Identification of Causal Diffusion Effects Using Stationary Causal Directed Acyclic Graphs*

Naoki Egami†

First Version: August 29, 2018
This Version: October 19, 2018

Abstract

Although social scientists have long been interested in the process through which ideas and behavior diffuse, the identification of causal diffusion effects, also known as peer effects, remains challenging. Many scholars consider the commonly used assumption of no omitted confounders to be untenable due to contextual confounding and homophily bias. To address this long-standing identification problem, I introduce a class of *stationary* causal directed acyclic graphs (DAGs), which represent the time-invariant nonparametric causal structure. I first show that this stationary causal DAG implies a new statistical test that can detect a wide range of biases, including the two types mentioned above. The proposed test allows researchers to empirically assess the contentious assumption of no omitted confounders. In addition, I develop a difference-in-difference style estimator that can directly correct biases under an additional parametric assumption. Leveraging the proposed methods, I study the spatial diffusion of hate crimes in Germany. After correcting large upward bias in existing studies, I find hate crimes diffuse only to areas that have a high proportion of school dropouts. To highlight the general applicability of the proposed approach, I also analyze the network diffusion of human rights norms. The proposed methodology is implemented in a forthcoming open source software package.

Keywords: Causal diagrams, Diffusion effects, Homophily bias, Peer effects, Social influence

*I thank Peter Aronow, Matt Blackwell, Tim Brackmann, Brandon de la Cuesta, Chris Felton, Romain Ferrali, Benjamin Fifield, Adam Glynn, Max Goplerud, Justin Grimmer, Andrew Guess, Robert Gulotty, Erin Hartman, Chad Hazlett, Amaney Jamal, Zhichao Jiang, In Song Kim, Gary King, Dean Knox, Shiro Kuriwaki, Chris Lucas, Ian Lundberg, Asya Magazinnik, Jonathan Mummolo, Santiago Olivella, Tyler Pratt, Marc Ratkovic, Ramona Rischke, James Robins, Tomoya Sasaki, Dustin Tingley, Rory Truex, Tyler VanderWeele, Yiqing Xu, Teppei Yamamoto, Soichiro Yamauchi, Xiang Zhou, Yang-Yang Zhou, members of Imai research group and Stewart Lab, and participants of seminars at Princeton and Harvard and the 2017 and 2018 Summer Meetings of the Political Methodology Society, for helpful comments. I am particularly grateful to Kosuke Imai, Rafaela Dancygier, and Brandon Stewart for their continuous encouragement and detailed comments.

†Ph.D. Candidate, Department of Politics, Princeton University, Princeton NJ 08544. Pre-doctoral Fellow, Department of Government, Harvard University, Cambridge MA 02138. Email: negami@princeton.edu, URL: http://scholar.princeton.edu/negami

1 Introduction

Social scientists have long been interested in how ideas and behavior diffuse across space, networks, and time. In political science, scholars study the spatial and temporal clusters of civil wars by analyzing how information, people, and goods move across space (Lake and Rothchild, 1998; Buhaug and Gleditsch, 2008). Political scientists have also investigated policy diffusion (Simmons and Elkins, 2004; Gilardi, 2010; Graham *et al.*, 2013), diffusion of social movements (Tarrow, 1994; Beissinger, 2007), democracies (Huntington, 1991; Pevehouse, 2002), norms (Keck and Sikkink, 1998; Hyde, 2011), and voting behavior (Nickerson, 2008; Sinclair, 2012; Alt *et al.*, 2018). Economists and sociologists have examined the diffusion of innovations (Rogers, 1962), job attainment (Granovetter, 1973), and school achievement (Sacerdote, 2001). Researchers in public health have focused on the spread of infectious disease (Halloran and Struchiner, 1995; Morozova *et al.*, 2018) and health behavior (Christakis and Fowler, 2007). In each of these research areas, a growing number of scholars aim to estimate the causal impact of diffusion dynamics. The goal is to learn about causal diffusion processes in which an outcome of one unit causes, not just correlates with, an outcome of another unit.

Despite its importance, the identification of causal diffusion effects, also known as peer effects or social influence, is one of the most challenging causal inference problems (Galton, 1889; Manski, 1993). Although commonly-used statistical methods, including conventional regression models and spatial econometric models (e.g., Anselin, 2013), require the assumption of no omitted confounders, this assumption is often untenable because both outcomes and confounders are interdependent across space and networks. In particular, two special types of confounding/bias are well-known (VanderWeele and An, 2013). When there exist some unobserved contextual factors that affect multiple units, we suffer from *contextual confounding* — we cannot distinguish whether units affect one another through diffusion processes or units are jointly affected by the shared unobserved contextual variables. *Homophily bias* arises when the formation of network ties is affected by some unobserved characteristics. We cannot discern whether connected units exhibit similar outcomes because of diffusion or because they selectively become connected with others who have similar unobserved characteristics. Emphasizing these biases, a number of papers across disciplines criticize existing diffusion studies (e.g., Buhaug and Gleditsch, 2008; Lyons, 2011). In fact, causal diffusion effects are often found to be overestimated by a large amount, for example, by 300 – 700% (Aral *et al.*, 2009; Eckles and Bakshy, 2017). Shalizi and Thomas (2011) argue that it is nearly impossible to credibly estimate causal diffusion effects from observational studies.

In this paper, I address this long-standing identification challenge by introducing a class of *stationary* causal directed acyclic graphs (DAGs), which represent the time-invariant nonparametric causal structure. Using this new class of stationary causal DAGs, I make two contributions: I propose a statistical test and an estimator to detect and correct a wide range of biases, including contextual confounding and homophily bias. The proposed approach provides a new way to credibly identify causal diffusion effects by directly addressing concerns about omitted confounders.

This paper proposes stationary causal DAGs (Section 2) to overcome a dilemma of existing approaches, which have either been agnostic about the underlying DAG or assumed full knowledge of its structure. On the one hand, causal diffusion analysis without any DAG structure has been intractable due to contextual confounding and homophily bias. On the other hand, researchers often cannot justify their full knowledge of the underlying DAG structure in applied contexts. Stationary causal DAGs use a simple time-invariant structure to formalize the underlying causal diffusion process, without assuming its full structure. They only require the existence of causal relationships among variables – not the effect or sign of such causal relationships – to be stable over time.

