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

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

Março 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 $\text{IVB} = -\theta^\star \times \pi$, 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. We further show that the common practice of replacing a contemporaneous treatment with a longer lag does not mechanically eliminate IVB: bias persists to the extent that the treatment is autocorrelated across periods. We characterize structural conditions—fixed effects absorbing between-unit variation in the collider channels and few treatment switchers inflating standard errors—that explain why IVB tends to be modest in typical panel applications. We illustrate the formula with six published studies spanning democratization, peacekeeping, fiscal capacity, infrastructure, and sovereign debt. For each study, we classify controls as collider candidates using DAG-based reasoning grounded in the published literature. Among theoretically motivated collider candidates, median IVBs are approximately 0.13 standard errors of the treatment effect. Only one candidate—GDP per capita in a growth regression—exceeds one standard error ($|\text{IVB/SE}| = 2.11$, corresponding to 58% of the treatment effect).

[*]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.

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 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). 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, 2009; 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

---

[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.

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).

Beyond the formula itself, we investigate the structural conditions under which IVB is large or small in panel settings with fixed effects. We show that two-way fixed effects (TWFE) absorb the between-unit component of the treatment-collider and outcome-collider channels, leaving only the within-unit variation to drive the IVB. This is a standard property of within-estimation, but its implication for the IVB is worth stating explicitly: when the association between the treatment and a candidate collider is primarily cross-sectional (e.g., richer countries have better governance in levels), fixed effects remove this association, and the residual within-unit IVB is small. Additionally, binary treatments with few switchers inflate standard errors relative to a constant population IVB, further reducing the bias in standard-error units.

We apply the IVB formula to six published studies in comparative politics and international political economy. For each study, we classify every control variable as a collider, confounder, mediator, or predetermined variable using DAG-based reasoning grounded in the published literature. Among the 14 theoretically motivated collider candidates across six studies, only one produces an IVB exceeding one standard error of the treatment effect: GDP per capita in Rogowski et al. (2022), with $|IVB/SE| = 2.11$, corresponding to 58% of $\hat{\beta}$. The median $|IVB/SE|$ across collider candidates is approximately 0.13—well within sampling noise. We examine two cases in detail: Leipziger (2024), where GDP per capita as a collider attenuates the democratization effect by approximately 16% of $\hat{\beta}$ but only half a standard error, and Rogowski et al. (2022), where the same variable generates the only IVB exceeding one standard error but its collider status is causally ambiguous.

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, including a result showing that replacing the treatment with a longer lag does not mechanically eliminate collider

bias. Section~5 uses Monte Carlo simulations to identify the structural conditions that determine IVB magnitude in panel data. Section~6 applies the formula to six published studies. 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 using linear structural equation models (LSEMs) and path analysis. Their sensitivity analysis demonstrated that pure M-Bias tends to be a `higher-order''` `effect`---`a product of four structural correlation coefficients`---`that is typically small` `unless all four correlations are simultaneously large. They further analyzed the`Butterfly-Structure,'' in which a variable is simultaneously a collider and a confounder, showing that conditioning on it reduces bias in approximately 75% of the feasible parameter space for that specific configuration. 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.

To the best of our knowledge, no prior work has derived the IVB formula for autoregressive distributed lag models or applied it as a diagnostic tool in TSCS research. The fact that none of the six published studies we replicate in Section~6—all appearing in leading journals—computed or reported the IVB for any of their control variables suggests that the tool, despite its algebraic simplicity, is not obvious to applied researchers.

Our paper builds on and extends this prior literature in several ways. First, whereas prior formulas are parameterized in terms of structural coefficients of the underlying LSEM—quantities that are typically unobserved and require full specification of the causal model—our FWL-based derivation expresses the bias in terms of regression coefficients that the applied researcher already has in hand ($\hat{\theta}^\star$ from the long regression and $\hat{\pi}$ from a simple auxiliary regression). This reparameterization is analogous to the contribution of Cinelli and Hazlett (2020) for omitted variable bias: just as their sensitivity analysis recast the OVB in terms of partial R^2 values that researchers can benchmark, our formula recasts collider bias in terms of coefficients that researchers already report. The practical asymmetry is notable: whereas the OVB formula requires observing

the omitted confounder—which, by definition, is typically unavailable—the IVB formula uses quantities that are already estimated, since the collider is by assumption included in the regression. 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 diagnostic recipe that requires only standard regression output, lowering the barrier to routine use by applied political scientists. Because the IVB formula is an algebraic identity of OLS, it holds exactly regardless of whether D is continuous, binary, or discrete—unlike LSEM-based formulas, which require separate derivations for different variable types (Ding and Miratrix, 2015).

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).

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.
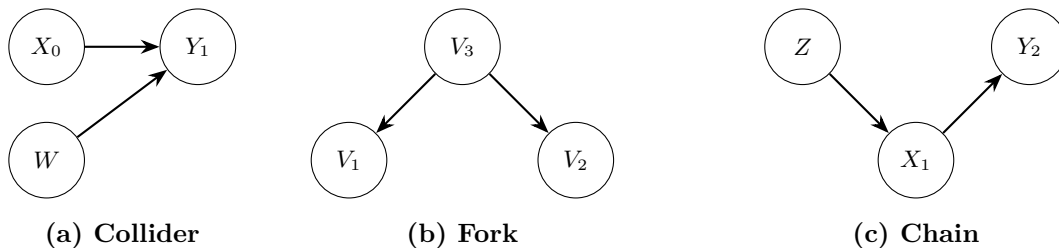

(a) Collider   **(b) Fork**   **(c) Chain**

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.
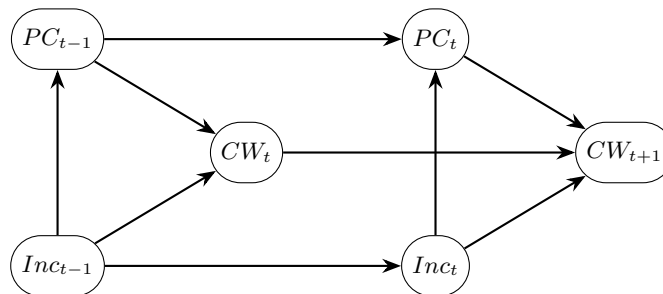

Figure 2: Temporal DAG for the effect of Political Change (PC) on Civil War (CW), with Per Capita Income (Inc) as a confounder. The temporal structure reveals that a lagged dependent variable (CW_t) is needed to block all backdoor paths to CW_{t+1}.

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.
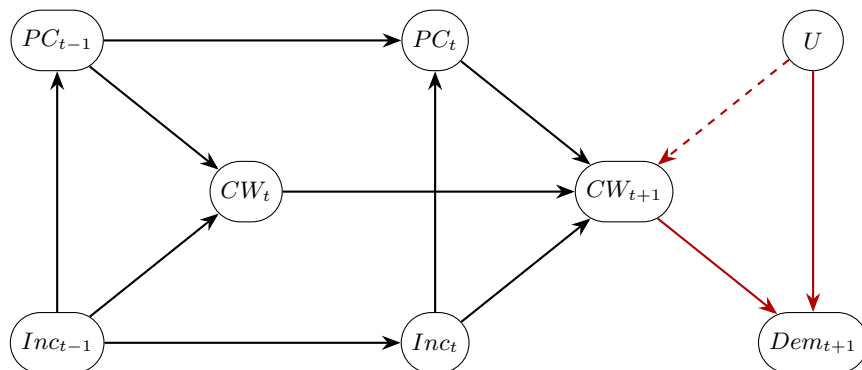

Figure 3: Complete DAG showing Democracy Level (Dem_{t+1}) as a collider. Civil War and an unobserved variable U both cause Democracy Level. Conditioning on Dem_{t+1} opens the collider $CW_{t+1} \to Dem_{t+1} \leftarrow U$, creating a spurious association between CW_{t+1} and U. Because U also causes 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. We use`foreign'' to denote that the relevant causal knowledge—that D causes Z and Y causes Z—often resides in literatures *outside* the researcher's substantive domain, not in the sense of cross-national. The phenomenon is not limited to economic variables: in studies of voter turnout (Y) and campaign spending (D), a researcher controlling for media coverage (Z) faces foreign collider bias if both spending and turnout causally affect media coverage—a relationship documented in the communications literature rather than the electoral behavior literature. 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.

