Included Variable Bias: A Formula for Collider Bias in Cross-Sectional and Time-Series Cross-Sectional Regressions

Manoel Galdino[*]   Davi Moreira[†]   Carolina Dolleans[‡]

February 2026

Abstract

The selection of control variables is a critical step in observational studies, yet standard heuristics can inadvertently lead researchers to include collider variables, introducing bias rather than removing it. While Directed Acyclic Graphs (DAGs) can diagnose whether a variable is a collider, they do not quantify the magnitude of the resulting bias. We derive the Included Variable Bias (IVB) formula—a closed-form expression that mirrors the classic Omitted Variable Bias (OVB) formula but addresses the opposite problem: the bias from erroneously including a variable. We show that $\mathrm{IVB} = -\theta^\star \times \pi$, where both components are directly estimable from data. We extend the result from cross-sectional models to the Autoregressive Distributed Lag (ADL) specifications common in time-series cross-sectional (TSCS) analysis. Monte Carlo simulations confirm the formula's exactness across multiple data-generating processes, and we illustrate its application with an example from the civil war literature.

1 Introduction

Applied researchers using observational data face a fundamental challenge: which variables should be included as controls in a regression model? The answer to this question is decisive for causal identification, yet the discipline's standard practices offer surprisingly little guidance for getting it right. Two widely used heuristics—what we call *control checking* and *confounding checking*—instruct researchers to include variables that cause the outcome or that jointly cause the treatment and the outcome. While these rules are sensible in many settings, they share a critical blind spot: they do not assess whether a candidate control variable is a collider, a variable that is caused by both the treatment and the outcome (or by the treatment and an

---

[*]Assistant Professor of Political Science, Universidade de São Paulo. E-mail: mgaldino@usp.br

[†]Professor of Quantitative Methods, Mitch Daniels School of Business, Purdue University.

[‡]Graduate Student of Political Science, Universidade Federal de Pernambuco.

unobserved common cause of the outcome). Including a collider in a regression opens a backdoor path and biases the estimate of the causal effect (Pearl, 2009; Elwert and Winship, 2014).

The problem is especially acute in time-series cross-sectional (TSCS) research designs, which are ubiquitous in comparative politics and international relations (Beck and Katz, 1995; Franzese and Hays, 2007). In TSCS settings, researchers routinely lag their control variables to avoid reverse causality. This seemingly prudent practice can, paradoxically, create collider bias. When a lagged control variable is itself caused by the outcome and by an unobserved variable that also affects the outcome, conditioning on it opens a spurious path—a phenomenon we term "foreign collider bias.''

Directed Acyclic Graphs (DAGs) have become an increasingly recognized tool for diagnosing such problems (Pearl and Mackenzie, 2018; Imbens, 2020). A growing literature in political science has begun to apply DAGs to improve causal identification (Montgomery, Nyhan and Torres, 2018; Blackwell and Glynn, 2018). DAGs provide a clear, qualitative answer to the question of *whether* including a particular variable introduces bias. However, they are silent on the equally important question of *how much* bias results. A researcher who discovers that a control variable is a collider still does not know the magnitude or even the direction of the bias without additional tools.

This paper fills that gap. We derive the **Included Variable Bias (IVB)** formula—a closed-form expression for the bias that arises when a collider variable is erroneously included in a regression model. The formula takes a remarkably simple form:

$$\text{IVB} = -\theta^\star \times \pi \tag{1}$$

where $\theta^\star$ is the coefficient on the collider in the misspecified ("long'') regression and π is the coefficient on the treatment in an auxiliary regression of the collider on the treatment and legitimate controls.[1] Both quantities are directly estimable from data, making the formula immediately operational for applied researchers.

The IVB formula mirrors the well-known Omitted Variable Bias (OVB) formula, but addresses the symmetrically opposite problem. Where OVB quantifies the cost of leaving out a relevant confounder, IVB quantifies the cost of putting in an irrelevant—or, more precisely, harmful—collider. We show that this result holds not only in standard cross-sectional regression but extends naturally to the Autoregressive Distributed Lag (ADL) models that are the workhorse of TSCS analysis. The extension relies on the Frisch–Waugh–Lovell (FWL) theorem (Frisch and Waugh, 1933; Lovell, 1963) and yields the same functional form, with the key quantities now partialled out by the dynamic controls (lagged dependent variables and lagged treatments).

---

[1] In the cross-sectional derivation, $\theta^\star$ corresponds to $\beta_2^\star$ and π corresponds to ϕ_1; we use the ADL notation throughout the paper as it is the more general form.

We validate the formula through Monte Carlo simulations across three data-generating processes: a pure cross-section, an ADL(1,0) panel, and a substantive civil war DGP inspired by the empirical literature on political change and conflict. In every scenario, the formula predicts the empirical bias with exact precision.

The remainder of this paper is organized as follows. Section~2 reviews the standard heuristics for selecting control variables and identifies their limitations. Section~3 introduces DAGs and illustrates collider bias in a TSCS context using a running example from the civil war literature. Section~4 presents the core theoretical contribution: the derivation of the IVB formula for cross-sectional and ADL models. Section~5 validates the formula through Monte Carlo simulations. Section~6 provides a practical walkthrough of the formula using the civil war example. Section~7 concludes with recommendations for applied researchers.

2   The Control Variable Problem

In observational studies, when using regression with a selection-on-observables strategy to identify a causal effect, including the right set of control variables is decisive. Which variables to include depends primarily on the researcher's scientific understanding of which variables are crucial to explain the phenomena of interest. A second criterion is that the model should be *identified*, meaning there is no bias in estimating the causal estimand of interest (Cinelli, Forney and Pearl, 2022). In practice, most scholars rely on informal heuristics to decide which controls to include. We review the two most common approaches below and argue that, while each has merit, neither is sufficient to prevent collider bias—and neither can quantify the resulting damage.

2.1   Control Checking

A standard heuristic is to review the relevant literature to search for potential causal variables of the outcome of interest and include them in the regression as controls. A typical example of such a heuristic comes from a paper on foreign aid:

> "As the previous literature on aid policy maintains, various other factors shape donor decisions about the allocation of aid resources, including other recipient characteristics and non-developmental donor goals. I include them as control [sic] to provide a fully specified model'' (Dietrich, 2016, p. 81).

One can argue in favor of this approach by noting that a thorough literature review may identify variables whose inclusion decreases the unexplained variance in the dependent variable, thereby improving the precision of the estimate of the Average Causal Effect (Hahn, 2004; Pearl, 2013; Cinelli, Forney and Pearl, 2022).

Unfortunately, this justification is not sufficient to decide whether a variable *should* be included as a control. As shown by Cinelli, Forney and Pearl (2022), it is quite possible to include a variable that decreases precision *and* induces bias, even if it is causally related to the outcome of interest. The control checking heuristic is agnostic about the causal structure linking the control to the treatment and outcome, and therefore cannot distinguish a helpful confounder from a harmful collider.

2.2 Confounding Checking

A more targeted heuristic is to survey the literature specifically for potential *confounding* variables—those that cause both the outcome and the treatment. In econometric parlance, the goal is to avoid omitted variable bias by including all appropriate confounders. A typical example of this approach comes from the literature on political regime change and civil war:

> "Our model does not attempt to present an inclusive theory of civil war, but level of democracy and political change do not provide a complete explanation. Therefore, we identify a number of control variables—Development, Ethnic Heterogeneity, Proximity of Independence, and International War in Country—whose omission might bias the results for the regime change variable''
> (Hegre et al., 2001, p. 37).

This heuristic is closer to the right intuition because it is explicitly focused on diminishing the primary source of bias in observational studies, namely, omitted variable bias. In DAG parlance, no omitted variable bias remains if the researcher blocks all backdoor paths by including all appropriate confounders. However—and this is a key point of this paper—this heuristic does not assess whether the researcher is inadvertently conditioning on a collider rather than a confounder. In the TSCS context, this shortcoming is particularly dangerous because the temporal structure of the data creates new opportunities for collider bias that are absent in pure cross-sections.

2.3 The Quantification Gap

What about formal approaches that go beyond heuristics? Several alternatives exist. Heckman (2008) argues that researchers should explicitly model the selection into treatment, which allows one to formally assess whether the causal effect is identified. While rigorous, this approach may be infeasible in practice due to tractability constraints or insufficient scientific knowledge about the selection process. Causal discovery algorithms (Spirtes and Zhang, 2016; Glymour, Zhang and Spirtes, 2019; Duarte et al., 2021) offer another path, but they presuppose that all relevant variables have been measured—precisely the problem we need to solve in the first place.

A third approach, advocated by Pearl (2009) and employed in this paper, is to draw a Directed Acyclic Graph (DAG) to determine whether the causal effect is identified. The DAG approach allows the researcher to reason about missing controls, erroneously included "bad controls,'' and even unobservable variables that set the limits of the study (Cinelli, Forney and Pearl, 2022).

Yet even the DAG approach, as typically used, has a limitation: it provides a *qualitative* diagnosis—
`this variable is a collider; do not condition on it''`---but not a `\textit{quantitative}` `one`---conditioning on this collider introduces a bias of magnitude x in direction d.'' This paper addresses precisely this gap. By deriving a closed-form formula for the Included Variable Bias, we give applied researchers a tool that complements DAGs: the graph tells them *whether* there is a problem, and the formula tells them *how much* of a problem it is.