Making use of this general class of stationary causal DAGs, I first propose a new statistical test of the no omitted confounders assumption (Section 3). I prove that the proposed test – using a lagged dependent variable – can detect a wide class of biases all at once, including contextual confounding and homophily bias. With this test, researchers can statistically evaluate whether they adjust for all relevant confounders, rather than simply assuming the validity of their confounder adjustment. Formally, this placebo test assesses whether a lagged dependent variable is conditionally independent of the treatment variable. Statistical properties of this test are based on a new theorem, which states that under stationary causal DAGs, the no omitted confounders assumption is equivalent to the conditional independence of a lagged dependent variable and the treatment variable.

Furthermore, I develop a new bias-corrected estimator that can directly remove biases under an additional parametric assumption (Section 4). It subtracts the bias detected by the placebo test from a biased estimator. I show that this estimator can correct biases under a parametric assumption that the effect and imbalance of unobserved confounders are constant over time. This method is complementary to the proposed placebo test – while it requires a stronger parametric assumption about the underlying causal DAGs, it can directly correct biases. I also demonstrate that this proposed estimator is closely connected to the widely-used difference-in-difference estimator (Card and Krueger, 1994).

This article builds on a growing literature of causal diffusion effects (e.g., Shalizi and Thomas, 2011; Goldsmith-Pinkham and Imbens, 2013; Ogburn and VanderWeele, 2014; O'Malley *et al.*, 2014; Shalizi and McFowland III, 2016). In addition to research on the use of experimental or quasi-experimental design (Duflo and Saez, 2003; Bramoullé *et al.*, 2009; Fowler and Christakis, 2010; An, 2015; Eckles *et al.*, 2016), a series of papers address potential problems of omitted confounders by deriving tests or bounds for causal diffusion effects. VanderWeele *et al.* (2012) show that after controlling for homophily bias and contextual confounding, the spatial autoregressive model can be used to test the null hypothesis of zero diffusion effects. Anagnostopoulos *et al.* (2008)'s test also evaluates the same null hypothesis of no diffusion effects. To compute bounds for diffusion effects, Ver Steeg and Galstyan (2010, 2013) examine a specific causal DAG only with homophily and diffusion, and VanderWeele (2011) proposes sensitivity analysis methods. This paper shares concerns about the no omitted confounders assumption. However, instead of testing the null hypothesis of zero diffusion effects or deriving bounds, this paper focuses on the point identification of causal diffusion effects.

This paper also draws upon emerging literature of negative controls, also known as placebo variables (Lipsitch *et al.*, 2010; Tchetgen Tchetgen, 2013). In particular, this paper extends recent studies using negative controls in panel data settings (Sofer *et al.*, 2016; Flanders *et al.*, 2017; Miao and Tchetgen Tchetgen, 2017) to the identification of causal diffusion effects. The proposed methods differ from the previous literature in that I introduce a placebo test and a bias-corrected estimator by exploiting a general class of stationary causal DAGs rather than one specific causal DAG. Finally, causal DAGs (Pearl, 2009) are useful not only for causal identification but also for asymptotic statistical inference. van der Laan (2014) and Ogburn *et al.* (2017) offer one of the first foundations to use causal directed acyclic graphs or nonparametric structural equation models for network data. Tchetgen Tchetgen *et al.* (2017) provide an alternative approach using chain graphs. Because these recent papers develop theories of statistical inference in a network asymptotic regime, they are complementary to the methods proposed in this paper that focus on the identification of causal diffusion effects.

Two Applications: Spatial and Network Diffusion

Leveraging the proposed methods, I study the spatial diffusion of hate crimes against refugees in Germany. In particular, I analyze a data set documented in Benček and Strasheim (2016) and expanded as part of a joint project (Dancygier, Egami, Jamal, and Rischke, 2018). Observing temporal and spatial clusters of hate crimes, existing studies have developed a number of theories on how hate crimes diffuse across space (e.g., Koopmans and Olzak, 2004; Myers, 2000). The central argument in such studies is that one incidence of hate crime can trigger another incidence, which again induces another, and can lead to waves of hate crimes (e.g., Braun, 2011). It is, therefore, of theoretical and policy interest to empirically estimate the spatial dynamics of hate crimes by separating out spurious correlations due to contextual confounding. Using the placebo test and the bias-corrected estimator, I find that the average effect of spatial diffusion is small, in contrast to existing studies (Braun, 2011; Jäckle and König, 2016). Further investigation of heterogeneous causal effects reveals that the spatial diffusion effect is large only for counties that have a high proportion of school dropouts. This finding suggests that the spatial diffusion of hate crimes is concentrated in places with low educational performance. This is consistent with rich qualitative and quantitative evidence that hate crime is often a problem of young people (Green *et al.*, 2001). Connecting to this application throughout the paper, I will describe how the proposed methods can facilitate spatial diffusion analysis. The main empirical analysis appears in Section 5.1.

To demonstrate that the proposed methods can be applicable to network diffusion problems in the same way, I analyze the diffusion of human rights norms (Greenhill, 2016). Extending an influential work (Johnston, 2001), Greenhill (2016) emphasizes that intergovernmental organizations (IGOs) offer forums in which high-level policymakers from different countries regularly meet, discuss policy issues, and learn from one another. He then argues that these interactions between government representatives "catalyze the process of international norm diffusion" (Greenhill, 2010, p. 129). The central research question is how much human rights norms diffuse among states through networks based on IGO connections. The original analysis recognizes concerns about homophily bias and carefully adjusts for relevant confounders, including network-related covariates, such as spatial and cultural similarity between states. However, the proposed methods reveal that a large amount of bias remains. After correcting the bias, an estimate of the average causal diffusion effect is close to zero, in contrast to the original findings. This reanalysis illustrates that in network diffusion studies, we can suffer from significant bias even after adjusting for network-related variables in addition to conventional control variables, such as GDP and the Polity score. Importantly, it is difficult to know the consequence of such bias on substantive findings without the proposed methods. Throughout the paper, I use this study as an example of network diffusion problems and demonstrate how the proposed approach can improve network diffusion analysis. The primary empirical analysis appears in Section 5.2.