The concept of foreign collider bias has broad applicability in TSCS research. Consider a researcher studying the effect of democratization on ethnic inequality who controls for GDP per capita (Leipziger, 2024). The `domestic'' literature on ethnic inequality identifies GDP as a correlate of inequality and a plausible confounder. But the`foreign'' literature on economic growth establishes that democracy causes GDP growth (Acemoglu et al., 2019), and that ethnic inequality causally reduces GDP (Grundler and Link, 2024). These two causal channels—$D \to Z$ and $Y \to Z$—make GDP per capita a collider as well as a confounder. Discovering this requires consulting the growth literature, not just the inequality literature. Similarly, in studies of public infrastructure and economic growth (Rogowski et al., 2022), GDP per capita is simultaneously the confounding variable that growth regressions must control for (conditional convergence) and a collider through the paper's core mechanism ($D \to Z$). Foreign collider bias is particularly insidious because researchers following best practices—reviewing the literature on potential confounders—may still miss it if they do not also investigate the determinants of each candidate control 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 y on D, the bias in the treatment effect estimate is:*

$$\boxed{IVB = \beta_1^\star - \beta_1 = -\beta_2^\star \times \phi_1} \tag{7}$$

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. Its simplicity, however, should not be mistaken for triviality. Prior work on collider bias quantification—including path-tracing formulas (Pearl, 2013), structural equation model derivations (Greenland, 2003), and the M-bias sensitivity analysis of Ding and Miratrix (2015)—parameterizes the bias in terms of structural coefficients of the underlying data generating process, which are not directly observable. None of these contributions made the bridge to FWL or expressed the result in terms of regression coefficients that the applied researcher already has in hand. It is precisely the simplicity of the IVB formula that makes it (i)~immediately operational as a diagnostic tool, (ii)~extensible to the ADL models prevalent in TSCS research (Propositions~2–4), and (iii)~a natural mirror of the OVB formula that applied researchers already use routinely.[2]

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	$\mathrm{sign}(\delta_1) \times \mathrm{sign}(\gamma)$	$-\mathrm{sign}(\beta_2^\star) \times \mathrm{sign}(\phi_1)$
Estimable?	Only if confounder is observed	Yes, both components from data

---

[2]The FWL-based derivation also extends straightforwardly beyond OLS. In two-stage least squares (2SLS) with a single instrument, the second stage is an OLS regression of y on $\hat{D}$ and controls, so the IVB formula applies directly to the second-stage coefficients: $\mathrm{IVB}_{\mathrm{IV}} = -\hat{\theta}_{\mathrm{IV}}^\star \times \hat{\pi}_{\mathrm{IV}}$, where $\hat{\pi}_{\mathrm{IV}}$ is the coefficient from regressing Z on the instrumented treatment $\hat{D}$. For nonlinear models such as logistic regression, the linear probability model (LPM) provides a first-order approximation of the average marginal effect, and the IVB formula applied to the LPM coefficients approximates the collider bias in the AME. We leave formal treatment of these extensions to future work.

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:

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

13

The short model coefficient (also via FWL) is:

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

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.

4.5 Interaction with Nickell Bias

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 within-transformation creates a mechanical correlation between the demeaned lagged dependent variable $\tilde{y}_{i,t-1} = y_{i,t-1} - \bar{y}_i$ and the demeaned error term $\tilde{e}_{it} = e_{it} - \bar{e}_i$. The bias is of order $O(1/T)$: it can be severe when the time dimension is short (e.g., $T < 10$) but becomes negligible as T grows (Judson and Owen, 1999).

Because the IVB formula is an algebraic identity—it decomposes the finite-sample difference between two nested OLS regressions—it remains exactly valid even in the presence of Nickell bias. The identity $\hat{\beta}^\star - \hat{\beta} = -\hat{\theta}^\star \times \hat{\pi}$ holds by construction whenever both regressions are estimated by OLS on the same sample. However, the *causal interpretation* of the components changes. In a panel with short T and unit fixed effects, both $\hat{\theta}^\star$ and $\hat{\pi}$ reflect a mixture of the true collider relationship and the dynamic bias introduced by the within-transformation. The IVB diagnostic therefore captures collider bias *within the OLS/FE framework as estimated*, not within an idealized framework free of dynamic panel bias.

To build intuition, consider how the contamination operates. In an ADL model with unit fixed effects, the coefficient on the lagged dependent variable $y_{i,t-1}$ is biased downward (toward zero) when T is small (Nickell, 1981). This downward bias on the autoregressive parameter propagates to the other coefficients in

the model, because the misestimated persistence of y distorts the partial correlations among all regressors. In particular, the coefficient $\hat{\theta}^\star$ on the candidate collider Z in the long model absorbs some of the "missing'' persistence of y, and the auxiliary-regression coefficient $\hat{\pi}$ is similarly distorted whenever Z co-moves with the lagged outcome. The net effect on the IVB product $-\hat{\theta}^\star\hat{\pi}$ is ambiguous in sign and depends on the specific DGP, but it is bounded by $O(1/T)$ just as the underlying Nickell bias is. Simulation evidence in Judson and Owen (1999) suggests that the distortion becomes practically negligible once T exceeds 20–30 periods.

The practical implication is straightforward. In panels with long time series ($T \geq 20$–30), the Nickell bias is small relative to sampling variability, and the IVB components can be interpreted as reflecting primarily the collider structure. In panels with short T (a common situation in comparative politics, where T may be 5–15 periods), researchers should be aware that the IVB estimate conflates collider bias with dynamic panel bias. In such settings, bias-corrected estimators (Arellano and Bond, 1991) or the half-panel jackknife may be needed before the IVB components can be given a purely collider-bias interpretation. Importantly, the *direction* of the IVB is less affected by Nickell bias than its *magnitude*: the sign of the product $-\hat{\theta}^\star\hat{\pi}$ tends to be informative about the direction of collider bias even when the magnitudes are contaminated by dynamic panel bias.

4.6 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:

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

16

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.

A sharper result obtains when the treatment is autoregressive and the collider depends only on the contemporaneous treatment:

Corollary 1 (IVB Decay Under AR(1) Treatment)**.** *Under the conditions of Proposition 4, suppose further that (i) the demeaned treatment follows a stationary AR(1) process,* $\widetilde{D}_t = \rho_D \widetilde{D}_{t-1} + u_t$ *with* $|\rho_D| < 1$*, and (ii) the collider depends only on the contemporaneous treatment,* $\pi_\ell = 0$ *for all* $\ell \geq 1$*. Then:*

$$IVB(k) \approx IVB(0) \cdot \rho_D^k.$$

Proof. With $\pi_\ell = 0$ for $\ell \geq 1$, the general formula reduces to $\text{IVB}(k) = -\theta^\star \pi_0 \omega_{0k}$. Under AR(1) treatment, $\omega_{0k} = \text{Cov}(\widetilde{D}_t, \widetilde{D}_{t-k})/\text{Var}(\widetilde{D}_{t-k}) = \rho_D^k$. The coefficient $\theta^\star$ is stable across k because D_{t-k} is a weak predictor of Z_t when k is large, so partitioning on D_{t-k} barely affects the projection of Y on Z. □

The corollary has immediate practical implications. For highly persistent treatments—such as democracy indices or GDP per capita, where $\rho_D \approx 0.95$—the bias decays slowly: $\text{IVB}(5) \approx 0.77 \times \text{IVB}(0)$. By contrast, for treatments with moderate autocorrelation ($\rho_D \approx 0.5$), the decay is rapid: $\text{IVB}(3) \approx 0.13 \times \text{IVB}(0)$. A useful rule of thumb is that $k^\star \approx \log(\text{SE}/|\text{IVB}(0)|)/\log(\rho_D)$ lags are needed for IVB to fall below one standard error of $\hat{\beta}$. An important limitation is that condition (ii) is restrictive: if Z depends on lags of D (i.e., $\pi_\ell \neq 0$ for some $\ell \geq 1$), the decay is governed by the weighted sum in the general formula and can be substantially slower. In the six empirical applications below, all treatments are highly persistent, which explains why lagging alone does not resolve the collider problem.

4.7 A Practical Recipe

We summarize the IVB formula as a simple diagnostic procedure for the applied researcher:

Step 0. Establish causal status. Using a DAG or substantive domain expertise, establish that Z is plausibly a collider—a variable caused by both D and Y (or by D and an unobserved common cause of Y). If Z is instead a confounder, the IVB formula quantifies omitted variable bias in the short model, and the *long* model is correctly specified. The steps below are informative as a collider diagnostic only if Z's collider status has been established through causal reasoning, not statistical testing. In Section 6, we illustrate how to combine literature review with DAG reasoning to assess collider plausibility before computing the IVB.

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:

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

The uncertainty of $\widehat{\text{IVB}}$ can be quantified via the delta method or cluster bootstrap; see Appendix A.6 for details and an upper bound on $\text{SE}(\widehat{\text{IVB}})$ that requires only standard regression output.

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.

4.8 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 formula quantifies the *arithmetic* difference between two nested specifications; it does not, by itself, determine which specification is causally correct. 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 three cases lead to fundamentally different conclusions:

- Z **is a collider.** Then the short regression is correctly specified, the IVB formula gives the bias introduced by including Z, and the researcher should exclude Z. This is the case the formula is designed to diagnose.

- Z **is a confounder.** Then the long regression is correctly specified, and the same algebraic decomposition gives the omitted variable bias in the short regression. In this case, excluding Z *introduces* bias rather than removing it. The formula is still informative—it tells the researcher how sensitive the treatment estimate is to including Z—but its sign is reversed: a large IVB is evidence that Z is an important confounder, not that it is a harmful collider.

- **Z is a mediator.** Then Z lies on the causal path from D to Y, and neither the short nor the long regression may recover the desired causal estimand without further assumptions (Acharya, Blackwell and Sen, 2016). The IVB formula captures the difference between the total effect and the direct effect, but interpreting this difference requires clarity about whether the target estimand is the total or direct effect.

Only the DAG determines which case applies. The IVB formula complements DAGs but cannot substitute for them: the graph tells the researcher *whether* Z is a collider, and only then does the formula quantify how much damage including it causes. We illustrate this interplay in Section~6, where we combine literature review with DAG reasoning to assess collider plausibility before computing the IVB. The Rogowski et al. application in particular shows that a large IVB does not automatically imply collider bias—it may instead reflect appropriate deconfounding.

Sufficient conditions for classifying Z as a collider include: (i)~credible evidence from separate studies that D causally affects Z and Y causally affects Z; (ii)~absence of evidence that Z causally affects either D or Y; and (iii)~the causal effect of D on Z and Y on Z operating through mechanisms that are not mediated by the other. When these conditions are not met—for instance, when Z is simultaneously a confounder and a collider—the IVB formula still quantifies the arithmetic difference between specifications, but its causal interpretation requires additional assumptions (see Ding and Miratrix 2015 for the "Butterfly-Structure'' case).