2.4   Related Work on Collider Bias Quantification

Several prior contributions have derived formulas for collider-related bias in specific settings. Greenland (2003) quantified collider-stratification bias in linear structural equation models, showing that conditioning on a common effect can induce associations between otherwise independent causes. Pearl (2013) provided a systematic treatment of bias formulas in linear models using path analysis, including expressions for collider bias that follow from Wright's path-tracing rules. Ding and Miratrix (2015) derived closed-form bias formulas for the M-bias (collider) configuration, establishing conditions under which the bias from conditioning on a collider exceeds the bias from omitting a confounder. Most recently, Jung et al. (2024) used the exact term "included variable bias'' in the context of disparate impact estimation, focusing on the legal and fairness implications of conditioning on post-treatment variables.

Our paper builds on and extends this body of work in several ways. First, we package the collider bias formula explicitly as an OVB/IVB parallel—a pedagogical device that makes the result immediately accessible to applied researchers trained in the omitted variable bias tradition. Second, whereas prior formulas have been derived for cross-sectional or static settings, we extend the result to the Autoregressive Distributed Lag (ADL) models that dominate TSCS research in political science, and introduce the concept of "foreign collider bias'' that arises from the temporal structure of panel data. Third, we distill the formula into a practical three-step diagnostic recipe that requires only standard regression output, lowering the barrier to routine use by applied political scientists.

3 DAGs and Collider Bias

3.1 Preliminaries

Directed Acyclic Graphs (DAGs) are part of an approach to causal inference first developed by Wright (1928) and Wright (1934), and substantially advanced by the work of Pearl (2009) and collaborators (Pearl and Mackenzie, 2018). DAGs have been applied in several domains, including epidemiology, computer science, and, more recently, political science (Imbens, 2020; Yao et al., 2021).

All DAGs comprise three primary structures: chains, forks, and colliders (or inverted forks), shown in Figure~1. Understanding these structures is essential for correct control variable selection.
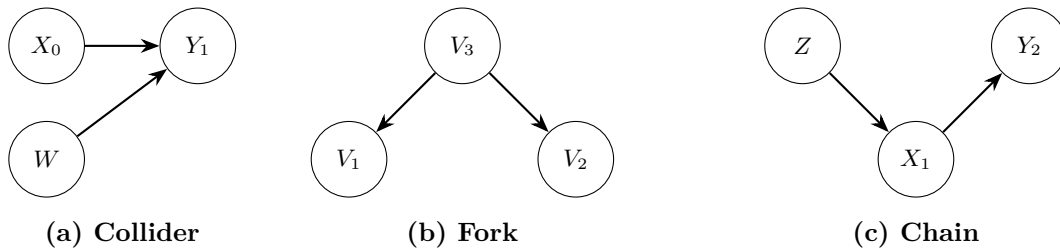

(a) Collider   **(b) Fork**   **(c) Chain**

Figure 1: The three elementary structures of DAGs: (a) a collider or inverted fork, where Y_1 is a common effect of X_0 and W; (b) a fork, where V_3 is a common cause of V_1 and V_2; and (c) a chain, where X_1 is a mediator between Z and Y_2.

In a **fork** (panel b), V_3 is a common cause of V_1 and V_2. If one is interested in the effect of V_1 on V_2, controlling for the common cause V_3 blocks the backdoor path and eliminates omitted variable bias. In a **chain** (panel c), X_1 mediates the effect of Z on Y_2. If one is interested in the total effect, controlling for the mediator blocks the causal path and leads to underestimation. In a **collider** (panel a), Y_1 is a common effect of X_0 and W. A collider, unlike chains and forks, does not induce an association between its parent variables. However, conditioning on a collider (or on its descendants) *creates* a spurious association between its causes (Elwert and Winship, 2014; Pearl and Mackenzie, 2018).

The heuristic behind "confounding checking'' can be understood as an attempt to block all backdoor paths by conditioning on forks. However, this heuristic says nothing about what happens when the candidate control is a collider. This is the problem we now illustrate in a TSCS context.

3.2 A TSCS Application: Political Change and Civil War

Consider a typical setup in comparative politics. A researcher wants to estimate the causal effect of political regime change on civil war, as in Hegre et al. (2001). Suppose the only measured confounder is per capita income, which plausibly causes both political change and civil war (Blattman and Miguel, 2010). The researcher controls for per capita income to block the backdoor path.

In a TSCS setting, variables must be indexed by time. The researcher decides to lag both per capita income and political change to address reverse causality concerns. The resulting model corresponds to the temporal DAG in Figure~2.
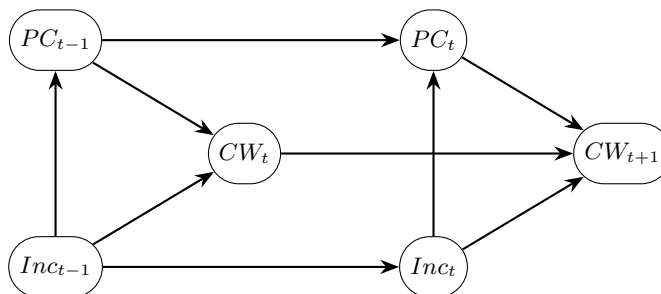

Figure 2: Temporal DAG for the effect of Political Change (PC) on Civil War (CW), with Per Capita Income (Inc) as a confounder. The temporal structure reveals that a lagged dependent variable (CW_t) is needed to block all backdoor paths to CW_{t+1}.

A key insight from this DAG is that the dynamic structure introduces backdoor paths through the lagged outcome (CW_t). To block these paths, the researcher must include CW_t as a control—leading to a dynamic panel (ADL) specification. Without the lagged dependent variable, the causal effect is biased.

3.3 Foreign Collider Bias

Now suppose the researcher, following the confounding checking heuristic, wants to add Democracy Level (Dem) as an additional control, as Hegre et al. (2001) suggest. At first glance, this seems harmless: democracy plausibly affects both political change and civil war. However, a more careful DAG reveals a critical problem.

It is plausible that civil war *causes* changes in democracy levels (countries that experience civil war tend to become less democratic). If, in addition, there exist unobserved factors U that affect both democracy levels and civil war—such as institutional quality, state capacity, or regional instability—then democracy level becomes a **collider**: it is a common effect of civil war and the unobserved U. Conditioning on it opens a backdoor path connecting U to civil war through democracy level. Figure~3 displays this structure.
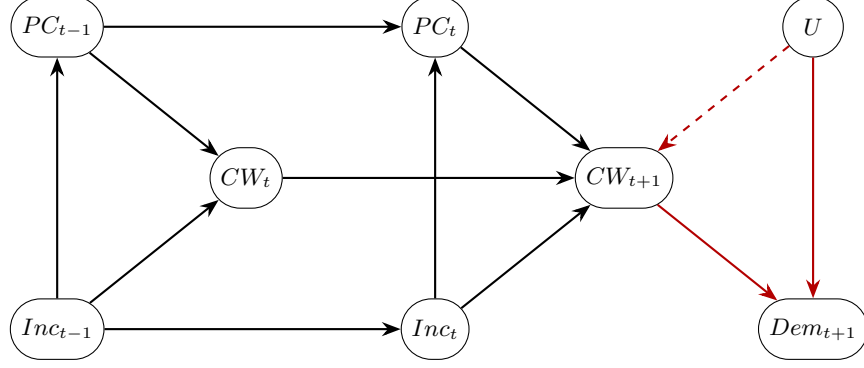
Figure 3: Complete DAG showing Democracy Level (Dem_{t+1}) as a collider. Civil War and an unobserved variable U both cause Democracy Level. Conditioning on Dem_{t+1} opens the collider $CW_{t+1} \rightarrow Dem_{t+1} \leftarrow U$, creating a spurious association between CW_{t+1} and U. Because U also causes CW_{t+1} (dashed arrow), this biases the estimate of the effect of Political Change on Civil War. Red arrows highlight the collider structure.

We call this *foreign collider bias* because the collider problem arises from a causal pathway that the researcher discovers only by looking at the "foreign'' literature on the determinants of democracy—a literature that is typically separate from the civil war literature motivating the original study. The standard heuristics for including controls review the literature on the *outcome* and the *treatment*, but they do not instruct the researcher to investigate whether the candidate control is itself caused by the treatment or outcome through an unobserved variable.

This example illustrates the power of DAGs: by making the causal structure explicit, they reveal that Democracy Level is a collider and that conditioning on it biases the regression. But the DAG alone does not tell us *how much* bias results. For that, we need a formula.

4 The Included Variable Bias Formula

In this section, we derive the Included Variable Bias (IVB) formula for the bias introduced by conditioning on a collider variable. We begin with the cross-sectional case, then compare IVB with the classical OVB formula, and finally extend the result to ADL models used in TSCS analysis.

4.1 Cross-Section Derivation

Consider the following setup. The true ("short'') model relating the outcome y to the treatment D is:

$$y = \beta_0 + \beta_1 D + e, \qquad \mathbb{E}[e \mid D] = 0 \tag{2}$$

where β_1 is the true causal effect of D on y.

Now suppose a researcher erroneously includes a collider variable Z in the regression, estimating the "long''
model:

$$y = \beta_0^\star + \beta_1^\star D + \beta_2^\star Z + e^\star \tag{3}$$

where the star superscript indicates coefficients from the misspecified model. Here $\mathbb{E}[e^\star \mid D] \neq 0$ in general,
because conditioning on the collider Z opens a spurious path.