Both applications highlight the large differences in substantive conclusions that can result from contextual confounding and homophily bias. By directly taking into account these biases, the proposed methods enable more credible and defensible causal diffusion analysis.

2 A Framework of Causal Diffusion Analysis

Causal diffusion refers to a process in which an outcome of one unit influences an outcome of another unit over time (Shalizi and Thomas, 2011; VanderWeele *et al.*, 2012). This definition generalizes and formalizes the standard definition in political science (Elkins and Simmons, 2005) and other social sciences, "diffusion as the process by which prior adoption of a trait or practice in a population alters the probability of adoption for remaining non-adopters" (Strang, 1991, p. 325). In this section, I first define the average causal diffusion effect and then describe challenges for its causal identification. Finally, I introduce a class of stationary causal directed acyclic graphs to represent the time-invariant nonparametric causal structure. I make use of this stationary causal DAG to develop a test of the no omitted confounders assumption in Section 3 and a bias-corrected estimator in Section 4.

2.1 The Setup

In causal diffusion analysis, we consider n units over T time periods. Let Y_{it} be the outcome for unit i at time t for $i \in \{1, \ldots, n\}$ and $t \in \{0, 1, \ldots, T\}$. Use $\mathbf{Y}_t$ to denote a vector $(Y_{1t}, \ldots, Y_{nt})$, which contains the outcomes at time t for n units. The outcome Y_{it} could be binary indicating whether county i experiences at least one hate crime in month t or continuous representing human rights performance of state i in year t, measured by the Personal Integrity Rights (PIR) score (Greenhill, 2016).

To encode spatial or network connections between these n units, I follow the standard spatial econometric literature (Anselin, 2013) and use a distance matrix $\mathbf{W}$. The ith row of this distance matrix, $\mathbf{W}_i$, represents connections between unit i and other units. In practice, researchers specify this distance matrix to reflect the underlying relationship responsible for the diffusion process they study. For instance, in the study of hate crime diffusion, it is of interest to estimate how much hate crimes in one county diffuse to other spatially proximate counties. Here, the distance matrix $\mathbf{W}$ could encode physical distance between counties where W_{ij} might be an inverse of the distance between district i and j, i.e., the closer are districts i and j, the larger is W_{ij}. In the study of human rights norms (Greenhill, 2016), diffusion is theorized to operate through the network connection that states form in intergovernmental organizations (IGOs). If states i and j share memberships in at least one IGO, W_{ij} takes a positive value and zero otherwise. When two states share more IGO memberships, W_{ij} is larger, which captures an idea that states are more strongly affected by other states with which they share more IGO memberships. As shown in these two examples, the distance matrix can naturally take into account the different strength of ties as a weighted matrix.

It can also incorporate directed connections, such as friendship networks, when necessary. Finally, I define *neighbors* $\mathcal{N}_i$ to be other units who are connected with a given unit i, i.e., $\mathcal{N}_i \equiv \{j : W_{ij} \neq 0\}$.

I rely on potential outcomes (Neyman, 1923; Rubin, 1974) to formally define causal diffusion effects. Based on the tradition of spatial econometrics and typical political science applications (Anselin, 2013; Franzese and Hays, 2007), this paper focuses on the weighted average of the neighbors' outcomes $\mathbf{W}_i^\top \mathbf{Y}_t$ as the treatment variable. Although I keep this setup throughout the paper, the methods I introduce in this paper can be easily applied to other definitions of the treatment variable. I use $D_{it} \equiv \mathbf{W}_i^\top \mathbf{Y}_t$ to denote the treatment variable and let $Y_{i,t+1}(d)$ represent the potential outcome variable of unit i at time $t+1$ if the unit receives the treatment $D_{it} = d$.

In the hate crime diffusion study, the treatment variable for each county is the weighted proportion of neighboring counties that experience hate crimes in month t. The binary potential outcome $Y_{i,t+1}(d)$ then represents whether county i experiences at least one hate crime in month $t+1$ if $d\%$ of neighboring counties suffer from hate crimes in month t. Similarly, in the study of human rights norms diffusion (Greenhill, 2016), the potential outcome $Y_{i,t+1}(d)$ is the PIR score of state i in year $t+1$ if the weighted average of the PIR scores of its IGO partners in year t (treatment) is set to d.

2.2 Definition of the Average Causal Diffusion Effect

Having set up the potential outcomes, I now introduce the causal diffusion effect as the comparison of two different potential outcomes. In the hate crime diffusion study, researchers might be interested in analyzing how the risk of hate crimes changes due to the incidence of hate crimes in neighboring counties. For example, they can compare the case when none of the neighboring counties experience hate crimes with the case when 30% of the neighboring counties suffer from hate crimes.

Formally, I define the *average causal diffusion effect* (ACDE) at time $t+1$ to be the average causal effect of the treatment variable D_{it} on the outcome at time $t+1$. It is the comparison between the potential outcome under a higher value of the treatment $D_{it} = d^H$ and the potential outcome under a lower value of the treatment $D_{it} = d^L$.

Definition 1 (The Average Causal Diffusion Effect)

The average causal diffusion effect (ACDE) at time $t+1$ is defined as,

$$\tau_{t+1}(d^H, d^L) \equiv \mathbb{E}[Y_{i,t+1}(d^H) - Y_{i,t+1}(d^L)], \tag{1}$$

where d^H and d^L are two constants specified by researchers.

The ACDE could quantify how much the risk of having hate crimes in the next month changes if we see more hate crimes in neighboring counties this month. This captures how much hate crimes diffuse across space over time. The ACDE could also represent how much human rights practices, measured by PIR scores, would change next year if the IGO partners have higher PIR scores this year. This quantity measures how much human rights norms diffuse over time through the IGO network connections.