The empirically prevalent case in which Z is simultaneously a collider and a confounder deserves particular attention. Ding and Miratrix (2015) analyzed one such configuration—the "Butterfly-Structure''—in cross-sectional LSEMs, finding that conditioning on Z reduces net bias in approximately 75% of the feasible parameter space. However, this result is specific to a single structural configuration with particular symmetry assumptions. In TSCS settings with fixed effects, the landscape is richer: FE absorbs between-unit confounding while leaving within-unit collider channels intact, potentially shifting the trade-off in ways that depend on the panel structure. A systematic FWL-based derivation of when collider bias dominates confounding bias under TWFE—yielding analytical results rather than relying solely on simulation—is a promising direction for future work that would extend both our framework and that of Ding and Miratrix (2015) to the panel setting.

5 When Does IVB Matter in Panel Data?

The IVB formula tells a researcher *how much* bias results from including a specific collider. But a natural prior question is: under what structural conditions should we expect the IVB to be large or small? This question is especially relevant for TSCS research, where two-way fixed effects (TWFE) absorb unit-level and time-level variation, potentially shrinking the components of the IVB.

We discuss three structural features of typical TSCS designs that help explain why empirical IVBs tend to be modest. The first two follow directly from the algebra of FWL projection under fixed effects; the third is documented with Monte Carlo simulations.

5.1 Fixed Effects Absorb the Between-Unit Collider Channels

The IVB is $-\hat{\theta}^\star \hat{\pi}$, where $\hat{\pi}$ comes from the auxiliary regression $Z \sim D + \text{FE}$ and $\hat{\theta}^\star$ from the long model. Under TWFE, unit fixed effects absorb any time-invariant component of the association between D and Z or between Y and Z. This is a standard property of within-estimation, but its implication for IVB is worth stating explicitly.

Suppose the collider depends on the treatment through both a between-unit and a within-unit channel: $Z_{it} = \gamma_D^{\text{btw}} \mu_i^D + \gamma_D^{\text{wth}}(D_{it} - \mu_i^D) + \gamma_Y Y_{it} + \eta_i + \mu_t + \nu_{it}$. Because $\gamma_D^{\text{btw}} \mu_i^D$ is collinear with the unit fixed effect η_i, it is absorbed, and the surviving auxiliary coefficient is $\pi = \gamma_D^{\text{wth}} + \gamma_Y \beta$—independent of γ_D^{btw}. An analogous argument applies to the $Y \to Z$ channel: $\theta^\star$ depends only on the within-unit association between Y and Z, not on persistent cross-country differences.

The practical implication is that even if D and Z are strongly correlated in the raw data, the IVB can be small if that correlation is primarily cross-sectional rather than temporal. In many TSCS applications in comparative politics, this is precisely the case: treatment variables like democratization or peacekeeping deployment vary primarily across countries in levels, and their within-country variation is modest.

5.2 Few Switchers Inflate Standard Errors Relative to the IVB

In many TSCS studies, the treatment is binary and slow-moving: democratization, peacekeeping deployment, policy adoption. Most units are either always-treated or never-treated, and only a small fraction of "switchers'' change treatment status during the observation period. With fewer switchers, the standard error of $\hat{\beta}$ increases sharply, while the population IVB remains roughly constant. The ratio $|\text{IVB}/\text{SE}(\hat{\beta})|$ therefore *decreases* as the number of switchers falls. In Monte Carlo simulations ($N = 200$, $T = 30$, $\beta = 1$, 500 replications), reducing the fraction of switchers from 70% to 10% increases the standard error by a factor

of 2–3, reducing |IVB/SE| proportionally.

An important caveat: the fact that |IVB/SE| is small does not mean the bias is harmless. The population IVB is constant regardless of the number of switchers. In a study with few switchers and a marginally significant result, the IVB could still be large enough relative to $\hat{\beta}$ to flip the sign or significance of the treatment effect, even though |IVB/SE| < 1. The researcher should therefore consider both |IVB/$\hat{\beta}$| (the bias as a share of the estimated effect) and |IVB/SE| (the bias relative to sampling uncertainty).

5.3   Summary: When Is IVB Small?

In applications typical of comparative politics and international relations—TWFE panels with binary treatments and slow-moving variables—two mechanisms operate together. Fixed effects absorb the between-unit collider channels, leaving only within-unit co-movement to drive the IVB. Few switchers inflate standard errors relative to a constant population IVB, making the bias small in standard-error units. These features characterize the majority of TSCS designs in comparative politics, which helps explain why empirical IVBs are often modest. Researchers using cross-sectional data, short panels, or continuous treatments should expect potentially larger IVBs.

These observations are consistent with the analytical results of Ding and Miratrix (2015), who showed in a cross-sectional LSEM that pure M-Bias is a "higher-order'' effect—a product of four structural correlations—that tends to be small unless all four are simultaneously large. The empirical applications in the next section confirm this pattern.

6   Empirical Applications

To assess the empirical relevance of included variable bias, we apply the IVB formula to six published studies spanning several substantive domains: democratization and public opinion (Claassen, 2020), democracy and ethnic inequality (Leipziger, 2024), UN peacekeeping and democratization (Blair, Di Salvatore and Smidt, 2022), fiscal capacity in Africa (Albers, Jerven and Suesse, 2023), postal infrastructure and growth (Rogowski et al., 2022), and sovereign bond denomination (Ballard-Rosa, Mosley and Wellhausen, 2021). For each study, we replicate the main TWFE or panel fixed-effects specification and compute the IVB decomposition for every control variable.

Consistent with the diagnostic recipe in Section~4.7, we begin with Step~0: for each study, we classify every control variable as a collider, confounder, mediator, or predetermined/neutral variable, drawing on the published literature to assess whether credible evidence exists for both the $D \to Z$ and $Y \to Z$ channels.

A control is classified as a collider candidate only when separate studies provide credible evidence that the treatment causes the control *and* the outcome causes the control. The detailed classification tables, with supporting references, are reported in Appendix~A.7. The full mechanical IVB decomposition for all 57 study-control combinations—regardless of collider plausibility—is also available in Appendix~A.7.

Table~2 reports the IVB only for controls classified as collider candidates. We use two complementary metrics. The primary metric is $|\text{IVB}/\text{SE}(\hat{\beta})|$, which measures the bias in standard-error units. This metric is always well-defined (the standard error is bounded away from zero for any valid estimate) and has a natural benchmark: $|\text{IVB}/\text{SE}| > 1$ means the bias exceeds the sampling uncertainty of the treatment effect. As a secondary metric, we report $\text{IVB}/\hat{\beta}$ (in percent) when the treatment effect is statistically significant ($|t| \geq 1.96$), which gives the bias as a share of the estimated effect. When the treatment effect is not significant, $\text{IVB}/\hat{\beta}$ is reported as "—" because dividing by a quantity indistinguishable from zero produces uninformative ratios.

Table 2: IVB for theoretically motivated collider candidates across six studies. Controls are classified as collider candidates based on DAG reasoning (see Appendix A.7). IVB/SE measures the bias in standard-error units; values exceeding 1 indicate that the bias surpasses sampling uncertainty. $\text{IVB}/\hat{\beta}$ is reported only when the treatment effect is statistically significant ($|t| \geq 1.96$); "—" indicates that $\hat{\beta} \approx 0$ renders the ratio uninformative. For Albers et al., standard errors are unavailable from the published replication, so IVB/SE is not reported. The full mechanical decomposition for all 57 study-control combinations is in Appendix A.7.

Study	Collider candidate	IVB	|IVB/SE|	IVB/$\hat{\beta}$ (%)
Claassen (2020, FE)	Log GDP p.c.	-0.0165	-0.08	—
Leipziger (2024, SEI)	Log GDP p.c.	0.0056	0.48	-15.9%
Leipziger (2024, ext.)	Civil war	0.0002	0.02	-0.7%
Leipziger (2024, ext.)	GDP growth	-0.0000	-0.00	0.0%
Blair et al. (2022)	Foreign Aid	0.0029	0.07	2.4%
Blair et al. (2022)	GDP per capita	0.0055	0.14	4.7%
Blair et al. (2022)	Refugees/IDPs	0.0050	0.13	4.2%
Albers et al. (2023)	Hyperinflation	0.0288	—	—
Albers et al. (2023)	GDP growth (YoY)	0.0163	—	—
Albers et al. (2023)	Liberal democracy	-0.0131	—	—
Rogowski et al. (2022)	Log GDP p.c.	0.0115	2.11	58.0%
Ballard-Rosa et al. (2021)	IMF program	-0.0086	-0.19	-7.9%
Ballard-Rosa et al. (2021)	Inflation crisis	-0.0125	-0.28	-11.4%
Ballard-Rosa et al. (2021)	Sov. debt crisis	-0.0038	-0.09	-3.5%

Three patterns emerge from the table. First, among the 14 collider candidates identified across six studies, only one produces an IVB exceeding one standard error of the treatment effect: log GDP per capita in Rogowski et al. (2022), with $|\text{IVB}/\text{SE}| = 2.11$, corresponding to 58% of $\hat{\beta}$. This is the only case in which the bias is both statistically and substantively relevant. Second, the median $|\text{IVB}/\text{SE}|$ across collider candidates

(excluding Albers, Jerven and Suesse 2023, where standard errors are unavailable) is approximately 0.13—the typical collider bias is about one-eighth of a standard error, well within sampling noise. Third, GDP per capita and related economic variables dominate the collider candidates across studies, consistent with their dual role as common effects of political treatments and socioeconomic outcomes.

