimates in Figure 7.

[11] In the appendix, we add additional controls listed in Brooks and Kurtz (2007, 710) but not in Jensen's original article. The results remain the same.

[12] Specifically, we re-sample 1,000 times with replacement from the distribution of first-differences for the respective quantity of interest (DEI, IEI, or TEI) and record the mean of each such bootstrap. This results in a distribution of 1,000 bootstrapped means, from which we construct the 95% confidence interval through normal approximation. This is implemented through R's `boot` package.

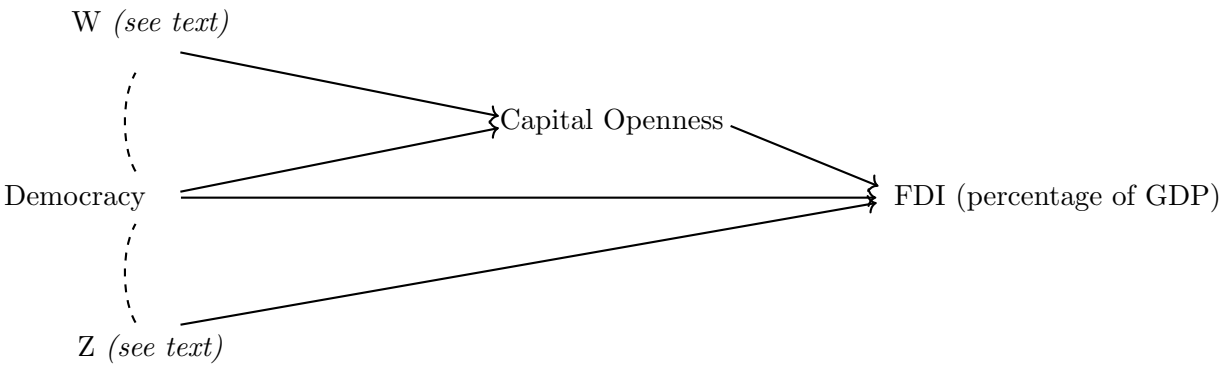


Figure 6: Causal path of democracy on FDI inflows. Adapted from Jensen (2003).

Estimates of the Direct, Indirect, and Total Effects

Estimand	Estimate	95% Confidence Interval
Direct effect of democracy ($\widehat{\text{DEI}}$)	0.45	[0.29,0.62]
Indirect effect of democracy ($\widehat{\text{IEI}}$) (through capital controls)	0.03	[-0.22,0.25]
Total effect of democracy ($\widehat{\text{TEI}}$)	0.48	[0.29,0.65]

Table 3: Estimates of the direct (DEI), indirect (IEI), and total (TEI) effect of democracy on FDI inflows (as a percentage of GDP). The mediating variable is an index of capital controls (Karcher and Steinberg, 2013). The set of variables in the vector **W** and **Z** are described in the text.

These results, to us, suggest that the effect of democracy on FDI does *not* primarily operate through more liberal policies. Instead, our findings are consistent with an interpretation based on audience costs and increased credibility through democratic institutions. This being said, remember that the DEI is the sum of all effects that are not modeled in the IEI. As such, the DEI may be a combination of various factors. We therefore did not demonstrate that audience costs are the sole mechanism through which democracy operates. But in the spirit of Popper, we showed that we can rule out the competing hypothesis that relies on democracies adopting the right kind of liberal policies.

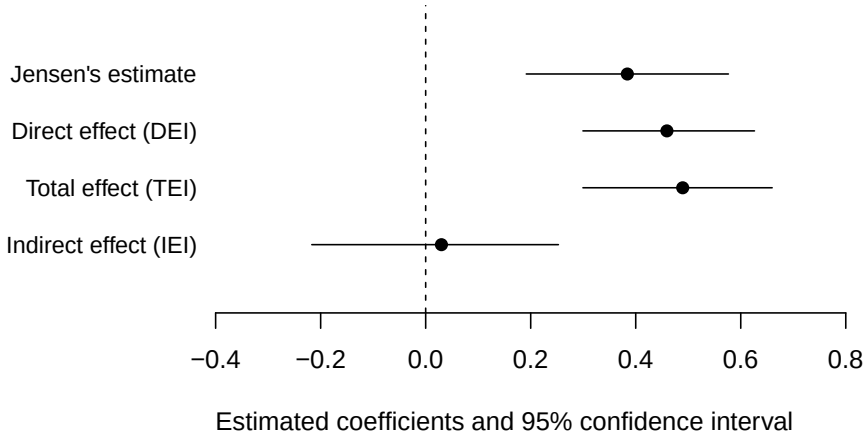


Figure 7: Comparison of Jensen's estimated effect of democracy on FDI inflows and estimated total, direct, and indirect effects from our proposed method. 95% confidence intervals for TEI, DEI, and IEI come from nonparametric bootstrapping.

8 Conclusion

Disentangling the different ways in which causal processes can operate is an important part of social science research. This is particularly true when the processes in question are diffuse. For instance, democracy does many things before it affects the outcomes we are interested in. We need appropriate tools to identify the causal channels that truly matter. Unfortunately, applied researchers are often trapped between a rock (post-treatment bias) and a hard place (omitted variable bias).