The collider Z is caused by both D and y:

$$Z = \gamma_0 + \gamma_1 D + \gamma_2 y + u, \qquad \mathbb{E}[u \mid D] = 0, \ \mathbb{E}[u \mid y] = 0 \tag{4}$$

The population coefficients from OLS satisfy:

$$\beta_1 = \frac{\mathrm{Cov}(D, y)}{\mathrm{Var}(D)}, \qquad \beta_1^\star = \frac{\mathrm{Cov}(D, y) - \beta_2^\star \, \mathrm{Cov}(D, Z)}{\mathrm{Var}(D)}$$

Subtracting, the bias in the estimated treatment effect is:

$$\mathrm{Bias} = \beta_1^\star - \beta_1 = -\beta_2^\star \frac{\mathrm{Cov}(D, Z)}{\mathrm{Var}(D)} \tag{5}$$

To express this in terms of estimable quantities, substitute the true model (2) into the collider equation (4):

$$Z = \gamma_0 + \gamma_1 D + \gamma_2(\beta_0 + \beta_1 D + e) + u$$

$$= \underbrace{(\gamma_0 + \gamma_2 \beta_0)}_{\phi_0} + \underbrace{(\gamma_1 + \gamma_2 \beta_1)}_{\phi_1} D + \underbrace{(\gamma_2 e + u)}_{\varepsilon} \tag{6}$$

The reduced-form coefficient ϕ_1 is simply the population coefficient from regressing Z on D:

$$\phi_1 = \frac{\mathrm{Cov}(D, Z)}{\mathrm{Var}(D)}$$

Substituting into (5) yields the IVB formula:

Proposition 1 (IVB, Cross-Section). *When a collider variable Z is erroneously included in a regression of*

9

y on D, the bias in the treatment effect estimate is:

$$IVB = \beta_1^\star - \beta_1 = -\beta_2^\star \times \phi_1 \qquad (7)$$

where $\beta_2^\star$ is the coefficient on Z in the long regression (3) *and ϕ_1 is the coefficient from regressing Z on D.*

Both $\beta_2^\star$ and ϕ_1 are directly estimable from data: run the long regression to obtain $\hat{\beta}_2^\star$, and regress Z on D to obtain $\hat{\phi}_1$. No knowledge of the structural parameters γ_1 or γ_2 is needed.

We note that Proposition~1 follows algebraically from the Frisch–Waugh–Lovell theorem and the standard short-versus-long regression decomposition. Our contribution lies not in the algebra per se, but in (i)~naming and packaging this result as a diagnostic tool that mirrors OVB, (ii)~making the direct estimability of both components explicit, and (iii)~extending the result to the ADL models prevalent in TSCS research.

4.2   OVB versus IVB

The IVB formula has a striking structural parallel with the classical Omitted Variable Bias formula. Table~1 presents the comparison side by side.

Table 1: Comparison of Omitted Variable Bias (OVB) and Included Variable Bias (IVB)

	OVB	IVB
Problem	Omitting a confounder	Including a collider
Correct model	Long (includes confounder)	Short (excludes collider)
Wrong model	Short (omits confounder)	Long (includes collider)
Formula	$\delta_1 \times \gamma$	$-\beta_2^\star \times \phi_1$
Components	δ_1: effect of omitted var in long reg	$\beta_2^\star$: effect of collider in long reg
	γ: regression of omitted var on D	ϕ_1: regression of collider on D
Sign of bias	$\text{sign}(\delta_1) \times \text{sign}(\gamma)$	$-\text{sign}(\beta_2^\star) \times \text{sign}(\phi_1)$
Estimable?	Only if confounder is observed	Yes, both components from data

The key difference lies in the direction of the problem. With OVB, the researcher runs a short regression when the long regression is correct, and the bias equals the product of two regression coefficients involving the omitted variable. With IVB, the researcher runs a long regression when the short regression is correct, and the bias is the *negative* product of two analogous coefficients involving the included collider. The negative sign in the IVB formula arises because the mechanisms operate in opposite directions: OVB captures the effect

of an omitted common cause channeled through the treatment-outcome relationship, while IVB captures the spurious association created by conditioning on a common effect.

A notable practical advantage of IVB is that its components are *always* estimable: the researcher has already included Z in the regression (otherwise there would be no bias), so $\beta_2^\star$ is available, and ϕ_1 comes from a simple auxiliary regression. By contrast, OVB requires observing the omitted variable—which, by definition, is often unavailable.

4.3 Extension to ADL(1,0) via FWL

We now extend the IVB formula to the dynamic models used in TSCS analysis. Consider an ADL(1,0) specification. The correct ("short'') model is:

$$y_t = \alpha + \beta D_t + \rho y_{t-1} + e_t, \qquad \mathbb{E}[e_t \mid D_t, y_{t-1}] = 0 \tag{8}$$

The misspecified ("long'') model erroneously includes a collider Z_t:

$$y_t = \alpha^\star + \beta^\star D_t + \rho^\star y_{t-1} + \theta^\star Z_t + v_t \tag{9}$$

Our goal is to express $\beta^\star - \beta$ in terms of estimable quantities. The key tool is the Frisch–Waugh–Lovell (FWL) theorem (Frisch and Waugh, 1933; Lovell, 1963).

Define the set of "legitimate'' controls $W_t = (1, y_{t-1})$ and let $\tilde{y}_t$, $\tilde{D}_t$, and $\tilde{Z}_t$ denote the residuals from projecting each variable on W_t. By FWL, the long regression (9) is equivalent to:

$$\tilde{y}_t = \beta^\star \tilde{D}_t + \theta^\star \tilde{Z}_t + \tilde{v}_t$$

In this bivariate residualized regression, the coefficient on $\tilde{D}_t$ satisfies:

$$\beta^\star = \frac{\mathrm{Cov}(\tilde{D}_t, \tilde{y}_t) - \theta^\star \, \mathrm{Cov}(\tilde{D}_t, \tilde{Z}_t)}{\mathrm{Var}(\tilde{D}_t)}$$

The short model coefficient (also via FWL) is:

$$\beta = \frac{\mathrm{Cov}(\tilde{D}_t, \tilde{y}_t)}{\mathrm{Var}(\tilde{D}_t)}$$

11

Subtracting:

$$\beta^\star - \beta = -\theta^\star \frac{\text{Cov}(\tilde{D}_t, \tilde{Z}_t)}{\text{Var}(\tilde{D}_t)}$$

Now consider the auxiliary regression:

$$Z_t = a + \pi D_t + \lambda y_{t-1} + \eta_t \tag{10}$$

By the algebra of partial regression, $\pi = \text{Cov}(\tilde{D}_t, \tilde{Z}_t)/\text{Var}(\tilde{D}_t)$. Substituting:

Proposition 2 (IVB, ADL(1,0)). *In an ADL(1,0) model, when a collider Z_t is erroneously included, the bias in the treatment effect is:*

$$\boxed{IVB(\beta) = \beta^\star - \beta = -\theta^\star \times \pi} \tag{11}$$

where $\theta^\star$ is the coefficient on Z_t in the long regression (9) and π is the coefficient on D_t in the auxiliary regression (10).

The formula has the same structural form as the cross-sectional case. The only difference is that all quantities are now *partial* effects, conditioned on the legitimate controls $(1, y_{t-1})$. The dynamics enter through the residualization, but the final formula remains a simple product.

4.4 Generalization to ADL(p,q)

The result extends naturally to general ADL(p,q) specifications. Consider the correct model:

$$y_t = \alpha + \sum_{i=1}^{p} \rho_i y_{t-i} + \sum_{j=0}^{q} \beta_j D_{t-j} + e_t$$

and the misspecified model including Z_t:

$$y_t = \alpha^\star + \sum_{i=1}^{p} \rho_i^\star y_{t-i} + \sum_{j=0}^{q} \beta_j^\star D_{t-j} + \theta^\star Z_t + v_t$$

Let W_t denote the vector of all legitimate controls (constant, $y_{t-1}, \ldots, y_{t-p}$, and the relevant lags of D other than D_{t-j}). The FWL argument carries through identically:

Proposition 3 (IVB, ADL(p,q)). *For any coefficient β_j in an ADL(p,q) model, the bias from including a collider Z_t is:*

$$\boxed{IVB(\beta_j) = \beta_j^\star - \beta_j = -\theta^\star \times \pi_j} \tag{12}$$

where π_j is the coefficient on D_{t-j} in the auxiliary regression $Z_t = a + \pi_j D_{t-j} + W_t'\delta + \eta_t$.

The generality of this result is noteworthy. Regardless of the number of lags in the outcome or the treatment, the IVB for any coefficient β_j is always the negative product of two estimable quantities: the collider coefficient $\theta^\star$ in the long model, and the partial association π_j between the treatment at lag j and the collider, controlling for all other legitimate regressors.

An important practical observation is that while the *form* of the bias is simple, the *magnitude* can vary substantially depending on the dynamics. When the treatment D_t is highly predictable from lags, the residualized treatment $\tilde{D}_t$ has little variation, and π can become unstable. When Z_t is a post-treatment variable that responds to y or to correlated shocks, both $\theta^\star \neq 0$ and $\pi \neq 0$ tend to arise mechanically.

A further consideration arises in panels with unit fixed effects and short time dimensions. In such settings, the OLS estimates of ADL coefficients suffer from Nickell bias (Nickell, 1981), which arises because the demeaned lagged dependent variable is mechanically correlated with the demeaned error term. The IVB formula remains an algebraic identity even in this setting—it correctly decomposes the difference between the short and long regression coefficients—but the components $\theta^\star$ and π will themselves reflect both collider bias and Nickell bias. Researchers using fixed-effects ADL models should therefore be aware that the IVB diagnostic captures collider bias within the OLS framework as estimated, and that separate corrections for dynamic panel bias (such as GMM estimators) may be needed before the IVB components can be given a purely causal interpretation.