The IVB/SE metric avoids the pitfalls of percentage-of-$\hat{\beta}$ ratios when the treatment effect is near zero. For instance, in Claassen's FE specification, the treatment effect is not statistically significant ($\hat{\beta} = -0.016$, SE $= 0.216$), making IVB/$\hat{\beta}$ uninformative. The IVB/SE metric reveals that the bias is just 8% of a standard error ($|\text{IVB/SE}| = 0.08$)—completely negligible. When the treatment effect *is* significant, IVB/$\hat{\beta}$ provides useful complementary information about substantive magnitude: in Rogowski et al. (2022), the IVB of 58% of $\hat{\beta}$ indicates that more than half the estimated effect could reflect collider bias (though it could equally reflect appropriate deconfounding—see Section~6.2).

A caveat on cross-study comparisons: standard errors vary substantially across studies (from 0.005 in Rogowski et al. to 0.216 in Claassen's FE specification), reflecting differences in sample size, clustering structure, and treatment variation rather than differences in bias magnitude. The |IVB/SE| benchmark is most informative *within* a given study—as a measure of whether collider bias exceeds that study's own sampling uncertainty—and should be interpreted with caution when comparing across studies with very different precision levels.

These aggregate results establish the empirical baseline. We now examine two studies in detail to illustrate how the IVB formula interacts with substantive reasoning about collider plausibility: Leipziger (2024), where the IVB is modest and the collider interpretation is well supported, and Rogowski et al. (2022), where the IVB is large but causally ambiguous. Full replication details for the remaining four studies are available in the Online Appendix.

6.1   Application 1: Democracy and Ethnic Inequality

Leipziger (2024) estimates the effect of democratic transitions on socioeconomic ethnic inequality using TWFE:

$$\text{Ethnic Inequality}_{it} = \beta\,\text{Democracy}_{i,t-1} + \theta\,\text{GDP p.c.}_{i,t-1} + \delta_i + \gamma_t + \varepsilon_{it}$$

where ethnic inequality is measured by three indices (V-Dem public services, nightlight-based income from Alesina et al., and education from Omoeva et al.), Democracy is a binary indicator based on the Lexical Index (≥ 5), and GDP per capita is the sole time-varying control in the baseline specification. Leipziger deliberately excludes additional controls from the baseline, citing "the risk of posttreatment bias'' (p. 1347)—a concern

consistent with the collider bias framework developed in this paper.

The replicated treatment coefficient is $\hat{\beta} = -0.035$ (SE $= 0.012$, $N = 15,191$), indicating that democratization reduces ethnic inequality in public services.

6.1.1 Collider Assessment

Is GDP per capita a collider in this setting? The question requires consulting two literatures.

First, does democracy (the treatment) cause GDP per capita? The evidence is strong. Acemoglu et al. (2019) use dynamic panel methods, system GMM, and regional democratization waves as instruments to estimate that democratization increases GDP per capita by approximately 20% over 25 years, operating through investment, education, and reduced social conflict. Papaioannou and Siourounis (2008) find approximately 1 percentage point of additional annual growth following democratization.

Second, does ethnic inequality (the outcome) cause GDP per capita? The evidence is also strong. Easterly and Levine (1997) showed that ethnic diversity explains poor growth performance in Africa through political instability and underinvestment in public goods. Grundler and Link (2024) use instrumental variables based on the artificiality of subnational borders in Africa to estimate a *causal* effect: moving from an equal distribution to complete concentration reduces GDP per capita by 12–15%.

Both channels—$D \to Z$ and $Y \to Z$—are supported by credible causal evidence, making GDP per capita a plausible collider. This is an instance of foreign collider bias: the researcher studying ethnic inequality must consult the growth economics literature to discover that GDP per capita is not merely a confounder but also a collider.

6.1.2 IVB Decomposition

Table 3: IVB decomposition for Leipziger (2024): Democracy $\to$ Ethnic Inequality (SEI), with GDP per capita as candidate collider.

	Estimate
$\hat{\beta}_{\text{short}}$ (without GDP p.c.)	-0.0407
$\hat{\beta}_{\text{long}}$ (with GDP p.c.)	-0.0352
$\hat{\theta}^*$ (GDP p.c. in long model)	-0.0497
$\hat{\pi}$ (Democracy in auxiliary reg.)	0.1123
IVB $= -\hat{\theta}^* \times \hat{\pi}$	0.0056
|IVB/SE|	0.48

	Estimate
$|\text{IVB}|/|\hat{\beta}_{\text{long}}|$	15.9%

GDP per capita produces an IVB of approximately 0.0056, corresponding to $|\text{IVB/SE}| = 0.48$—about half a standard error—and approximately 16% of the treatment effect. The direction is consistent with the literature: $\hat{\theta}^{\star} < 0$ (richer countries have less ethnic inequality) and $\hat{\pi} > 0$ (democracy increases income), so $\text{IVB} = -\hat{\theta}^{\star}\hat{\pi} > 0$, pushing the treatment coefficient toward zero. In other words, controlling for GDP per capita *attenuates* the estimated effect of democratization on ethnic inequality. The short model (without GDP per capita) yields a somewhat larger effect.

In the extended specification that includes additional controls, civil war is a second collider candidate: democracy reduces civil war risk (Cederman, Gleditsch and Buhaug, 2013) and ethnic inequality increases it (Cederman, Gleditsch and Buhaug, 2013), satisfying both the $D \to Z$ and $Y \to Z$ channels. However, its IVB is negligible ($|\text{IVB/SE}| = 0.02$).

This application illustrates two key points. First, even a single control retained for valid confounding-adjustment reasons can introduce a measurable IVB when it is also a collider. Leipziger's careful exclusion of other controls to avoid post-treatment bias was well motivated—but the one control that was retained (GDP per capita) is itself a candidate collider. Second, the IVB of approximately 16% is substantively notable but does not threaten the main finding: at half a standard error, it falls well within sampling uncertainty. This is consistent with the discussion in Section~5: TWFE absorbs the between-unit component of the democracy–GDP association, leaving only the modest within-unit co-movement to drive the IVB.

6.2 Application 2: Postal Infrastructure and Economic Growth

Rogowski et al. (2022) estimate the effect of public postal infrastructure on economic development using a TWFE country panel model. The main specification is:

$$\text{GDP growth}_{i,t+1} = \beta \, \text{Post Offices}_{i,t-1} + \mathbf{Z}'_{it}\gamma + \delta_i + \mu_t + \varepsilon_{it}$$

where GDP growth is the 5-year growth rate (one period forward), Post Offices is the log cumulative stock (5-year lag), and $\mathbf{Z}_{it}$ includes four controls: log GDP per capita (5-year lag), log population, urbanization rate, and Polity2 democracy score.

The replicated treatment coefficient is $\hat{\beta} = 0.0198$ (SE $= 0.0054$, $N = 1,540$). Post offices have a statistically

significant positive effect on subsequent GDP growth.

6.2.1 Collider Assessment

The key collider candidate is log GDP per capita. The paper's core argument is that post offices facilitate economic development $(D \rightarrow Z)$—indeed, this is precisely the mechanism being tested. At the same time, GDP growth (the outcome) mechanically determines future GDP per capita levels $(Y \rightarrow Z$, through the accounting identity that growth cumulates into levels). This creates a collider structure: both D and Y cause Z.

However, GDP per capita is *simultaneously* a strong confounder in growth regressions, through conditional convergence: richer countries grow more slowly, holding other factors constant. This is a well-established empirical regularity in the growth literature (Acemoglu et al., 2019). Whether controlling for GDP per capita introduces collider bias (through the $D \rightarrow Z$ and $Y \rightarrow Z$ channels) or removes confounding bias (through the convergence channel) depends on the relative strength of these effects—a question the IVB formula alone cannot answer.

6.2.2 IVB Decomposition

Table 4: IVB decomposition for Rogowski et al. (2022): Post Offices $\rightarrow$ GDP Growth.

Control as z	$\hat{\beta}_{\mathrm{short}}$	$\hat{\beta}_{\mathrm{long}}$	$\hat{\theta}^*$	$\hat{\pi}$	IVB	|IVB/SE|	IVB/$\hat{\beta}$ (%)
Log GDP p.c.	0.0083	0.0198	-4.0130	0.0029	0.0115	2.11	58.0
Log population	0.0206	0.0198	-1.9465	- 0.0004	- 0.0008	0.15	-4.0
Urbanization	0.0192	0.0198	-3.1421	0.0002	0.0006	0.11	3.0
Polity2	0.0191	0.0198	0.0341	- 0.0221	0.0008	0.14	3.8

Log GDP per capita dominates the IVB decomposition, with |IVB/SE| = 2.11—the only case across all six studies where the collider bias exceeds one standard error of the treatment effect. This corresponds to approximately 58% of $\hat{\beta}_{\mathrm{long}}$. Including GDP per capita *inflates* the treatment effect from 0.0083 (without

GDP p.c.) to 0.0198 (with GDP p.c.). The mechanism is clear: $\hat{\theta}^\star < 0$ (convergence—higher GDP per capita predicts lower subsequent growth) and $\hat{\pi} > 0$ (more post offices predict higher GDP per capita), so IVB $= -\hat{\theta}^\star\hat{\pi} > 0$. The other controls produce IVBs well below one standard error.

6.2.3 The Limits of the IVB Formula

This application illustrates the crucial caveat from Section~4.8: **a large IVB does not automatically imply collider bias**. The 58% IVB for GDP per capita could reflect either of two causal structures:

- If GDP per capita is primarily a **collider** (because post offices raise GDP through the paper's core mechanism, and growth mechanically cumulates into GDP levels), then including it biases the treatment effect upward, and the short model's estimate of 0.0083 is closer to the truth.

- If GDP per capita is primarily a **confounder** (because richer countries both build more post offices and grow more slowly due to convergence), then excluding it biases the treatment effect downward, and the long model's estimate of 0.0198 is closer to the truth.