Finally, I introduce an assumption about the measurement of outcomes. I assume that we observe the outcome variable at time $t = 0$ and then sequentially observe the outcome variables at time $t = 1, \ldots, T$. This assumption requires that we avoid the temporal aggregation problem (Granger, 1988) that can mask the dynamics of the underlying diffusion process.

Assumption 1 (Sequential Consistency)
For each unit, we observe the outcome variable at time $t = 0$. Then, for every unit at every time period $t = 1, \ldots, T$, one of the potential outcome variables is observed, and the realized outcome variable for unit i at time $t + 1$ is denoted by

$$Y_{i,t+1} = Y_{i,t+1}(D_{it}). \tag{2}$$

The violation of this assumption implies simultaneity bias, that is, the treatment variable and the outcome variable simultaneously cause each other (Danks and Plis, 2013; Hyttinen *et al.*, 2016). In the literature of causal diffusion analysis, the importance of this assumption has recently received much attention because without it, the causal order of the treatment and outcome becomes ambiguous and causal diffusion effects are no longer well-defined (Lyons, 2011; Ogburn and VanderWeele, 2014; Ogburn *et al.*, 2017). See Joffe and Robins (2009) and Zhang *et al.* (2011) for a similar problem in the structural nested model and g-estimation. I maintain this assumption throughout the paper unless otherwise noted. In practice, researchers can make this assumption more plausible by measuring outcomes frequently enough. For example, the assumption could be more tenable when scholars can measure the incidence of hate crimes monthly rather than annually.

2.3 Identification under No Omitted Confounders Assumption

In this section, I describe the widely used identification assumption of no omitted confounders and explain pervasive concerns about its violation. This assumption states that all relevant confounders are in a selected set of control variables. In practice, this assumption could imply that the number of hate crimes in neighboring counties is as-if random given control variables.

Formally, the assumption states that the potential outcomes at time $t + 1$ are independent of a joint distribution of neighbors' outcomes at time t given control variables.

Assumption 2 (No Omitted Confounders)
For $i = 1, 2, \ldots, n$,

$$Y_{i,t+1}(d) \perp\!\!\!\perp \{Y_{jt}\}_{j \in \mathcal{N}_i} \mid \mathbf{C}, \tag{3}$$

for $d \in \mathcal{D}$ where $\mathcal{D}$ is the support of D_{it}, and $\mathbf{C}$ is a set of pretreatment variables, which I call a *control set.*

Under this assumption of no omitted confounders, we can estimate the ACDE without bias.

Result 1 (Identification under No Omitted Confounders Assumption)
Under Assumptions 1 and 2, the ACDE at time $t+1$ is identified as,

$$\tau_{t+1}(d^H, d^L) \;\; = \;\; \int_{\mathcal{C}} \left\{ \mathbb{E}[Y_{i,t+1} \mid D_{it} = d^H, \mathbf{C} = \mathbf{c}] - \mathbb{E}[Y_{i,t+1} \mid D_{it} = d^L, \mathbf{C} = \mathbf{c}] \right\} dF_{\mathbf{C}}(\mathbf{c}),$$

where $F_{\mathbf{C}}(\mathbf{c})$ is the cumulative distribution function of $\mathbf{C}$ and we assume that $\Pr(D_{it} = d^H | \mathbf{C} = \mathbf{c}) > 0$ and $\Pr(D_{it} = d^L | \mathbf{C} = \mathbf{c}) > 0$ for $i = 1, \ldots, n$ and all $\mathbf{c} \in \mathcal{C}$ where $\mathcal{C}$ is the support of $\mathbf{C}$ (Imbens and Rubin, 2015). Control set $\mathbf{C}$ includes summary statistics of $\mathbf{W}_i$, such as the number of neighbors (Shalizi and Thomas, 2011).

This result implies that as long as the no omitted confounders assumption is satisfied, researchers can estimate the ACDE by estimating the conditional expectation $\mathbb{E}[Y_{i,t+1}|D_{it}, \mathbf{C}]$ and then averaging it over the empirical distribution of control variables $\mathbf{C}$. In practice, we can estimate $\mathbb{E}[Y_{i,t+1}|D_{it}, \mathbf{C}]$ through regression, matching, weighting, or other approaches (Imbens, 2004; Ho *et al.*, 2007).

Although many empirical studies of diffusion make the assumption of no omitted confounders, it is widely known that the assumption is often questionable in practice (Manski, 1993; Shalizi and Thomas, 2011; VanderWeele and An, 2013). Indeed, there are numerous papers across disciplines criticizing existing observational diffusion studies for their implausible assumptions of no omitted confounders; to name a few, Buhaug and Gleditsch (2008) and Houle *et al.* (2016) in political science, Gibbons and Overman (2012) in economics, and Lyons (2011) and Shalizi and Thomas (2011) in public health.

This concern over the assumption of no omitted confounders (Assumption 2) is pervasive mainly because it implies the absence of two well-known types of biases: contextual confounding and homophily bias. *Contextual confounding* – the primary focus of the spatial diffusion literature – can exist when units share some unobserved contextual factors. For example, in the study of hate crime diffusion, the risk of having hate crimes is likely to be affected by some economic policies, which often affect multiple counties at the same time. In this case, researchers might observe spatial clusters of hate crimes even without diffusion. When researchers fail to account for this common contextual variable, their estimates suffer from contextual confounding, and they might overestimate the diffusion effect.

Another well-known type of bias is *homophily bias* – the main concern in the network diffusion literature. This bias arises when units become connected due to their unobserved characteristics. For example, in the study of norm diffusion (Greenhill, 2016), human rights practices in a given state might be affected by its political culture, which is also likely to influence what kinds of IGOs that the state joins. As a result, even without any diffusion, states that share IGO memberships can have similar cultures and hence similar levels of human rights practices. When we cannot adjust for such variables that affect both IGO memberships and human rights practices, an estimate of causal diffusion effects suffers from homophily bias.

In the next subsection, I introduce a new class of causal directed acyclic graphs as a framework of causal diffusion analysis. This graph representation clarifies how contextual confounding and homophily bias differ although both are examples of omitted variable bias. Most importantly, leveraging the framework, I develop new statistical methodologies to address both contextual confounding and homophily bias in Sections 3 and 4.