4.5 Lagging the Treatment Is Not a Mechanical Cure

Applied work sometimes replaces a contemporaneous treatment D_t with a longer lag (e.g., D_{t-3}) under the intuition that this should make collider bias less likely. The IVB decomposition clarifies that this is not true in general.

Proposition 4 (Lag Substitution Without Lagged Outcome)**.** *Consider a model without lagged dependent variables. For any target lag $k \geq 0$, define:*

$$y_t = \alpha + \beta_k D_{t-k} + W_t'\gamma + e_t$$

as the correct (short) model, and

$$y_t = \alpha^\star + \beta_k^\star D_{t-k} + W_t'\gamma^\star + \theta^\star Z_t + v_t$$

as the long model including collider Z_t. Then:

$$\beta_k^\star - \beta_k = -\theta^\star \pi_k,$$

where π_k is the coefficient on D_{t-k} in

$$Z_t = a + \pi_k D_{t-k} + W_t'\delta + \eta_t.$$

Now suppose the collider process is:

$$Z_t = a + \pi_0 D_t + \pi_1 D_{t-1} + W_t'\delta + \eta_t.$$

Let $\widetilde{X}$ denote residualization of X on W_t, and define

$$\widetilde{D}_{t-\ell} = \omega_{\ell k}\widetilde{D}_{t-k} + r_{\ell t}, \qquad \ell \in \{0,1\},$$

with $\mathrm{Cov}(\widetilde{D}_{t-k}, r_{\ell t}) = 0$. Then:

$$\pi_k = \pi_0\omega_{0k} + \pi_1\omega_{1k},$$

so that

$$\beta_k^\star - \beta_k = -\theta^\star \left(\pi_0\omega_{0k} + \pi_1\omega_{1k}\right). \tag{13}$$

In particular, for $k = 3$:

$$\beta_3^\star - \beta_3 = -\theta^\star \left(\pi_0\omega_{03} + \pi_1\omega_{13}\right).$$

Equation~(13) shows that moving from D_{t-1} to D_{t-3} does not mechanically eliminate IVB. Bias disappears only in special cases, such as $\theta^\star = 0$, $\omega_{03} = \omega_{13} = 0$, or knife-edge cancellation $\pi_0\omega_{03} + \pi_1\omega_{13} = 0$.

More generally, if

$$Z_t = a + \sum_{\ell=0}^{L} \pi_\ell D_{t-\ell} + W_t'\delta + \eta_t,$$

then for any target lag k:

$$\pi_k = \sum_{\ell=0}^{L} \pi_\ell\omega_{\ell k} \quad \Rightarrow \quad \beta_k^\star - \beta_k = -\theta^\star \sum_{\ell=0}^{L} \pi_\ell\omega_{\ell k}.$$

Hence, longer lags can attenuate IVB when the $\omega_{\ell k}$ terms are small, but they do not guarantee zero collider

bias.

4.6 A Practical Recipe

We summarize the IVB formula as a simple three-step procedure for the applied researcher:

1. **Run the long regression.** Estimate the model that includes the suspected collider Z:

$$y = \ldots + \theta^\star Z + \text{other controls} + v$$

   Record $\hat{\theta}^\star$.

2. **Run the auxiliary regression.** Regress the suspected collider on the treatment and all legitimate controls:

$$Z = \ldots + \pi D + \text{same controls as in step 1, minus } Z + \eta$$

   Record $\hat{\pi}$.

3. **Compute the bias.** The estimated IVB is:

$$\widehat{\text{IVB}} = -\hat{\theta}^\star \times \hat{\pi}$$

If $|\widehat{\text{IVB}}|$ is large relative to the estimated treatment effect, the researcher has strong reason to exclude the collider from the specification. The sign of the IVB indicates the *direction* of the bias: whether including the collider attenuates or amplifies the treatment effect.

4.7 Interpretation Caveats

An important caveat concerns the interpretation of the IVB formula. The decomposition $\beta_1^\star - \beta_1 = -\beta_2^\star \times \phi_1$ is an algebraic identity: it holds whenever the short and long regressions are nested, regardless of the causal role of Z. The product $\beta_2^\star \times \phi_1$ will be non-zero whenever Z is correlated with both D and y conditional on other controls—a condition satisfied not only when Z is a collider, but also when Z is a confounder or a mediator.

The *causal* interpretation of this quantity—that it represents collider bias, and that the short regression recovers the correct causal effect—requires prior knowledge that Z is indeed a collider. This knowledge must come from a DAG or from substantive domain expertise; the formula alone cannot supply it. If Z is actually a confounder, the same algebraic decomposition gives the omitted variable bias in the short regression, and

it is the *long* regression that is correctly specified. If Z is a mediator, neither specification may recover the desired causal estimand without further assumptions.

In short, the IVB formula complements DAGs but cannot substitute for them. The DAG tells the researcher *whether* Z is a collider, and only then does the formula quantify the damage from conditioning on it. Researchers should therefore use the three-step recipe from Section~4.6 only after having established—through causal reasoning, not statistical tests—that the variable in question is plausibly a collider.

5   Monte Carlo Validation

We validate the IVB formula through Monte Carlo simulations across three data-generating processes (DGPs): a pure cross-section, an ADL(1,0) panel, and a substantive civil war DGP. In every case, we compare the empirical bias (the difference between the long and short regression estimates) to the bias predicted by the IVB formula. We run 500 replications for each DGP.

5.1   DGP 1: Cross-Section

The first DGP is a simple cross-sectional model with $n = 10,000$ observations. The treatment effect is $\beta_1 = 1$, and the collider is generated as $Z = 0.6D + 0.4Y + u$, where $u \sim \mathcal{N}(0,1)$.

5.2   DGP 2: ADL(1,0) Panel

The second DGP is an ADL(1,0) panel with $N = 200$ units and $T = 20$ periods. The treatment effect is $\beta = 1$, the autoregressive coefficient is $\rho = 0.5$, and the collider is generated as $Z_t = 0.6D_t + 0.4Y_t + u_t$.

5.3   DGP 3: Civil War with Foreign Collider Bias

The third DGP is a substantive model based on the civil war example from Sections 3 and 6. The true effect of Political Change on Civil War is $\beta_{PC} = 5$. The model includes Per Capita Income as a confounder and a lagged dependent variable. Democracy Level is generated as a collider, caused by Civil War ($\gamma_{CW} = -0.6$) and an unobserved variable U ($\gamma_U = 1.0$) that also affects Civil War ($\gamma_{U \to CW} = 0.5$).

5.4   Results

Figure~4 displays the cross-sectional results. Each point represents one simulation replication, with the formula-predicted bias on the horizontal axis and the empirical bias on the vertical axis. All points fall exactly on the 45-degree line, confirming that the IVB formula predicts the bias with exact precision.
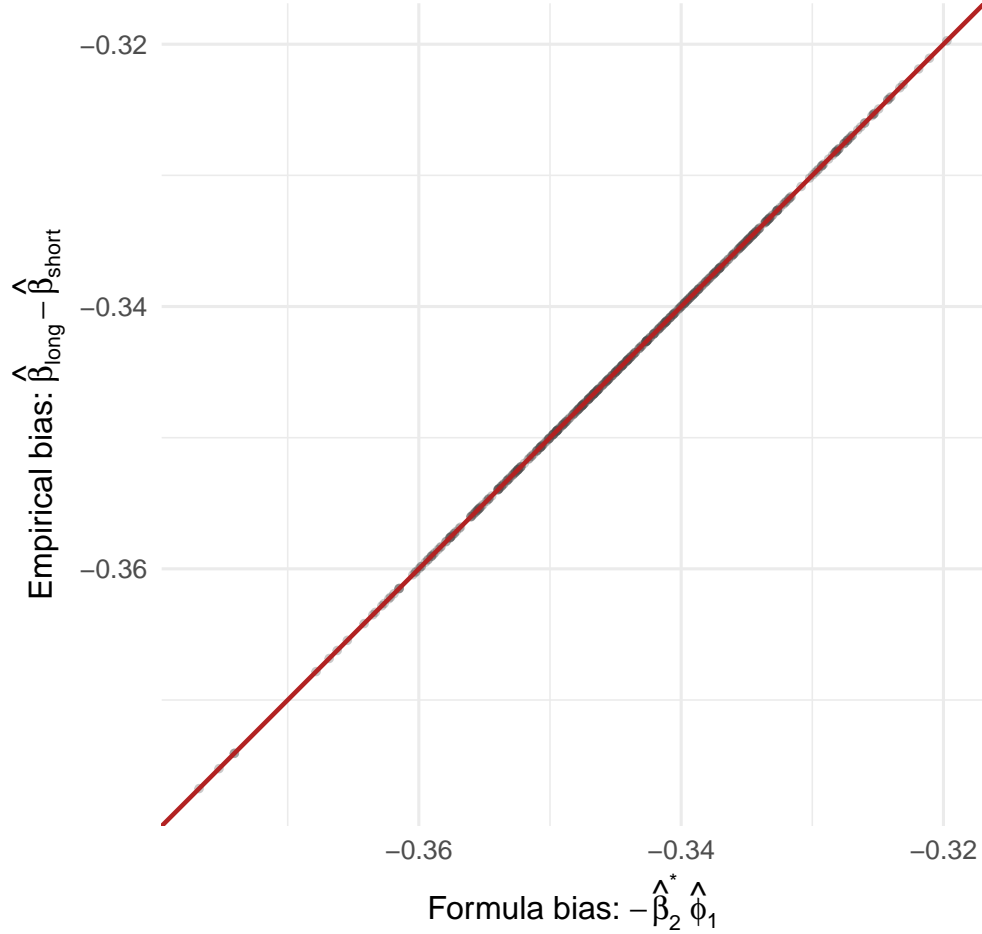
Figure 4: Cross-section: Empirical bias versus IVB formula prediction across 500 simulations. All points fall on the 45-degree line, confirming the formula's exactness.