In reality, GDP per capita is likely *both* a confounder and a collider, and the net effect depends on the relative magnitudes of these channels. The IVB formula cannot resolve this ambiguity—it can only quantify the arithmetic difference between the two specifications. The researcher must use domain knowledge and causal reasoning to determine which specification is more appropriate.

This case underscores that the IVB formula is a *diagnostic* tool, not a decision rule. It tells the researcher exactly how much the treatment estimate changes when a candidate collider is included. Whether that change represents the introduction of bias (collider) or the removal of bias (confounder) requires a DAG.

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, 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.'' A further corollary shows that the common practice of replacing a contemporaneous treatment with a longer lag does not mechanically cure IVB: because the treatment is typically autocorrelated, the collider retains a partial association with the lagged treatment, and bias persists. Longer lags may attenuate IVB when these autocorrelations decay, but they do not guarantee its elimination.

Third, we provide a diagnostic recipe that any applied researcher can implement, and we investigate both the structural conditions under which IVB is large and the empirical magnitude of IVB in published studies. We show that fixed effects absorb the between-unit component of the collider channels, and that binary treatments with few switchers inflate standard errors relative to the IVB. In typical TWFE applications in comparative politics, these features operate together, which helps explain why empirical IVBs are often modest. Across six published studies spanning democratization, peacekeeping, fiscal capacity, infrastructure, and sovereign debt, we classify each control variable using DAG-based reasoning grounded in the published literature and compute the IVB for theoretically motivated collider candidates. Among 14 such candidates across six studies, the median |IVB/SE| is approximately 0.13—the typical collider bias is about one-eighth of a standard error, well within sampling noise. Only one candidate—GDP per capita in a growth regression (Rogowski et al., 2022)—exceeds one standard error (|IVB/SE| = 2.11, corresponding to 58% of $\hat{\beta}$), and even this case is causally ambiguous because GDP per capita is simultaneously a confounder. These conditions characterize the majority of TSCS designs in comparative politics, but researchers using cross-sectional data, short panels, or continuous treatments should expect potentially larger IVBs.

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.

The empirical applications also highlight the importance of what we call `foreign collider bias.''
`In both replications, the candidate collider (GDP per capita) is a variable whose causal`

`relationships with treatment and outcome are documented in literatures *outside* the`
`focal study's domain. A researcher studying ethnic inequality or postal infrastructure`
`may not spontaneously consult the literatures on democracy and economic growth or on`
`convergence dynamics. Yet it is precisely these`foreign'' literatures that reveal the collider structure. This suggests that the standard heuristic of reviewing the outcome and treatment literatures should be supplemented with systematic attention to the determinants of each control variable.

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. Future work should investigate whether the linear IVB provides a useful approximation in such settings or whether a separate nonlinear formula is needed. Critically, the IVB formula is an algebraic identity that quantifies the *arithmetic* difference between two nested specifications; only a DAG or substantive domain expertise can determine whether that difference represents collider bias, omitted variable bias, or mediation. The formula does not help with the prior step of identifying colliders, for which causal reasoning remains 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.

These limitations point to two natural directions for future work. First, 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?'' Second, and more ambitiously, the FWL framework that underpins our formula could be extended to derive *analytical* results for the case in which Z is simultaneously a collider and a confounder under TWFE. Ding and Miratrix (2015) analyzed this trade-off for a specific cross-sectional configuration (the Butterfly-Structure), but the TSCS setting introduces richer possibilities: fixed effects differentially absorb between-unit confounding and collider channels, potentially shifting the balance in ways that depend on the panel structure. An FWL-based derivation of when collider bias dominates confounding bias under TWFE—yielding closed-form conditions rather than relying solely on simulation—would provide applied researchers with stronger analytical guidance. We leave both extensions for future research. We also provide the `compute_ivb_multi()` function as supplementary material, enabling researchers to compute IVB decompositions for any TWFE specification with minimal effort.

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

Acemoglu, Daron, Suresh Naidu, Pascual Restrepo and James A. Robinson. 2019. "Democracy Does Cause Growth." *Journal of Political Economy* 127(1):47–100.

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

Albers, Thilo, Morten Jerven and Marvin Suesse. 2023. "The Fiscal State in Africa: Evidence from a Century of Growth." *International Organization* 77(4):823–856.

Alesina, Alberto. 1987. "Macroeconomic Policy in a Two-Party System as a Repeated Game." *The Quarterly Journal of Economics* 102(3):651–678.

Alesina, Alberto and Guido Tabellini. 1990. "A Positive Theory of Fiscal Deficits and Government Debt." *Review of Economic Studies* 57(3):403–414.

Arellano, Manuel and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *The Review of Economic Studies* 58(2):277–297.

Ballard-Rosa, Cameron, Layna Mosley and Rachel L. Wellhausen. 2021. "Coming to Terms: The Politics of Sovereign Bond Denomination." *International Organization* 75(4):1063–1092.

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.

Bermeo, Sarah Blodgett. 2016. "Aid Is Not Oil: Donor Utility, Heterogeneous Aid, and the Aid-Democratization Relationship." *International Organization* 70(1):1–32.

Besley, Timothy and Torsten Persson. 2011. *Pillars of Prosperity: The Political Economics of Development Clusters.* Princeton University Press.

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.

Blair, Robert A., Jessica Di Salvatore and Hannah M. Smidt. 2022. "UN Peacekeeping and Democratization in Conflict-Affected Countries." *American Political Science Review* 116(4):1215–1232.

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

Cederman, Lars-Erik, Kristian Skrede Gleditsch and Halvard Buhaug. 2013. *Inequality, Grievances, and Civil War.* Cambridge University Press.

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.

Claassen, Christopher. 2020. "Does Public Support Help Democracy Survive?" *American Journal of Political Science* 64(1):118–134.

Costalli, Stefano, Jessica Di Salvatore and Andrea Ruggeri. 2024. "Do UN Peace Operations Help Forcibly Displaced People?" *Journal of Peace Research* 61(1):95–110.

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* .

Easterly, William and Ross Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *The Quarterly Journal of Economics* 112(4):1203–1250.

Eichengreen, Barry, Ricardo Hausmann and Ugo Panizza. 2005. The Pain of Original Sin. In *Other People's Money: Debt Denomination and Financial Instability in Emerging Market Economies*, ed. Barry Eichengreen and Ricardo Hausmann. University of Chicago Press pp. 13–47.

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

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.

Grundler, Klaus and Sebastian Link. 2024. "The Causal Effect of Ethnic Inequality on Income Per Capita." *The Economic Journal* 134(661):1985–2016.

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.

Hibbs, Douglas A. 1977. "Political Parties and Macroeconomic Policy." *American Political Science Review* 71(4):1467–1487.

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.

Judson, Ruth A. and Ann L. Owen. 1999. "Estimating Dynamic Panel Data Models: A Guide for Macroeconomists." *Economics Letters* 65(1):9–15.

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.

Leipziger, Lukas E. 2024. "Does Democracy Reduce Ethnic Inequality?" *American Journal of Political Science* 68(4):1335–1352.

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.

Moore, Will H. and Stephen M. Shellman. 2004. "Fear of Persecution: Forced Migration, 1952–1995." *Journal of Conflict Resolution* 48(5):723–745.

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

Papaioannou, Elias and Gregorios Siourounis. 2008. "Democratisation and Growth." *The Economic Journal* 118(532):1520–1551.

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.

Rogowski, Jon C., John Gerring, Matthew Maguire and Lee Cojocaru. 2022. "Public Infrastructure and Economic Development: Evidence from Postal Systems." *American Journal of Political Science* 66(4):885–901.

Ross, Michael L. 2001. "Does Oil Hinder Democracy?" *World Politics* 53(3):325–361.

Sargent, Thomas J. and Neil Wallace. 1981. "Some Unpleasant Monetarist Arithmetic." *Federal Reserve Bank of Minneapolis Quarterly Review* 5(3):1–17.

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

Vreeland, James Raymond. 2003. *The IMF and Economic Development.* Cambridge University Press.

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.

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:

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

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

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

Step 3 (Bias). Subtracting:

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

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 $\tilde{D}_t$:

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

Step 5 (Result).

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

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 $\mathrm{IVB}(\beta_j) = -\theta^\star \times \pi_j$ 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