In this paper, we offered a simple tool to solve this difficult dilemma. Under relatively benign assumptions, we can recover estimates of the direct, indirect, and total effects of an intervention. This creates valuable opportunities for research in other areas. Oftentimes, theories do not compete in terms of their main predictions. Instead, they differ with respect to the causal channels that they believe are more relevant. But as we know, adjudicating between different causal channels remains a major challenge. We believe that the tools provided in this paper will help researchers venture in this difficult terrain.

At the same time, we make it abundantly clear that the quality of the inferences that we can draw with our method is only as good as the theory underpinning the analysis. We heartily support recent calls for tighter overlaps between theory and empirical work (Samii, 2016). In particular,

we believe that thinking hard about causal channels and opening up the "black box of causality" is essential (Imai et al., 2011).

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490): 493–505.

Acharya, Avidit, Matthew Blackwell, and Maya Sen. forthcoming. "Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects." *American Political Science Review* .

Alvarez, Michael R., José Antonio Cheibub, Fernando Limongi, and Adam Przeworski. 1996. "Classifying Political Regimes." *Studies in Comparative International Development* 31 (2): 3–36.

Baum, Matthew A., and David A. Lake. 2003. "The Political Economy of Growth: Democracy and Human Capital." *American Journal of Political Science* 47 (2): 333–347.

Brooks, Sarah M., and Marcus J. Kurtz. 2007. "Capital, Trade, and the Political Economies of Reform." *American Journal of Political Science* 51 (4): 703–720.

Chaudoin, Stephen. 2014. "Promises or Policies? An Experimental Analysis of International Agreements and Audience Reactions." *International Organization* 68 (1): 235–256.

Chowdhury, Abdur, and George Mavrotas. 2006. "FDI and Growth: What Causes What?" *World Economy* 29 (1): 9–19.

de Mello, Luiz R. 1997. "Foreign Direct Investment in Developing Countries and Growth: A Selective Survey." *Journal of Development Studies* 34 (1): 1–34.

de Mello, Luiz R. 1999. "Foreign Direct Investment-Led Growth: Evidence From Time Series and Panel Data." *Oxford Economic Papers* 51 (1): 133–151.

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

Frye, Timothy, and Edward D. Mansfield. 2004. "Timing is Everything: Elections and Trade Liberalization in the Postcommunist World." *Comparative Political Studies* 37 (4): 371–398.

Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W.W. Norton.

Goldstein, Judith L., Douglas Rivers, and Michael Tomz. 2007. "Institutions in International Relations: Understanding the Effects of the GATT and the WTO on World Trade." *International Organization* 61 (1): 37–67.

Greene, William H. 2008. *Econometric Analysis*. 6th ed. New York: Pearson.

Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies." *American Political Science Review* 105 (4): 765–789.

Jensen, Nathan M. 2003. "Democratic Governance and Multinational Corporations: Political Regimes and Inflows of Foreign Direct Investment." *International Organization* 57 (3): 587–616.

Karcher, Sebastian, and David A Steinberg. 2013. "Assessing the Causes of Capital Account Liberalization: How Measurement Matters." *International Studies Quarterly* 57 (1): 128–137.

King, Gary. 2010. "A Hard Unsolved Problem? Post-treatment Bias in Big Social Science Questions." Presented at the "Hard Problem in Social Science" Symposium, Harvard University.

King, Robert G., and Ross Levine. 1993. "Finance and Growth: Schumpeter Might be Right." *Quarterly Journal of Economics* 108 (3): 717–737.

Kobrin, Stephen J. 1987. "Testing the Bargaining Hypothesis in the Manufacturing Sector in Developing Countries." *International Organization* 41 (4): 609–638.

Lucas, Robert E. 1976. "Econometric Policy Evaluation: A Critique." In *The Phillips Curve and Labor Markets*, ed. Karl Brunner, and Alan Melzer. Amsterdam: North Holland.

Milner, Helen V., and Keiko Kubota. 2005. "Why the Move to Free Trade? Democracy and Trade Policy in the Developing Countries." *International Organization* 59 (1): 107–143.

Pearl, Judea. 2009. *Causality*. Cambridge University Press.

Pepinsky, Thomas B. 2014. "The Politics of Capital Flight in the Global Economic Crisis." *Economics & Politics* 26 (3): 431–456.

Rose, Andrew K. 2004. "Do We Really Know That the WTO Increases Trade?" *American Economic Review* 94 (1): 98–114.

Rosenbaum, Paul R. 1984. "The Consequences of Adjustment for a Concomitant Variable that has been Affected by the Treatment." *Journal of the Royal Statistical Society* 147 (5): 656–666.

Samii, Cyrus. 2016. "Causal Empiricism in Quantitative Research." *Journal of Politics* 78 (3): 941–955.

Tomz, Michael, Judith L Goldstein, and Douglas Rivers. 2007. "Do We Really Know That the WTO Increases Trade? Comment." *American Economic Review* 97 (5): 2005–2018.