2.4 Stationary Causal Directed Acyclic Graphs

In this paper, I formalize underlying diffusion processes using a new class of stationary causal directed acyclic graphs (DAGs), which represent the time-invariant nonparametric causal structure. Conventionally, researchers have taken one of the two approaches in causal diffusion analysis: be agnostic about the underlying DAG or assume full knowledge of its structure. On the one hand, researchers may wish to avoid any assumption about the underlying DAG. However, causal diffusion analysis is then intractable due to contextual confounding and homophily bias. Methods agnostic about the underlying causal structure have lacked statistical guarantees. On the other hand, researchers are often unable to justify their full knowledge of the underlying DAG structure in applied contexts, even though the identification problem becomes mathematically straightforward. In contrast to these previous approaches, the proposed class of stationary causal DAGs uses a simple time-invariant structure to formalize the underlying causal diffusion process, without assuming its full structure, so that methods can be widely applicable in empirical settings. In Sections 3 and 4, by making use of this additional structural stationarity, I develop a new statistical test of the no omitted confounders assumption and a bias-corrected estimator. I review basic causal DAG terminologies in Appendix A. Comprehensive introductions to causal DAGs can be found in Pearl (2009).

Without loss of generality, I consider a class of causal DAGs that contain at least all the variables observed by a researcher. Within this class, I define *stationary* causal DAGs to be causal DAGs with a nonparametric causal structure that is time-invariant. Here, the causal structure refers to the existence of causal relationships among variables (the existence of arrows in causal DAGs), not the effect or sign of such causal relationships. For example, in the study of hate crime diffusion, suppose that the unemployment rate is a part of the causal model for hate crimes in one month. Then, stationary causal DAGs require that the unemployment rate should stay a part of the causal model for hate crimes in the next month. Importantly, the effect of the unemployment rate can be changing over time; the only requirement is about the existence of the causal relationship. I provide a formal definition below.

Definition 2 (Stationary Causal Directed Acyclic Graphs)
Consider variables in a causal directed acyclic graph $\mathcal{G}$ that have more than one child or have at least one parent. Among these variables, distinguish two types; the time-independent variable Z_i and the time-dependent variable X_{it}. A causal directed acyclic graph $\mathcal{G}$ is said to be stationary when it meets the following conditions.

(2.1) $X_{it} \in \text{PA}(X_{i,t+1}) \quad \text{for } i \in \{1, \ldots, n\} \text{ and } \ t = 0, \ldots, T-1.$

(2.2) For $i, i' \in \{1, \ldots, n\},\ \exists\ t, k \text{ s.t. } X_{it} \in \text{PA}(\widetilde{X}_{i',t+k}) \Rightarrow\ X_{it'} \in \text{PA}(\widetilde{X}_{i',t'+k})$ for all $t' = 0, \ldots, T-k.$

(2.3) For $i, i' \in \{1, \ldots, n\},\ \exists\ t \text{ s.t. } Z_i \in \text{PA}(X_{i't}) \ \Rightarrow\ Z_i \in \text{PA}(X_{i't'})$ for all $t' = 0, \ldots, T,$

where $A \in \text{PA}(B)$ indicates that variable A is a parent of variable B.

Condition 2.1 requires that all time-dependent variables that have at least one parent be affected by their own lagged variables. This condition is more plausible when the time intervals are shorter. Condition 2.2 means that if two time-dependent variables have a child-parent relationship at one time period, the same causal relationship should exist for all other time periods. Similarly, Condition 2.3 requires that if a time-independent variable is a parent of a time-dependent variable at one time period, the same child-parent relationship should exist at all other time periods. The last two requirements are at the heart of the stationary causal DAG – the existence of causal relationships should be stable over time. I provide examples of stationary causal DAGs in Figure 1, representing basic diffusion dynamics, contextual confounding and homophily bias.

In many social and biomedical science applications, scholars often assume that the underlying causal structure is stable over time, even if causal effects might change substantially over time. Hence, this structural stationarity is often a natural requirement for DAGs. In fact, causal DAGs in several important papers about causal diffusion effects (Shalizi and Thomas, 2011; O'Malley *et al.*, 2014; Ogburn and VanderWeele, 2014) are examples of the proposed general class of stationary causal DAGs. Causal DAGs in the causal discovery literature often impose a similar but stronger condition (Danks and Plis, 2013; Hyttinen *et al.*, 2016). They often assume that variables are affected only by one-time lag (also known as the first order Markov assumption) and this structure is time-invariant. In contrast, stationary DAGs introduced here allow for any higher order temporal dependence (Condition 2.2).

When the causal structure changes at some time, the underlying DAG is not stationary. However, if researchers know the time when the underlying structure changes, we can still make use of stationary DAGs separately for before and after this time point. For example, the network diffusion dynamics of human rights norms might be radically different before and after the end of the Cold War. Then, we can analyze data before and after 1991 separately, each as a stationary causal DAG.

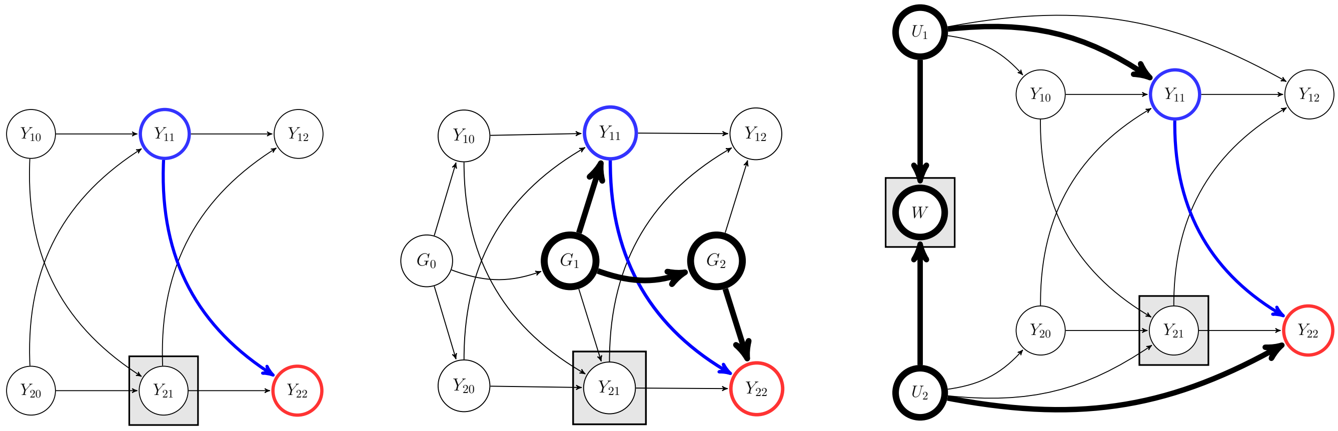
Illustration of Contextual Confounding and Homophily Bias