Table~2 summarizes the results across all three DGPs. In every scenario, the mean empirical bias and the mean formula-predicted bias coincide to at least three decimal places.

Table 2: Monte Carlo summary across three DGPs (500 replications each). The IVB formula predicts the empirical bias exactly in every scenario.

Scenario	True effect	Mean estimate (with collider)	Mean empirical bias	Mean formula bias
Cross-section	1	0.656	-0.344	-0.344
ADL(1,0)	1	0.654	-0.346	-0.346
Civil War	5	4.596	-0.404	-0.404

Figure~5 shows the distribution of the estimated effect of Political Change on Civil War with and without Democracy Level as a control. The correct specification (without the collider) is centered on the true effect of 5, while the misspecified model (with the collider) is shifted to the left, reflecting the negative IVB introduced by conditioning on Democracy Level.
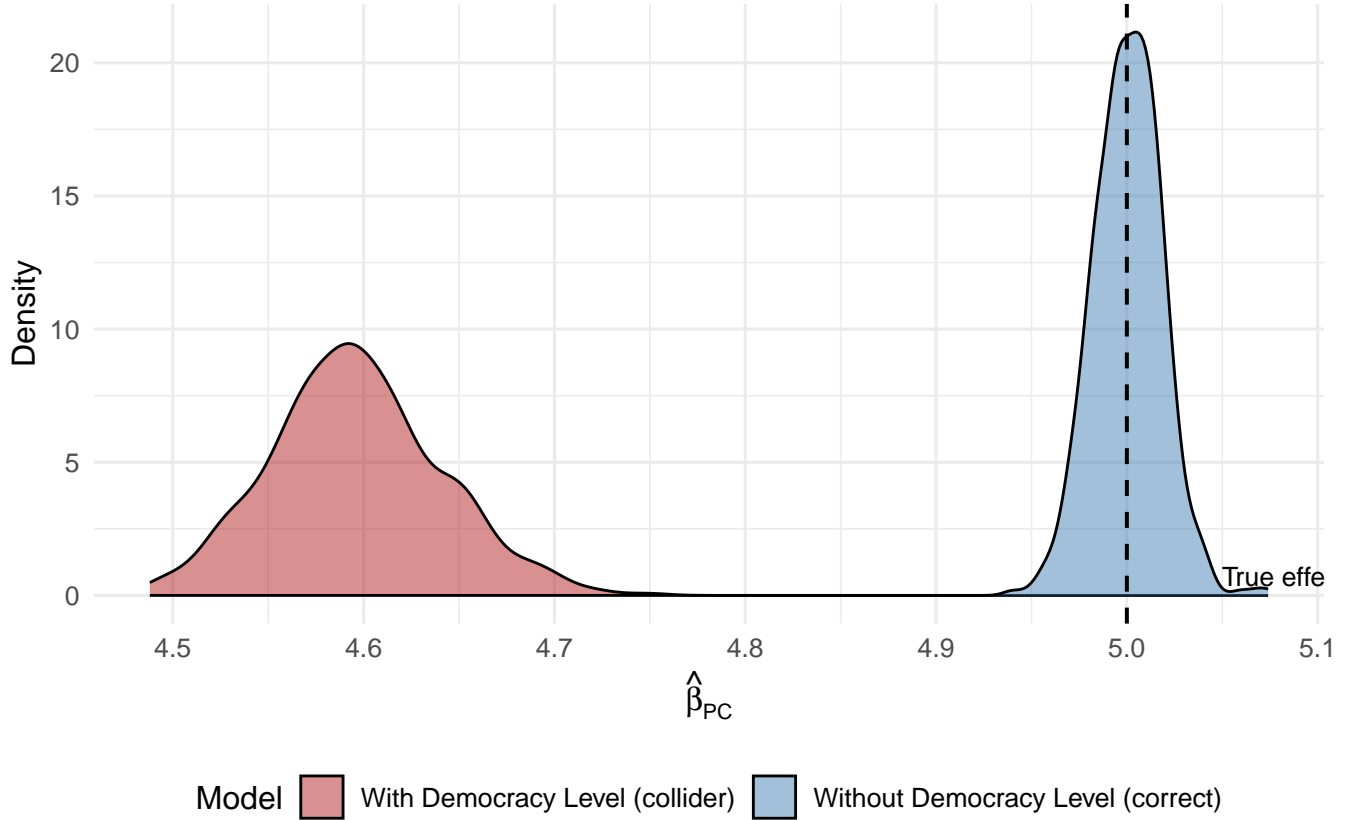
Figure 5: Civil War DGP: Distribution of the estimated effect of Political Change on Civil War across 500 simulations. The correct specification (without Democracy Level, in blue) is centered on the true effect of 5. The misspecified model (with Democracy Level as collider, in red) is biased downward.

The results are unambiguous: the IVB formula $\widehat{\text{IVB}} = -\hat{\theta}^\star \times \hat{\pi}$ matches the empirical bias *exactly* in every replication and every DGP. This is not an approximation; it is an algebraic identity that holds in finite samples whenever the OLS estimates are computed on the same data.

6 Application: Political Change and Civil War

We now return to the civil war example to illustrate how a researcher would apply the IVB formula in practice. The scenario is as follows: the researcher has estimated an ADL model of Civil War on Political Change, controlling for Per Capita Income and lagged Civil War. A reviewer suggests adding Democracy Level as an additional control. The researcher suspects that Democracy Level may be a collider but wants to quantify the potential damage.

Following the three-step recipe from Section~4.6:

Step 1: Run the long regression. The researcher estimates:

$$CW_{t+1} = \alpha^\star + \beta^\star_{PC} \, PC_t + \rho^\star \, CW_t + \alpha^\star_{Inc} \, Inc_t + \theta^\star \, Dem_{t+1} + v_t$$

In this simulation draw, $\hat{\theta}^\star = -0.125$. The coefficient is negative, reflecting the fact that civil war tends to reduce democracy levels. Including Democracy Level in the regression partially "controls away'' the variation in Civil War that operates through this channel, distorting the estimate of the Political Change effect.

Step 2: Run the auxiliary regression. The researcher estimates:

$$Dem_{t+1} = a + \pi \, PC_t + \lambda \, CW_t + \delta \, Inc_t + \eta_t$$

The estimated coefficient is $\hat{\pi} = -2.971$. This coefficient captures the partial association between Political Change and Democracy Level, net of lagged Civil War and Income.

Step 3: Compute the IVB. The estimated bias is:

$$\widehat{\text{IVB}} = -\hat{\theta}^\star \times \hat{\pi} = -(-0.125) \times (-2.971) = -0.371$$

We can verify: the correct estimate (without Democracy Level) is $\hat{\beta}_{PC} = 5.007$, and the collider-contaminated estimate is $\hat{\beta}^\star_{PC} = 4.636$. The difference is $4.636 - 5.007 = -0.371$, which matches the IVB formula prediction.

This example illustrates several practical lessons. First, the IVB formula allows researchers to *quantify* the bias from including a suspected collider, providing concrete numbers rather than a qualitative warning. Second, the sign of the bias is informative: in this case, including Democracy Level *attenuates* the estimated effect of Political Change on Civil War, pushing it toward zero. A researcher who inadvertently includes Democracy Level would underestimate the true effect. Third, the three-step procedure is simple enough to be implemented as a routine diagnostic check whenever a variable's causal status is uncertain.

More broadly, the IVB formula serves as a bridge between DAG-based qualitative reasoning and the quantitative demands of applied research. The DAG tells the researcher that Democracy Level is a collider; the formula tells her exactly how much damage including it would cause.

7 Conclusion

This paper makes three contributions to the methodology of causal inference in observational studies. First, we derive the Included Variable Bias (IVB) formula, a closed-form expression for the bias introduced when a researcher erroneously includes a collider variable in a regression model. The formula, $\text{IVB} = -\theta^\star \times \pi$, is the mirror image of the classic Omitted Variable Bias formula and shares its appealing properties: it is a simple product of two estimable regression coefficients, it reveals both the magnitude and the direction of the bias, and it is immediately operational for applied researchers.

Second, we show that the IVB formula extends naturally from cross-sectional models to the Autoregressive Distributed Lag (ADL) specifications that are the workhorse of time-series cross-sectional research in political science. The extension relies on the Frisch–Waugh–Lovell theorem and preserves the same functional form, with the key quantities now partialled out by the dynamic controls. This result is particularly relevant because TSCS settings are especially susceptible to collider bias: the practice of lagging control variables to avoid reverse causality can inadvertently create collider structures, a phenomenon we call "foreign collider bias.''

Third, we provide a simple three-step diagnostic recipe that any applied researcher can implement. Whenever a variable's causal status is uncertain—is it a confounder that should be included, or a collider that should be excluded?—the researcher can run the long regression, run the auxiliary regression, and compute the IVB. If the estimated bias is large relative to the treatment effect, the researcher has quantitative grounds for excluding the variable.

Our work connects to and extends several strands of the literature. It complements Cinelli, Forney and Pearl (2022), who provide a systematic taxonomy of good and bad controls using DAGs, by adding a quantitative dimension: not just *whether* a control is bad, but *how bad* it is. It complements Blackwell and Glynn (2018), who propose sequential g-estimation and marginal structural models as solutions to post-treatment bias in TSCS, by providing a diagnostic formula that does not require the researcher to adopt a different estimation framework. And it extends the OVB tradition exemplified by Cinelli and Hazlett (2020), who develop sensitivity analysis tools for omitted variable bias, to the symmetric problem of included variable bias.