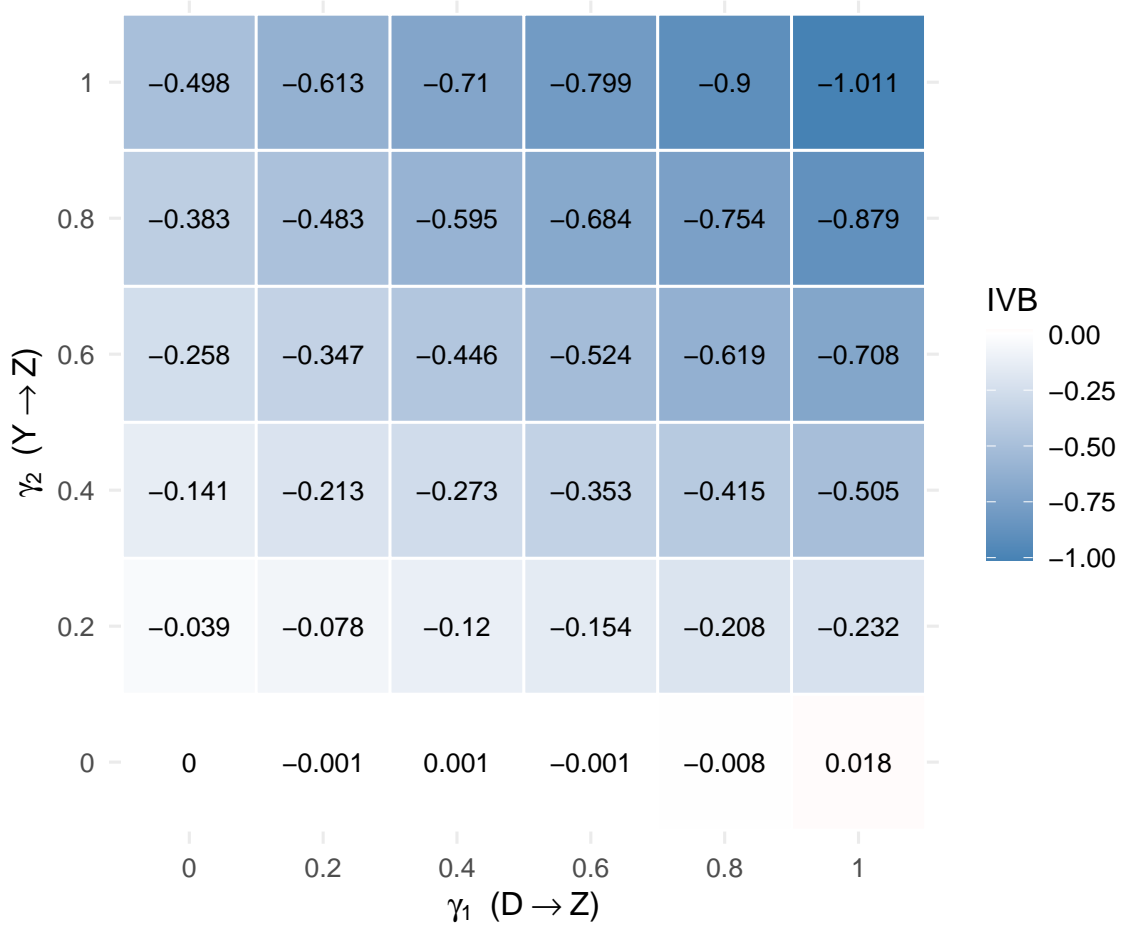Figure A1: Cross-section IVB as a function of the collider parameters γ_1 (D $\rightarrow$ Z) and γ_2 (Y $\rightarrow$ Z). The bias increases in magnitude with both parameters.
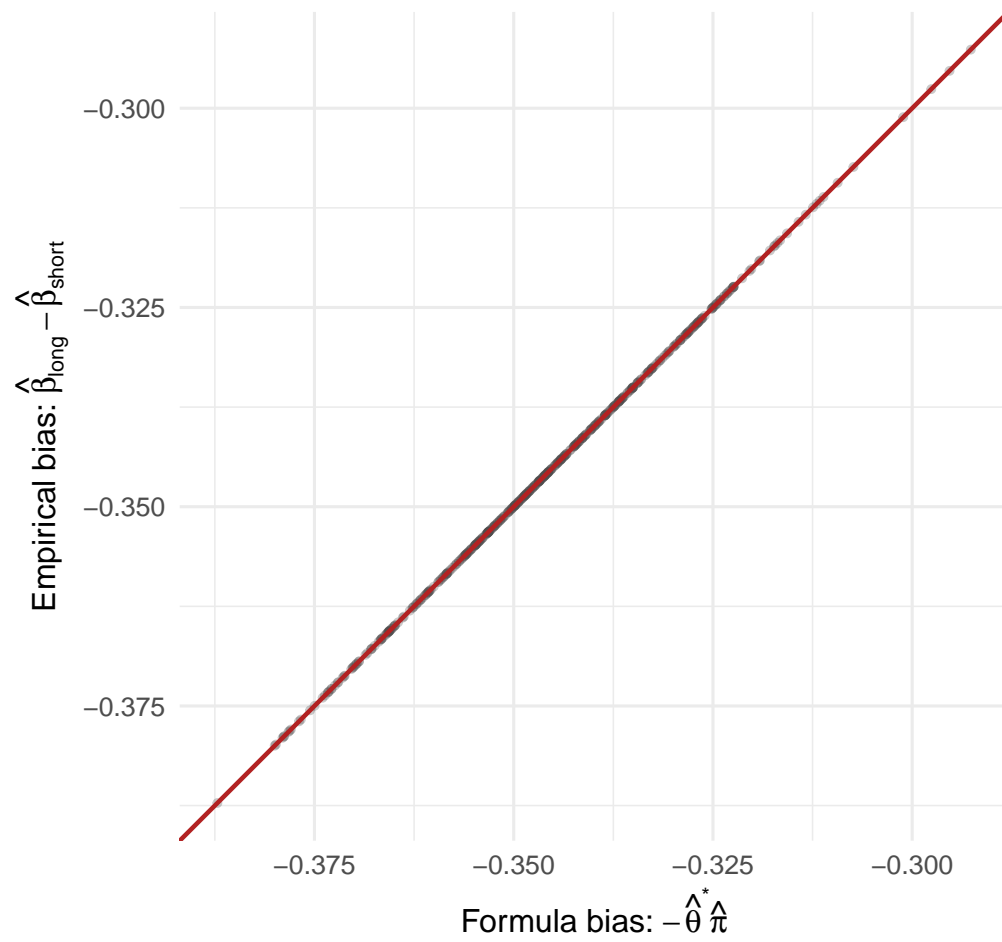
Figure A2: ADL(1,0): Empirical bias versus IVB formula prediction across 500 simulations.
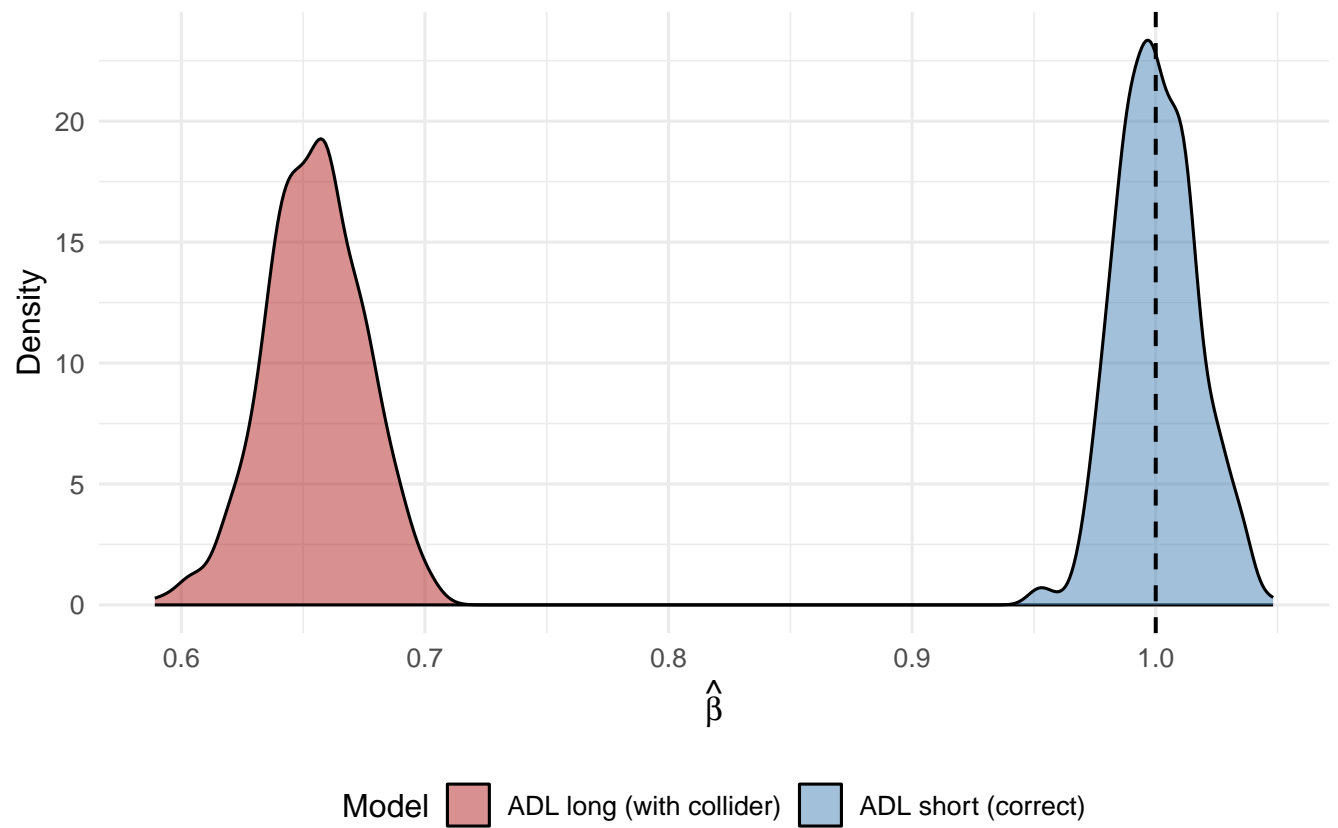
Figure A3: ADL(1,0): Distribution of treatment effect estimates with and without the collider across 500 simulations.
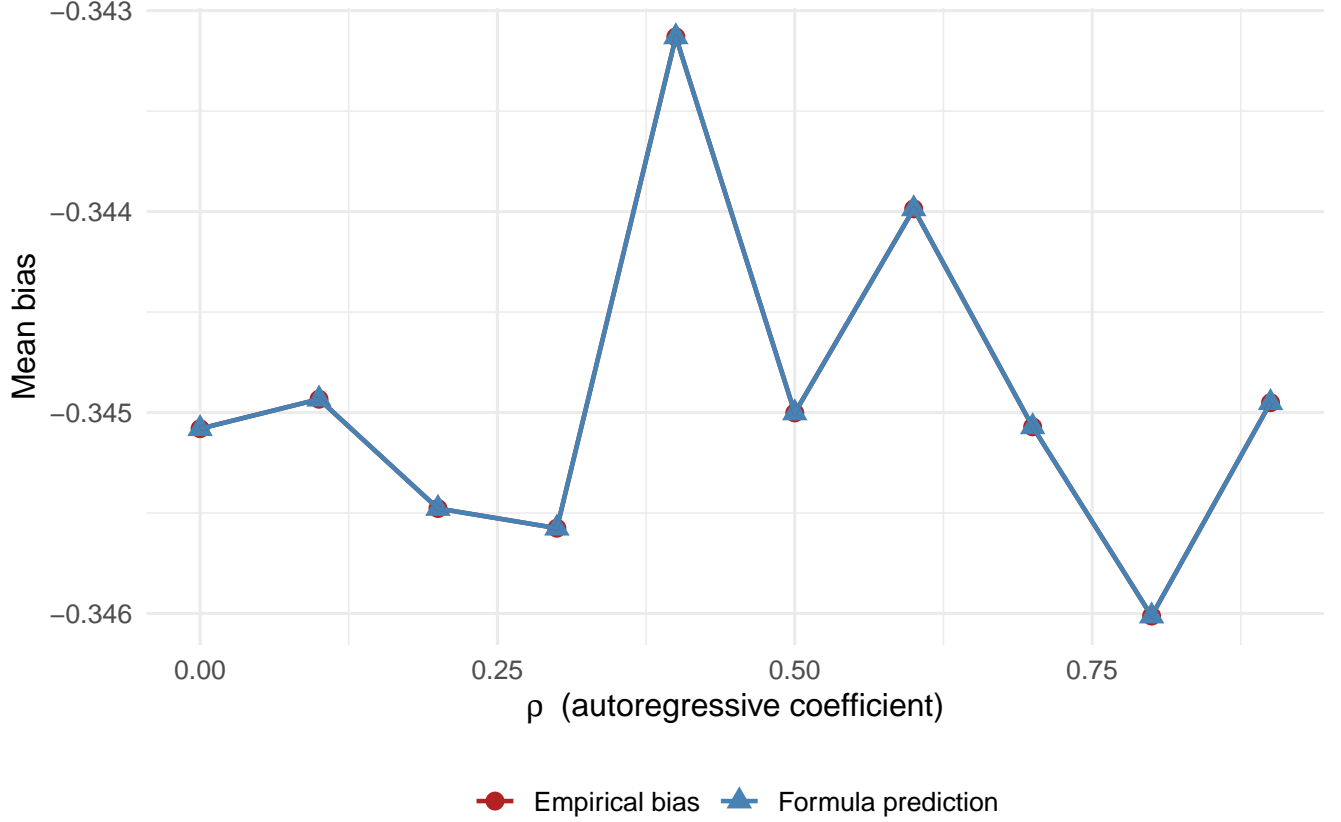
Figure A4: ADL(1,0): Effect of the autoregressive parameter ρ on the magnitude of the IVB. Both the empirical bias and the formula prediction are shown.