Here, I use stationary causal DAGs to illustrate contextual confounding and homophily bias. With this DAG representation, it becomes clear that both types of biases are due to unblocked back-door paths that have different structural roots.

I begin by introducing a stationary causal DAG for a simple diffusion model (Figure 1 (a)) where there exists only diffusion. This causal DAG in Figure 1 (a) has six nodes representing outcome variables Y_{it} for two units $i \in \{1, 2\}$ over three time periods $t \in \{0, 1, 2\}$. The arrows between these six nodes represent direct causal relationships where $A \rightarrow B$ means that variable A can have a direct causal effect on variable B. Without loss of generality, I focus on the causal diffusion effect of Y_{11} on Y_{22} where Y_{11} is the treatment variable (blue), Y_{22} is the outcome variable (red), and the causal arrow of interest $Y_{11} \rightarrow Y_{22}$ is colored blue. I use gray square boxes to indicate variables that are adjusted for.

First, this causal DAG in Figure 1 (a) naturally encodes stationarity: $Y_{i,t+1}$ is directly affected by Y_{it} and the existence of causal diffusion effects is time-invariant, e.g., $Y_{11} \rightarrow Y_{22}$ and $Y_{10} \rightarrow Y_{21}$. This causal DAG also shows that we can satisfy the assumption of no omitted confounders (Assumption 2) by adjusting for Y_{21} (the gray square box), which blocks all back-door paths from Y_{11} to Y_{22}. When diffusion is the only mechanism behind correlated outcomes among connected units, researchers only need to adjust for previous outcomes.

Next, I introduce a slightly more complicated stationary causal DAG to consider contextual confounding. As defined in the previous subsection, contextual confounding results from unadjusted contextual factors. In addition to six nodes in Figure 1 (a), a causal DAG in Figure 1 (b) has three additional contextual variables G_t with $t \in \{0, 1, 2\}$. In the study of



(a) Diffusion-Only

(b) Contextual Confounding

(c) Homophily Bias

Figure 1: Stationary Causal Directed Acyclic Graphs for Diffusion, Contextual Confounding, and Homophily Bias. Note: The causal diffusion effect of Y_{11} (blue) on Y_{22} (red) is the quantity of interest and the causal arrow $Y_{11} \to Y_{22}$ is colored blue. Gray square boxes indicate control variables. Figure (a) shows the diffusion-only model. Figure (b) illustrates contextual confounding as a back-door path $Y_{11} \leftarrow G_1 \to G_2 \to Y_{22}$ (the thick black path). Figure (c) shows homophily bias as a back-door path $Y_{11} \leftarrow U_1 \to \boxed{W} \leftarrow U_2 \to Y_{22}$ (the thick black path).

hate crime diffusion, this contextual variable could be some economic policies affecting multiple counties. This graph shows that when researchers only adjust for the previous outcome Y_{21}, they suffer from contextual confounding, which is represented as a back-door path $Y_{11} \leftarrow G_1 \to G_2 \to Y_{22}$ (the thick black path in Figure 1 (b)). If researchers can adjust for G_1 or G_2, this back-door path is blocked, and the assumption of no omitted confounders is satisfied. Although this graph is the simplest representation of contextual confounding, it shows that researchers have to adjust for all contextual variables that are spatially correlated. Especially in spatial diffusion analysis, because many variables are naturally correlated across space (e.g., Buhaug and Gleditsch, 2008), it is difficult to observe all relevant contextual variables and avoid contextual confounding.

Finally, I examine a stationary causal DAG representing homophily bias (Figure 1 (c)). In addition to six nodes in the first causal DAG, this causal DAG contains two unit-level variables U_i with $i \in \{1, 2\}$ and variable W representing the connection between two units. This variable W takes 1 when two units are connected, and it takes 0 otherwise. In the study of human rights norm diffusion, variable U could represent political culture and variable W could indicate whether two states share memberships in at least one IGO.

The core of the homophily problem is that this connection variable W is affected by two unit-level characteristics U_1 and U_2. Shalizi and Thomas (2011) show that the connection variable W is always, but often implicitly, adjusted for in any diffusion analysis so that researchers can compare observations with similar spatial/network pre-treatment characteristics. This is why there is a gray square box around W in Figure 1 (c). Technically speaking, we need to include directed arrows from W to six outcome variables, but those arrows cannot form any unblocked back-door path because W is always adjusted for as explained above. Therefore, we exclude them in Figure 1 (c) for the sake of visual simplicity. Shalizi and Thomas (2011) also show that when researchers do not adjust for unit-level variables explaining the connection W, they suffer from homophily bias. In the causal DAG, this homophily bias is represented as a collider bias through a back-door path, $Y_{11} \leftarrow U_1 \rightarrow \boxed{W} \leftarrow U_2 \rightarrow Y_{22}$ (the thick black path in Figure 1 (c)). This simple graph illustrates that the assumption of no omitted confounders requires observing all variables explaining connections between units. In practice, because subjects of the study often form their network connections long before researchers observe their behavior, it is difficult to adjust for variables explaining their connections and avoid homophily bias.