Several limitations deserve mention. The IVB formula is derived for linear models, which covers a large share of applied TSCS research—including OLS with panel-corrected standard errors, two-way fixed effects, and difference-in-differences designs with continuous outcomes. However, the Frisch–Waugh–Lovell theorem does not hold for nonlinear link functions, so the formula as stated does not apply to logistic or probit regression.

This matters for studies with binary outcomes, such as civil war onset, where nonlinear specifications are standard. Future work should investigate whether the linear IVB provides a useful approximation in such settings or whether a separate nonlinear formula is needed. The formula also quantifies bias under the assumption that the researcher *knows* which variable is the collider; it does not help with the prior step of identifying colliders, for which DAGs remain essential. Additionally, we do not develop a formal sensitivity analysis framework for IVB—an extension that would allow researchers to assess how sensitive their results are to potential collider bias even when the collider is not directly observed.

This last limitation points to a natural direction for future work. Just as Cinelli and Hazlett (2020) developed sensitivity analysis tools that extend the OVB formula to scenarios where the omitted variable is unobserved, a parallel development for IVB would allow researchers to ask: "How strong would the collider relationship need to be to change my conclusions?'' We leave this extension for future research.

In the meantime, we encourage applied researchers to add the IVB formula to their diagnostic toolkit. The combination of DAGs for qualitative reasoning and the IVB formula for quantitative assessment provides a powerful framework for making informed decisions about control variable selection—decisions that are critical for the credibility of causal claims in observational research.

References

Beck, Nathaniel and Jonathan N. Katz. 1995. "What to Do (and Not to Do) with Time-Series Cross-Section Data." *American Political Science Review* 89(3):634–647.

Blackwell, Matthew and Adam N. Glynn. 2018. "How to Make Causal Inferences with Time-Series Cross-Sectional Data under Selection on Observables." *American Political Science Review* 112(4):1067–1082.

Blattman, Christopher and Edward Miguel. 2010. "Civil War." *Journal of Economic Literature* 48(1):3–57.

Cinelli, Carlos, Andrew Forney and Judea Pearl. 2022. "A Crash Course in Good and Bad Controls." *Sociological Methods & Research* 51(3):1071–1104.

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

Dietrich, Simone. 2016. "Donor Political Economies and the Pursuit of Aid Effectiveness." *International Organization* 70(1):65–102.

Ding, Peng and Luke W. Miratrix. 2015. "To Adjust or Not to Adjust? Sensitivity Analysis of M-Bias and Butterfly-Bias." *Journal of Causal Inference* 3(1):41–57.

Duarte, Guilherme, Noam Finkelstein, Dean Knox, Jonathan Mummolo and Ilya Shpitser. 2021. "An Automated Approach to Causal Inference in Discrete Settings." *arXiv preprint arXiv:2109.13471* .

Elwert, Felix and Christopher Winship. 2014. "Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable." *Annual Review of Sociology* 40:31–53.

Franzese, Robert J. and Jude C. Hays. 2007. "Spatial Econometric Models of Cross-Sectional Interdependence in Political Science Panel and Time-Series-Cross-Section Data." *Political Analysis* 15(2):140–164.

Frisch, Ragnar and Frederick V. Waugh. 1933. "Partial Time Regressions as Compared with Individual Trends." *Econometrica* 1(4):387–401.

Glymour, Clark, Kun Zhang and Peter Spirtes. 2019. "Review of Causal Discovery Methods Based on Graphical Models." *Frontiers in Genetics* 10:524.

Greenland, Sander. 2003. "Quantifying Biases in Causal Models: Classical Confounding vs Collider-Stratification Bias." *Epidemiology* 14(3):300–306.

Hahn, Jinyong. 2004. "Functional Restriction and Efficiency in Causal Inference." *Review of Economics and Statistics* 86(1):73–76.

Heckman, James J. 2008. "Econometric Causality." *International Statistical Review* 76(1):1–27.

Hegre, Håvard, Tanja Ellingsen, Scott Gates and Nils Petter Gleditsch. 2001. "Toward a Democratic Civil Peace? Democracy, Political Change, and Civil War, 1816–1992." *American Political Science Review* 95(1):33–48.

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.

Jung, Jongbin, Sam Corbett-Davies, Johann D. Gaebler, Ravi Shroff and Sharad Goel. 2024. "Mitigating Included- and Omitted-Variable Bias in Estimates of Disparate Impact." arXiv:1809.05651v4.

Lovell, Michael C. 1963. "Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis." *Journal of the American Statistical Association* 58(304):993–1010.

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

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49(6):1417–1426.

Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference.* 2nd ed. Cambridge University Press.

Pearl, Judea. 2013. "Linear Models: A Useful "Microscope" for Causal Analysis." *Journal of Causal Inference* 1(1):155–170.

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

Spirtes, Peter and Kun Zhang. 2016. Causal Discovery and Inference: Concepts and Recent Methodological Advances. In *Applied Informatics.* Vol. 3 SpringerOpen pp. 1–28.

Wright, Philip G. 1928. *Tariff on Animal and Vegetable Oils.* Macmillan Company.

Wright, Sewall. 1934. "The Method of Path Coefficients." *The Annals of Mathematical Statistics* 5(3):161–215.

Yao, Liuyi, Zhixuan Chu, Sheng Li, Yaliang Li, Jing Gao and Aidong Zhang. 2021. "A Survey on Causal Inference." *ACM Transactions on Knowledge Discovery from Data* 15(5):1–46.

A Online Appendix

A.1 Appendix A: Full Cross-Section Derivation

We provide the step-by-step algebra for the cross-section IVB derivation.

Setup. The true model is $y = \beta_0 + \beta_1 D + e$ with $\mathbb{E}[e \mid D] = 0$. The wrong model includes a collider Z: $y = \beta_0^\star + \beta_1^\star D + \beta_2^\star Z + e^\star$. The collider relationship is $Z = \gamma_0 + \gamma_1 D + \gamma_2 y + u$ with $\mathbb{E}[u \mid D, y] = 0$.

Step 1. From the standard OLS formula for a multivariate regression:

$$\beta_1^\star = \frac{\mathrm{Cov}(D, y) \cdot \mathrm{Var}(Z) - \mathrm{Cov}(D, Z) \cdot \mathrm{Cov}(Z, y)}{\mathrm{Var}(D) \cdot \mathrm{Var}(Z) - [\mathrm{Cov}(D, Z)]^2}$$

However, it is more convenient to use the Frisch–Waugh–Lovell approach. By FWL, $\beta_1^\star$ equals the coefficient from regressing y on D after partialling out Z, or equivalently:

$$\beta_1^\star = \frac{\mathrm{Cov}(D, y) - \beta_2^\star \mathrm{Cov}(D, Z)}{\mathrm{Var}(D)}$$

This gives:

$$\beta_1^\star - \beta_1 = -\beta_2^\star \frac{\mathrm{Cov}(D, Z)}{\mathrm{Var}(D)}$$

Step 2. Substitute $y = \beta_0 + \beta_1 D + e$ into $Z = \gamma_0 + \gamma_1 D + \gamma_2 y + u$:

$$Z = \gamma_0 + \gamma_1 D + \gamma_2(\beta_0 + \beta_1 D + e) + u$$

$$= (\gamma_0 + \gamma_2 \beta_0) + (\gamma_1 + \gamma_2 \beta_1)D + (\gamma_2 e + u)$$

Define $\phi_0 = \gamma_0 + \gamma_2 \beta_0$, $\phi_1 = \gamma_1 + \gamma_2 \beta_1$, and $\varepsilon = \gamma_2 e + u$.

Step 3. Since $\mathbb{E}[\varepsilon \mid D] = \gamma_2 \mathbb{E}[e \mid D] + \mathbb{E}[u \mid D] = 0$:

$$\frac{\mathrm{Cov}(D, Z)}{\mathrm{Var}(D)} = \phi_1$$

Step 4. Therefore:

$$\mathrm{IVB} = \beta_1^\star - \beta_1 = -\beta_2^\star \times \phi_1 \qquad \blacksquare$$

A.2 Appendix B: Full FWL/ADL Derivation

We provide the complete derivation for the ADL(1,0) case using FWL.

Setup. The correct model is $y_t = \alpha + \beta D_t + \rho y_{t-1} + e_t$ with $\mathbb{E}[e_t \mid D_t, y_{t-1}] = 0$. The wrong model includes a collider: $y_t = \alpha^\star + \beta^\star D_t + \rho^\star y_{t-1} + \theta^\star Z_t + v_t$.

Step 1 (FWL residualization). Let $W_t = (1, y_{t-1})$. Define residuals:

$$\tilde{y}_t := y_t - \hat{\Pi}_y W_t$$

$$\tilde{D}_t := D_t - \hat{\Pi}_D W_t$$

$$\tilde{Z}_t := Z_t - \hat{\Pi}_Z W_t$$

By FWL, the long regression of y_t on D_t, y_{t-1}, and Z_t is equivalent to the regression of $\tilde{y}_t$ on $\tilde{D}_t$ and $\tilde{Z}_t$.