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.

A.6 Appendix F: Variance of the IVB Estimator

The IVB diagnostic $\widehat{IVB} = -\hat{\theta}^\star \hat{\pi}$ is a product of two estimators. Its sampling variance can be approximated via the delta method. Write $IVB = -\theta^\star \pi$ and let $\hat{\theta}^\star$ and $\hat{\pi}$ denote the OLS estimators from the long model and the auxiliary regression, respectively. By the delta method:

$$\text{Var}(\widehat{IVB}) \approx \pi^2 \, \text{Var}(\hat{\theta}^\star) + \theta^{\star 2} \, \text{Var}(\hat{\pi}) + 2\theta^\star \pi \, \text{Cov}(\hat{\theta}^\star, \hat{\pi}). \tag{14}$$

The first two terms on the right-hand side are readily available from the standard output of the long regression $(\text{Var}(\hat{\theta}^\star))$ and the auxiliary regression $(\text{Var}(\hat{\pi}))$. The cross-covariance term $\text{Cov}(\hat{\theta}^\star, \hat{\pi})$ is not standard, since the two estimators come from different regressions. It can be estimated via seemingly unrelated regressions

(SUR) or, more simply, via cluster bootstrap.

For a conservative bound that avoids estimating the cross-covariance, apply the Cauchy–Schwarz inequality:

$$\text{SE}(\widehat{\text{IVB}}) \leq |\hat{\pi}| \cdot \text{SE}(\hat{\theta}^{\star}) + |\hat{\theta}^{\star}| \cdot \text{SE}(\hat{\pi}). \tag{15}$$

This upper bound is computable from standard regression output alone, without any additional estimation. If the bound is small relative to $\text{SE}(\hat{\beta})$, the researcher can conclude that the IVB is precisely estimated as small (or large). If the bound is inconclusive—i.e., the upper bound on $\text{SE}(\widehat{\text{IVB}})$ is of similar magnitude to $|\widehat{\text{IVB}}|$ itself—then a cluster bootstrap is recommended to obtain a tighter interval.

We recommend the following two-step procedure: (1) compute the upper bound in Equation~(15) as a quick screening tool; (2) if the bound is inconclusive, run a cluster bootstrap that jointly resamples both regressions to obtain $\text{SE}(\widehat{\text{IVB}})$ and a confidence interval directly.

A.7   Appendix G: Control Classification and Full IVB Results

Following the diagnostic recipe in Section~4.7 (Step~0), we classify each control variable in the six replicated studies according to its causal role relative to the treatment (D) and outcome (Y). A control is classified as a **collider** when credible evidence from separate studies supports both $D \to Z$ and $Y \to Z$; as a **confounder** when the evidence supports $Z \to D$ and/or $Z \to Y$ (but not $D \to Z$); as a **mediator** when $D \to Z \to Y$; and as **predetermined/neutral** when no contemporaneous causal pathway is plausible (e.g., geographic or demographic variables, or lagged variables where the lag structure blocks the collider channel). Variables with evidence supporting both confounding and collider channels are labeled **collider + confounder**. The classification criteria are consistent with the sufficient conditions stated in Section~4.8.

A.7.1 Claassen (2020) — D: public support, Y: liberal democracy

Table A1: Control classification for Claassen (2020).

Control	$D \to Z$	$Y \to Z$	Classification	Key references
Log GDP p.c.	Weak (institutional)	Democracy $\to$ GDP	Collider (weak)	Acemoglu et al. (2019)
GDP growth	Same channel	Same channel	Collider (weak)	Acemoglu et al. (2019)
Regional democracy	Diffusion	Composition	Predetermined	—
% Muslim	None	None (absorbed by FE)	Predetermined	—
Resource dep.	None	Resource curse is $Z \to Y$	Confounder	Ross (2001)

Collider candidate for Table~2: Log GDP p.c. GDP growth shares the same mechanism but is excluded from the main table to avoid double-counting the same causal channel. Both GDP variables are classified as "Collider (weak)'' because the $D \to Z$ channel (public support $\to$ GDP) is indirect and operates through institutional quality rather than a direct causal mechanism.

A.7.2 Leipziger (2024) — D: binary democracy, Y: ethnic inequality (SEI)

Table A2: Control classification for Leipziger (2024).

Control	$D \to Z$	$Y \to Z$	Classification	Key references
Log GDP p.c.	Democracy $\to$ GDP	Ethnic ineq. $\to$ GDP	**Collider**	Acemoglu et al. (2019); Grundler and Link (2024)
Oil income p.c.	None	Geological	Predetermined	—
Civil war	Democracy $\to$ civil war	Ethnic ineq. $\to$ civil war	**Collider**	Cederman, Gleditsch and Buhaug (2013)
Ethnic frac.	None	Demographic	Predetermined	—
GDP growth	Same channel as GDP p.c.	Same channel	**Collider**	Acemoglu et al. (2019)

Collider candidates: Log GDP p.c. (baseline spec.), Civil war and GDP growth (extended spec.).

A.7.3 Blair et al. (2022) — D: UN peacekeeping, Y: democratization

Table A3: Control classification for Blair et al. (2022).

Control	$D \to Z$	$Y \to Z$	Classification	Key references
Population	Demographic	Demographic	Predetermined	—
Foreign Aid	PKO → bundled aid	Democracy → aid	**Collider**	Bermeo (2016)
GDP p.c.	PKO → stability → GDP	Democracy → GDP	**Collider**	Acemoglu et al. (2019)
Refugees/IDPs	PKO → fewer displaced	Democracy → less repression	**Collider + Confounder**	Costalli, Di Salvatore and Ruggeri (2024); Moore and Shellman (2004)
Literacy	Education stock (decades)	Weak	Predetermined	—
Fuel exports	Geological	Geological	Predetermined	—

Collider candidates: Foreign Aid, GDP per capita, Refugees/IDPs. Refugees/IDPs is classified as collider + confounder: PKO mandates reduce displacement ($D \to Z$; Costalli, Di Salvatore and Ruggeri, 2024) and democratization reduces refugee-generating repression ($Y \to Z$; Moore and Shellman, 2004), but refugee crises also trigger PKO deployment ($Z \to D$), making the variable simultaneously a confounder (as with GDP per capita in Rogowski et al. 2022).

A.7.4 Albers et al. (2023) — D: government turnover, Y: tax collection

Table A4: Control classification for Albers et al. (2023).