3 A Placebo Test to Detect Biases

Identification of the ACDE is challenging in practice. The central concern is that the commonly-used identification assumption of no omitted confounders is often difficult to justify, as discussed in the previous section. In this section, I make use of stationary DAGs to develop a new statistical test of this contentious assumption. I prove that the proposed test – using a lagged dependent variable as a general placebo outcome – can detect a wide class of biases, including contextual confounding and homophily bias. This placebo test helps the credible identification of causal diffusion effects by statistically assessing the validity of the confounder adjustment. Although the main theorem behind the placebo test utilizes theories of causal DAGs, the application of the method does not require familiarity with causal DAGs or full knowledge of the underlying causal DAG. An investigator, however, needs to know the time ordering of outcomes and observed confounders. I first introduce the main theorem along with its basic intuition (Section 3.1) and then describe a general procedure of the placebo test (Section 3.2). In Section 3.3, I provide some extensions.

3.1 Equivalence Theorem

The proposed statistical test of the no omitted confounders assumption exploits a lagged dependent variable as a placebo outcome. In particular, it tests the assumption of no omitted confounders by assessing whether a lagged dependent variable is conditionally independent of the treatment variable. I show that if a lagged dependent variable is conditionally independent of the treatment variable, it provides statistical evidence for the assumption of no omitted confounders. Conversely, when a lagged dependent variable and the treatment are conditionally dependent, it implies that some relevant confounders are omitted. This placebo test is based on an equivalence theorem (Theorem 1), which states that the conditional independence of the lagged outcome and the treatment is equivalent to the no omitted confounders assumption under stationary causal DAGs. Proof of this theorem makes use of the time-invariant structure of stationary causal DAGs.

Before deriving this main theorem, I develop some intuition to understand why a lagged dependent variable can serve as a placebo outcome to detect biases under stationary causal DAGs. The basic idea is simple: the structure of spurious correlation between the main outcome and the treatment is similar to the one between the placebo outcome (a lagged dependent variable) and the treatment. Using this structural similarity, we can assess the existence of spurious correlation between the main outcome and the treatment (whether the no omitted confounders assumptions holds) by checking whether there exists spurious correlation between the placebo outcome and the treatment.

Using causal DAG terminologies, the same intuition can be restated that back-door paths between the main outcome and the treatment are similar to those between the placebo outcome and the treatment. For example, a back-door path in Figure 1 (b) representing contextual confounding ($Y_{11} \leftarrow G_1 \rightarrow G_2 \rightarrow Y_{22}$) is similar to the one between the treatment and the placebo outcome ($Y_{11} \leftarrow G_1 \rightarrow Y_{21}$), although they are not the same. Another back-door path in Figure 1 (c) representing homophily bias ($Y_{11} \leftarrow U_1 \rightarrow \boxed{W} \leftarrow U_2 \rightarrow Y_{22}$) is also similar to a back-door path to the placebo outcome ($Y_{11} \leftarrow U_1 \rightarrow \boxed{W} \leftarrow U_2 \rightarrow Y_{21}$). The time-invariant structure of stationary DAGs enables the structural similarity of these back-door paths.

Exploiting this structural similarity, the proposed placebo test checks whether there is any unblocked back-door path between the main outcome and the treatment by testing the existence of unblocked back-door paths between the placebo outcome and the treatment. Because the treatment has no causal effect on the placebo outcome, the treatment should be conditionally independent of the placebo outcome if there is no unblocked back-door path between the main outcome and the treatment (i.e., the assumption of no omitted confounders holds). This conditional independence of the placebo outcome and treatment is what the proposed test checks. If the conditional independence does not hold, it statistically implies the violation of the no omitted confounders assumption. If instead, the conditional independence holds, it provides strong statistical evidence for the no omitted confounders assumption.

Now, I formally prove the main theoretical result justifying the placebo test (details of its procedure are in Section 3.2). It states that, when the underlying causal DAG is stationary, the assumption of no omitted confounders is equivalent to the conditional independence of the simultaneous outcomes given *a placebo set.* Researchers can derive this placebo set by a simple deterministic rule introduced below. The following formal result implies that researchers can use a lagged dependent variable as a general placebo outcome to statistically assess the assumption of no omitted confounders.

Theorem 1 (Equivalence between No Omitted Confounders Assumption and Conditional Independence of Simultaneous Outcomes) Under Assumption 1, for every causal model faithful to stationary causal DAGs,

$$Y_{i,t+1}(d) \perp\!\!\!\perp \{Y_{jt}\}_{j \in \mathcal{N}_i} \mid \mathbf{C} \iff Y_{it} \perp\!\!\!\perp \{Y_{jt}\}_{j \in \mathcal{N}_i} \mid \mathbf{C}^P, \tag{4}$$

where a placebo set $\mathbf{C}^P$ is defined as

$$\mathbf{C}^P \equiv \{\mathbf{C}, \mathbf{C}^{(-1)}, \{Y_{j,t-1}\}_{j \in \mathcal{N}_i}\} \setminus \mathrm{Des}(Y_{it}), \tag{5}$$

where $\mathbf{C}^{(-1)}$ is a lag of the time-dependent variables in $\mathbf{C}$, $\{Y_{j,t-1}\}_{j \in \mathcal{N}_i}$ is a lag of the treatment variable, and $\mathrm{Des}(Y_{it})$ is a descendant of Y_{it}, i.e., a set of variables affected by Y_{it}. The violation of the no omitted confounders assumption, if any, is assumed to be proper in a sense defined below.

Proof is in Appendix C.2. Now I discuss several important implications of the theorem. First, in Equation (4), the assumption of no omitted confounders (the left-hand side) is proven to be equivalent to the conditional independence of the observed outcome of individual i and her neighbors' outcomes at the same time period given a placebo set (the right-hand side). Because this right-hand side is observable and testable, this theorem directly implies that we can statistically assess the assumption of no omitted confounders by testing the conditional independence of the simultaneous outcomes on the right-hand side of the equation. This theorem implies a procedure of the placebo test I introduce in the next subsection.