Step 2 (Two-variable regression formula). In the residualized regression:

$$\beta^\star = \frac{\text{Cov}(\tilde{D}_t, \tilde{y}_t) - \theta^\star \text{Cov}(\tilde{D}_t, \tilde{Z}_t)}{\text{Var}(\tilde{D}_t)}$$

The short model (by FWL applied to the regression of y_t on D_t and y_{t-1}) gives:

$$\beta = \frac{\text{Cov}(\tilde{D}_t, \tilde{y}_t)}{\text{Var}(\tilde{D}_t)}$$

Step 3 (Bias). Subtracting:

$$\beta^\star - \beta = -\theta^\star \frac{\text{Cov}(\tilde{D}_t, \tilde{Z}_t)}{\text{Var}(\tilde{D}_t)}$$

Step 4 (Auxiliary regression). Consider $Z_t = a + \pi D_t + \lambda y_{t-1} + \eta_t$. By FWL, π equals the coefficient from regressing $\tilde{Z}_t$ on $\tilde{D}_t$:

$$\pi = \frac{\text{Cov}(\tilde{D}_t, \tilde{Z}_t)}{\text{Var}(\tilde{D}_t)}$$

Step 5 (Result).

$$\text{IVB}(\beta) = \beta^\star - \beta = -\theta^\star \times \pi \qquad \blacksquare$$

ADL(_p_,_q_) extension. For general ADL(_p_,_q_), the legitimate control set becomes $W_t = (1, y_{t-1}, \ldots, y_{t-p}, D_{t-1}, \ldots, D_{t-q})$ (excluding D_{t-j} when computing the bias for β_j). The FWL argument carries through identically, yielding $\text{IVB}(\beta_j) = -\theta^\star \times \pi_j$ where π_j is the coefficient on D_{t-j} in the auxiliary regression $Z_t = a + \pi_j D_{t-j} + W_t' \delta + \eta_t$.
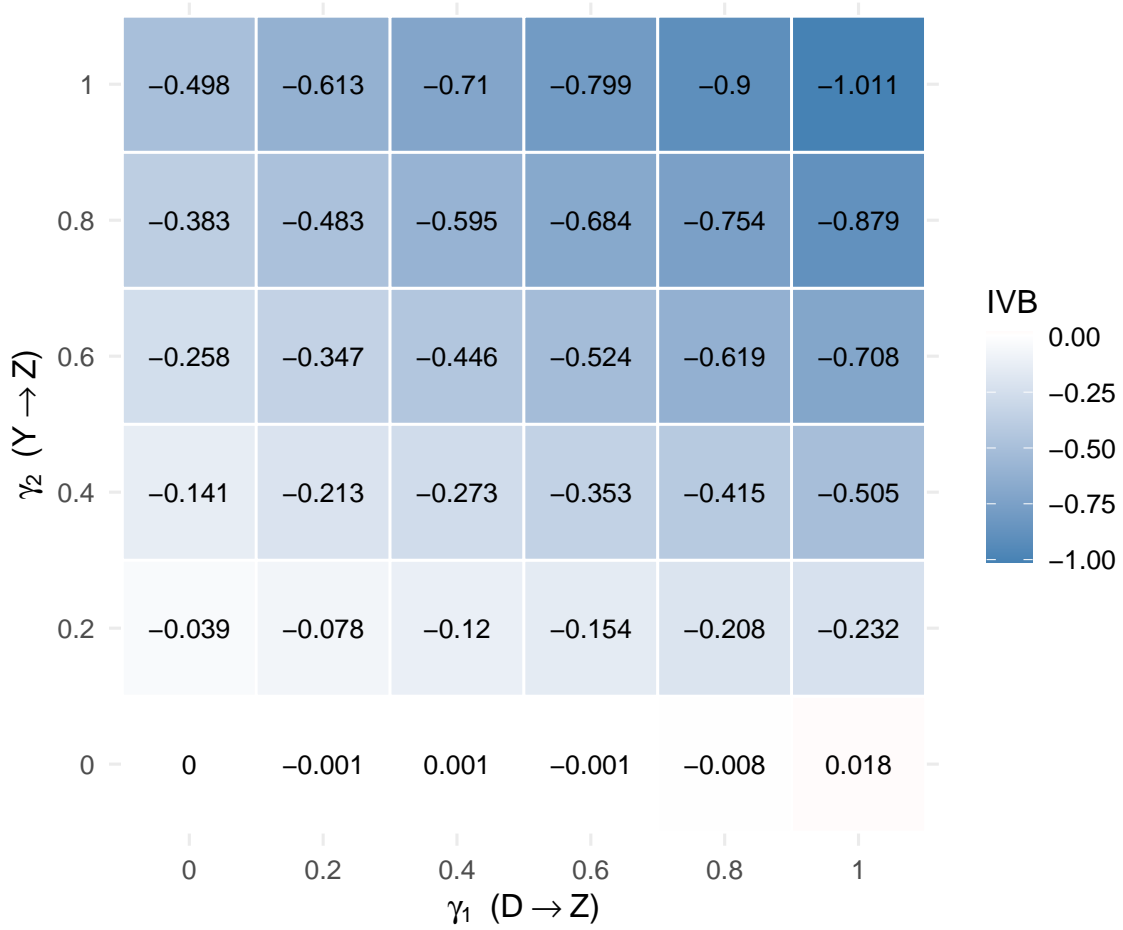
A.3 Appendix C: Additional Simulation Plots

Figure A1: Cross-section IVB as a function of the collider parameters γ_1 (D → Z) and γ_2 (Y → Z). The bias increases in magnitude with both parameters.
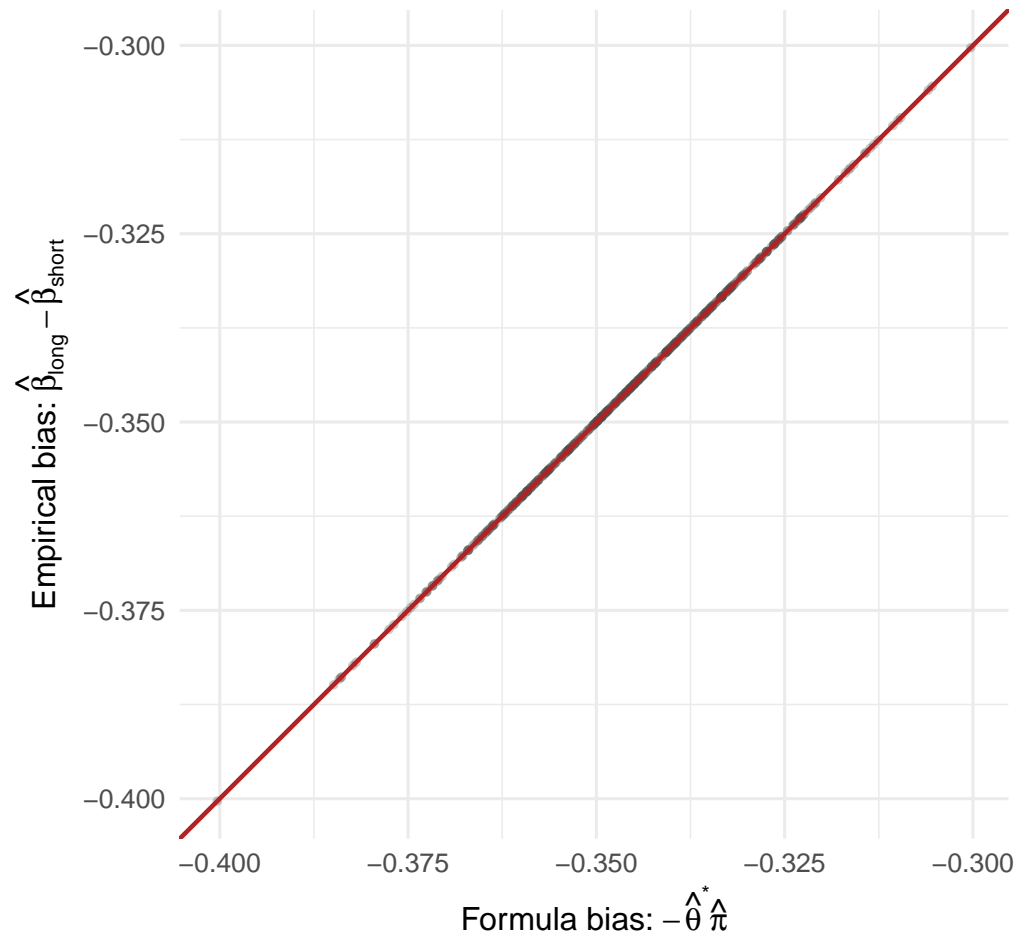
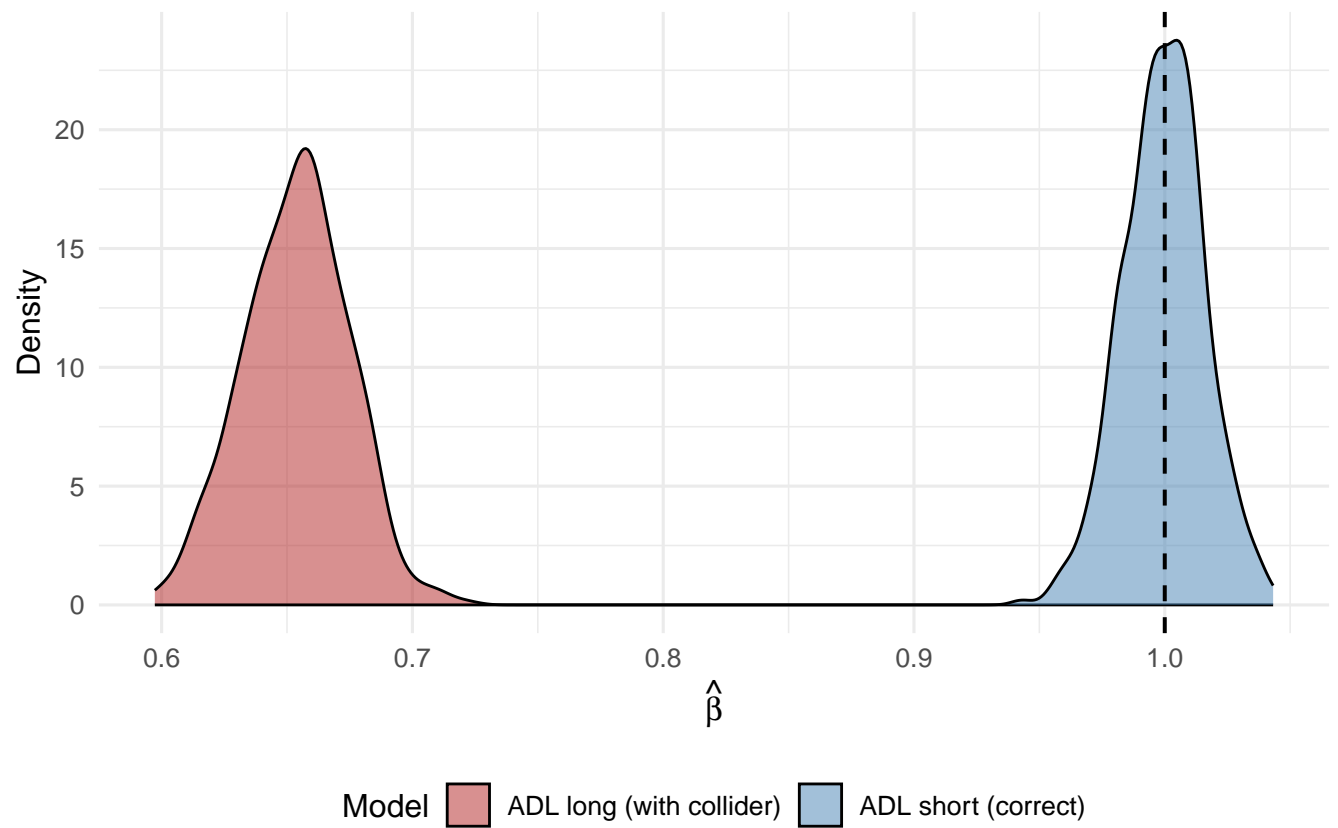Figure A2: ADL(1,0): Empirical bias versus IVB formula prediction across 500 simulations.