Control	$D \to Z$	$Y \to Z$	Classification	Key references
Independence	Historical	Structural	Predetermined	—
Hyperinflation	Turnover → fiscal → inflation	Tax → fiscal balance	**Collider**	Sargent and Wallace (1981)
Civil war (lag)	Lag protects	Lag protects	Predetermined	—
GDP growth	Turnover → uncertainty → GDP	Tax → investment → GDP	**Collider**	—
Drought (lag)	Exogenous	Exogenous	Predetermined	—
Liberal democracy	Turnover → regime openness	Fiscal bargaining	**Collider (weak)**	Besley and Persson (2011)
Secession	Ethnic-political	Structural	Predetermined	—
Socialist	Ideological	Structural	Predetermined	—
Alliance (G5)	Weak	Weak	Predetermined	—
Market access	Weak	Creditworthiness	Ambiguous	—
Ext. default	Weak	Fiscal capacity	Ambiguous	—
Intl. war (lag)	Geopolitical	Geopolitical	Predetermined	—

Collider candidates: Hyperinflation, GDP growth (YoY), Liberal democracy. Note: standard errors for Albers et al. are not available from the published replication materials, so IVB/SE cannot be computed.

A.7.5 Rogowski et al. (2022) — D: log post offices, Y: GDP growth

Table A5: Control classification for Rogowski et al. (2022).

Control	$D \to Z$	$Y \to Z$	Classification	Key references
Log GDP p.c.	PO → GDP (paper's mechanism)	Growth → levels (identity)	**Collider + Confounder**	Acemoglu et (2019)
Log population	Population → PO (demand)	Pop. → growth (Solow)	Confounder	—
Urbanization	Weak	Growth → urbanization	Weak	—
Polity2	PO → democracy (slow)	Growth → democratization	Weak	—

Collider candidate: Log GDP p.c. (with caveat that it is simultaneously a confounder through conditional convergence; see Section~6.2).

A.7.6 Ballard-Rosa et al. (2021) — D: left government, Y: domestic-currency bonds

Table A6: Control classification for Ballard-Rosa et al. (2021).

Control	$D \rightarrow Z$	$Y \rightarrow Z$	Classification	Key references
Inflation crisis	Left $\rightarrow$ inflation	Bonds $\rightarrow$ inflation expect.	**Collider**	Hibbs (Alesina (1987)
IMF program	Left $\rightarrow$ fiscal policy $\rightarrow$ IMF	Bonds $\rightarrow$ fiscal crisis $\rightarrow$ IMF	**Collider**	Vreeland (Eichengreen, mann and Pa (2005)
Sov. debt crisis	Left $\rightarrow$ fiscal expansion $\rightarrow$ crisis	Bonds $\rightarrow$ currency mismatch $\rightarrow$ crisis	**Collider**	Alesina and Tab (1990); Eichen Hausmann Panizza (2005)
Log GDP p.c.	Left $\rightarrow$ GDP (weak, lagged)	Weak	Predetermined	—
GDP growth	Weak	Weak	Predetermined	—
Ext. debt/GDP	Weak	Weak	Predetermined	—
Curr. acct/GDP	Weak	Weak	Predetermined	—
Trade/GDP	Structural	Structural	Predetermined	—
Oil rents/GDP	Geological	Geological	Predetermined	—
FDI/GDP	Weak	Weak	Predetermined	—
US Treasury 10yr	Exogenous	Exogenous	Predetermined	—
Pegged XR	Structural	Weak	Predetermined	—
High CBI	Institutional	Weak	Predetermined	—
Chinn-Ito open.	Structural	Weak	Predetermined	—
Currency crisis	Weak	Weak	Predetermined	—
Democracy (V-Dem)	Weak	Weak	Predetermined	—

Collider candidates: Inflation crisis, IMF program, Sovereign debt crisis. All three share a common mechanism: left-government fiscal policies create fiscal stress ($D \rightarrow Z$; Alesina and Tabellini 1990; Alesina 1987), while the currency denomination of sovereign bonds affects crisis vulnerability through balance-sheet mismatches ($Y \rightarrow Z$; Eichengreen, Hausmann and Panizza 2005). For IMF programs specifically, Vreeland (2003) shows that domestic political factors—including government ideology—drive program participation beyond economic necessity.

A.7.7 Full Mechanical IVB Decomposition

Table~A7 reports the complete IVB decomposition for all 57 study-control combinations, regardless of collider plausibility. This table is provided for completeness and transparency; the substantive interpretation of any individual IVB as collider bias requires the causal reasoning documented in the classification tables above.

Table A7: Full mechanical IVB decomposition. All controls are included regardless of collider plausibility. IVB/$\hat{\beta}$ is reported only when the treatment effect is significant ($|t| \geq 1.96$); "—" indicates $\hat{\beta} \approx 0$ or SE unavailable.

Study	Control	IVB	|IVB/SE|	IVB/$\hat{\beta}$ (%)
Claassen (OLS)	Log GDP p.c.	-0.0010	-0.01	-0.4%
Claassen (OLS)	GDP growth	-0.0009	-0.01	-0.3%
Claassen (OLS)	Regional democracy	-0.0092	-0.11	-3.4%
Claassen (OLS)	% Muslim	0.0010	0.01	0.4%
Claassen (OLS)	Resource dep.	0.0083	0.10	3.1%
Claassen (FE)	Log GDP p.c.	-0.0165	-0.08	—
Claassen (FE)	GDP growth	0.0123	0.06	—
Claassen (FE)	Regional democracy	-0.0020	-0.01	—
Claassen (FE)	Resource dep.	-0.0001	-0.00	—
Leipziger (SEI)	Log GDP p.c.	0.0056	0.48	-15.9%
Leipziger (GRG)	Log GDP p.c.	-0.0003	-0.02	0.9%
Leipziger (GGINI)	Log GDP p.c.	-0.0008	-0.12	—
Leipziger (SEI ext.)	Log GDP p.c.	0.0048	0.41	-15.4%
Leipziger (SEI ext.)	Oil income p.c.	-0.0000	-0.00	0.1%
Leipziger (SEI ext.)	Civil war	0.0002	0.02	-0.7%
Leipziger (SEI ext.)	Ethnic frac.	0.0007	0.06	-2.1%
Leipziger (SEI ext.)	GDP growth	-0.0000	-0.00	0.0%
Blair et al.	Population	-0.0005	-0.01	-0.5%
Blair et al.	Foreign Aid	0.0029	0.07	2.4%
Blair et al.	GDP per capita	0.0055	0.14	4.7%
Blair et al.	Refugees/IDPs	0.0050	0.13	4.2%
Blair et al.	Literacy	0.0000	0.00	0.0%
Blair et al.	Fuel exports	-0.0010	-0.03	-0.8%
Albers et al.	Independence	-0.0409	—	—
Albers et al.	Hyperinflation	0.0288	—	—

Table A7: Full mechanical IVB decomposition. All controls are included regardless of collider plausibility. IVB/$\hat{\beta}$ is reported only when the treatment effect is significant ($|t| \geq 1.96$); "—" indicates $\hat{\beta} \approx 0$ or SE unavailable. *(continued)*

Study	Control	IVB	|IVB/SE|	IVB/$\hat{\beta}$ (%)
Albers et al.	Civil war (PRIO)	0.0210	—	—
Albers et al.	GDP growth (YoY)	0.0163	—	—
Albers et al.	Drought	-0.0150	—	—
Albers et al.	Liberal democracy	-0.0131	—	—
Albers et al.	Secession	0.0124	—	—
Albers et al.	Socialist	0.0070	—	—
Albers et al.	Alliance (G5)	0.0061	—	—
Albers et al.	Market access	0.0055	—	—
Albers et al.	External default	-0.0014	—	—
Albers et al.	Intl war (PRIO)	-0.0007	—	—
Rogowski et al.	Log GDP p.c.	0.0115	2.11	58.0%
Rogowski et al.	Log population	-0.0008	-0.15	-4.0%
Rogowski et al.	Urbanization	0.0006	0.11	3.0%
Rogowski et al.	Polity2	0.0008	0.14	3.8%
Ballard-Rosa et al.	Log GDP p.c.	-0.0057	-0.13	-5.2%
Ballard-Rosa et al.	GDP growth	0.0003	0.01	0.3%
Ballard-Rosa et al.	Ext. debt/GDP	0.0057	0.13	5.2%
Ballard-Rosa et al.	Curr. acct/GDP	-0.0018	-0.04	-1.7%
Ballard-Rosa et al.	Trade/GDP	0.0024	0.05	2.2%
Ballard-Rosa et al.	Oil rents/GDP	0.0002	0.00	0.2%
Ballard-Rosa et al.	FDI/GDP	-0.0007	-0.02	-0.6%
Ballard-Rosa et al.	US Treasury 10yr	0.0012	0.03	1.1%
Ballard-Rosa et al.	Pegged XR	-0.0001	-0.00	-0.1%
Ballard-Rosa et al.	High CBI	-0.0005	-0.01	-0.4%
Ballard-Rosa et al.	Chinn-Ito open.	0.0010	0.02	0.9%
Ballard-Rosa et al.	IMF program	-0.0086	-0.19	-7.9%
Ballard-Rosa et al.	Currency crisis	0.0006	0.01	0.6%
Ballard-Rosa et al.	Inflation crisis	-0.0125	-0.28	-11.4%
Ballard-Rosa et al.	Sov. debt crisis	-0.0038	-0.09	-3.5%

Table A7: Full mechanical IVB decomposition. All controls are included regardless of collider plausibility. IVB/$\hat{\beta}$ is reported only when the treatment effect is significant ($|t| \geq 1.96$); "—" indicates $\hat{\beta} \approx 0$ or SE unavailable. *(continued)*

Study	Control	IVB	|IVB/SE|	IVB/$\hat{\beta}$ (%)
Ballard-Rosa et al.	Democracy (V-Dem)	0.0004	0.01	0.3%