Figure A3: ADL(1,0): Distribution of treatment effect estimates with and without the collider across 500 simulations.
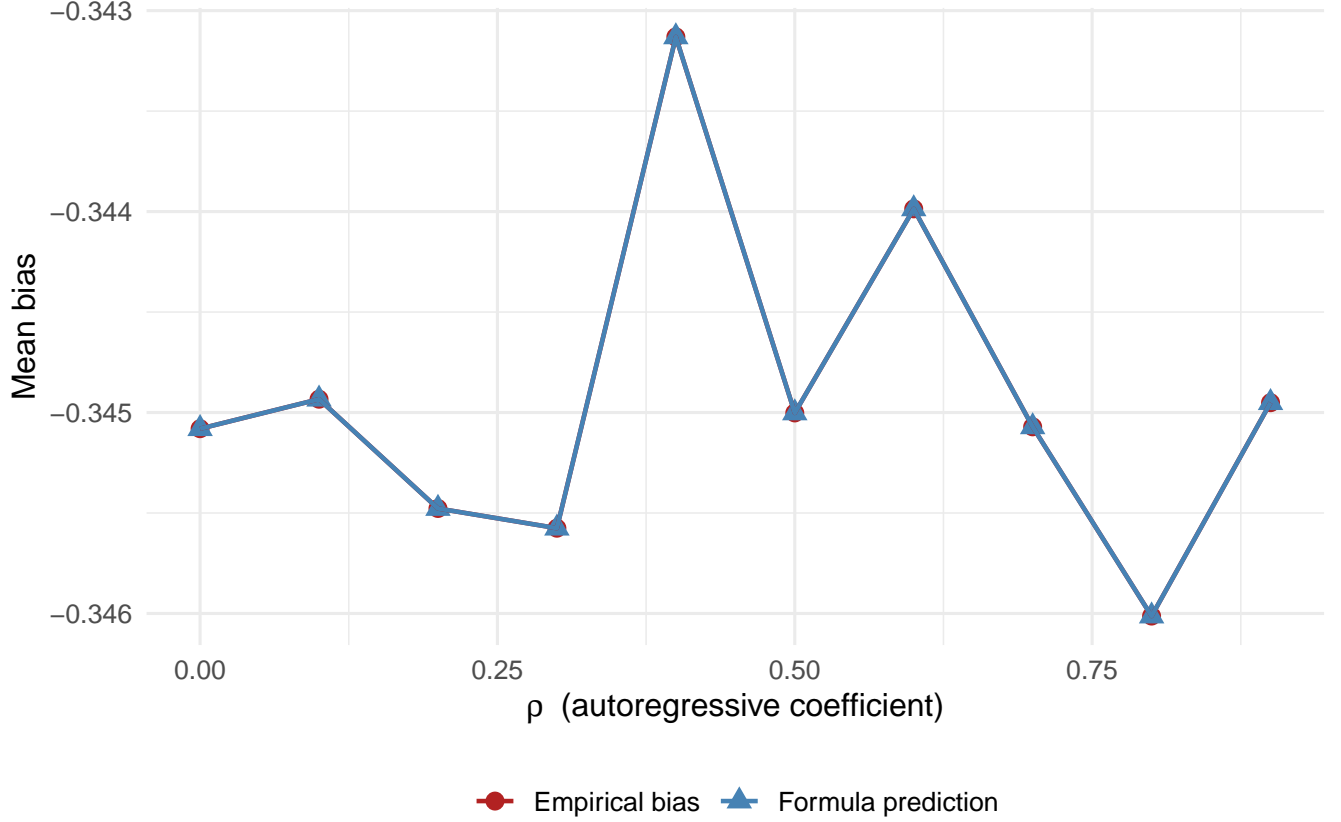
Figure A4: ADL(1,0): Effect of the autoregressive parameter ρ on the magnitude of the IVB. Both the empirical bias and the formula prediction are shown.

A.4   Appendix D: Simulation Code Description

All simulations in this paper are implemented in R and embedded in the R Markdown source file. The code is fully reproducible: all random number generators are seeded, and results can be exactly replicated by compiling the source document.

DGP 1 (Cross-section): We generate $n = 10{,}000$ observations per replication with $D \sim \mathcal{N}(0,1)$, $e \sim \mathcal{N}(0,1)$, and $u \sim \mathcal{N}(0,1)$. The outcome is $Y = \beta_1 D + e$ with $\beta_1 = 1$. The collider is $Z = \gamma_1 D + \gamma_2 Y + u$ with $\gamma_1 = 0.6$ and $\gamma_2 = 0.4$.

DGP 2 (ADL): We simulate a panel with $N = 200$ units and $T = 20$ periods. The outcome follows $Y_t = \rho Y_{t-1} + \beta D_t + e_t$ with $\rho = 0.5$ and $\beta = 1$. The collider is $Z_t = \delta_d D_t + \delta_y Y_t + u_t$ with $\delta_d = 0.6$ and $\delta_y = 0.4$.

DGP 3 (Civil War): We simulate a richer panel with $N = 200$ units and $T = 20$ periods. The DGP includes Political Change (PC), Civil War (CW), Per Capita Income (Inc), Democracy Level (Dem), and

an unobserved variable U. The key parameters are: $\beta_{PC} = 5$, $\rho_{CW} = 0.3$, $\rho_{PC} = 0.4$, $\alpha_{Inc} = 0.5$, $\rho_{Inc} = 0.8$, $\gamma_{CW \to Dem} = -0.6$, $\gamma_{U \to Dem} = 1.0$, $\gamma_{U \to CW} = 0.5$.

The full simulation code is available in the companion file `sim_ivb_completa.R`.

A.5 Appendix E: Connection to Potential Outcomes

The IVB formula can be connected to the potential outcomes framework as follows. Under the standard assumptions of SUTVA and a linear constant-effect model, the potential outcomes are $Y_i(d) = \alpha + \beta_1 d + e_i$, where β_1 is the Average Treatment Effect (ATE). The observed outcome is $Y_i = Y_i(D_i) = \alpha + \beta_1 D_i + e_i$.

When the researcher conditions on a collider Z_i—a variable that is caused by both D_i and Y_i—the conditional independence assumption $Y_i(d) \perp\!\!\!\perp D_i \mid Z_i$ is violated. This is because conditioning on a common effect of D and Y creates a spurious association between treatment assignment and potential outcomes. In the language of potential outcomes, including a collider introduces selection bias *within strata of Z*.

The IVB formula quantifies this selection bias. The term $\beta_2^\star$ reflects the (spurious) partial association between Z and Y conditional on D in the misspecified model, while ϕ_1 reflects the association between D and Z that arises because both are causes (or functions of causes) of Z. Their product, with a negative sign, equals the bias in the ATE estimate.

This connection underscores a key insight: the IVB formula bridges the DAG-based and potential outcomes-based perspectives on causal inference. Where DAGs provide the qualitative diagnosis (conditioning on a collider violates d-separation), the formula provides the quantitative assessment (the bias equals $-\beta_2^\star \times \phi_1$), expressed entirely in terms of observable regression coefficients.