The difference between a control set $\mathbf{C}$ and a placebo set $\mathbf{C}^P$ is to guarantee that unblocked back-door paths between the main outcome and the treatment are as similar as possible to those between the placebo outcome and the treatment. To derive this placebo set, we only need to know which variables in the control set are time-dependent and which variables are affected by outcomes at time t. The former information is often readily available, and the latter one is essentially the same as the information we usually use to avoid post-treatment bias in the standard causal inference problem.

Finally, the theorem requires one technical condition – the violation of the no omitted confounders assumption, if any, is *proper*. It means that bias (i.e., the violation of the no omitted confounders assumption) is in fact driven by omitted variables. Bias is not proper when the only source of bias is the misadjustment of the lag structure of observed covariates. Intuitively, the proposed placebo test is designed to detect bias from omitted variables and hence, it cannot detect biases that merely come from misunderstandings about the lag structure of observed variables. Importantly, contextual confounding and homophily bias are proper, and hence within the scope of this theorem. I provide a formal definition and examples in Appendix C.1.

3.2 Procedure

The proposed placebo test is easy to implement – it has only two steps. After introducing a general procedure of the placebo test, I describe each step in order.

A Placebo Test

For a given control set $\mathbf{C}$, the following test statistically assesses whether the control set contains all confounders, i.e., whether the assumption of no omitted confounders holds (Assumption 2).

Step 1: Derive a placebo set $\mathbf{C}^P$ from a selected control set $\mathbf{C}$ based on Equation (5).

Step 2: Test the conditional independence, $Y_{it} \perp\!\!\!\perp \{Y_{jt}\}_{j \in \mathcal{N}_i} \mid \mathbf{C}^P$.

In Step 2, if the conditional independence does not hold, it implies the violation of the no omitted confounders assumption. In contrast, if the conditional independence holds, it provides strong statistical evidence for the no omitted confounders assumption.

Deriving A Placebo Set The first step of the proposed placebo test is to derive a placebo set $\mathbf{C}^P$ from a selected control set $\mathbf{C}$ (Equation (5)). This intermediate step guarantees that unblocked back-door paths between the main outcome and the treatment are similar to those between the placebo outcome and the treatment. This step follows a simple deterministic rule: (1) add lags of existing control variables and a lag of the treatment variable and then (2) remove all the variables affected by outcomes at time t. As one simple example, I consider the hate crime diffusion study where the main outcome is the incidence of hate crimes at time $t+1$ and the treatment is the proportion of neighbors that experienced hate crimes at time t, i.e., $Y_{i,t+1} \equiv \texttt{Hate Crimes}_{t+1}$ and $D_{it} \equiv \texttt{Prop of Neighbors with Hate Crimes}_t$. Suppose we adjust for two variables: the unemployment rate at time t and whether a county belongs to East Germany, i.e., $\mathbf{C} \equiv \{\texttt{Unemployment}_t,\ \texttt{East}\}$. Then, a corresponding placebo set adds two variables to the original control variables: the unemployment rate at time $t-1$ (a lag of the existing control variable) and the proportion of neighbors that had hate crimes at time $t-1$ (a lag of the treatment variable), i.e., $\mathbf{C}^P \equiv \{\texttt{Unemployment}_t,\ \texttt{Unemployment}_{t-1},\ \texttt{East},$ $\texttt{Prop of Neighbors with Hate Crimes}_{t-1}\}$. This placebo set is for the case where the unemployment rate at time t is not affected by the incidence of hate crimes at time t. Otherwise, we need to remove the unemployment rate at time t and thus a placebo set should instead be $\mathbf{C}^P \equiv \{\texttt{Unemployment}_{t-1},\ \texttt{East},\ \texttt{Prop of Neighbors with Hate Crimes}_{t-1}\}$.

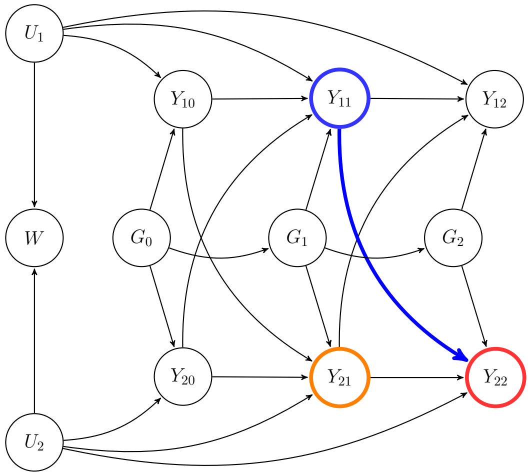
Testing Conditional Independence Although there are many ways to implement the second step of the placebo test, this paper proposes a parametric test based on the spatial autoregressive (SAR) model (e.g., Anselin, 2013; Beck *et al.*, 2006; Cressie, 2015; Franzese and Hays, 2007). For example, when outcomes are continuous, we can implement the placebo test by the following linear spatial autoregressive model.

$$Y_{it} \quad = \quad \alpha_0 + \delta \mathbf{W}_i^\top \mathbf{Y}_t + \gamma_0^\top \mathbf{C}^P + \epsilon_{it}, \tag{6}$$

where $\mathbf{W}_i^\top \mathbf{Y}_t \equiv D_{it}$ is the treatment variable, $\mathbf{C}^P$ is a placebo set, and ϵ_{it} is an error term.

The placebo outcome Y_{it} is conditionally independent of the treatment variable when the assumption of no omitted confounders (Assumption 2) holds. Therefore, the spatial autoregressive coefficient δ serves as a test statistic of the placebo test. If the assumption of no omitted confounders holds, this coefficient is zero, and its p-value should follow a uniform distribution. By testing whether this spatial autoregressive coefficient is zero, researchers can assess the no omitted confounders assumption and thus detect biases, which include contextual confounding and homophily bias. When outcomes take different forms (e.g., binary, counts), we can use corresponding spatial autoregressive models, such as probit or poisson versions.

It is important to note that if the parametric assumptions of the model are violated, the spatial autoregressive coefficient in Equation (6) can be zero even when omitted variable bias remains. As any other statistical tests, a specific parametric placebo test can fail if its underlying parametric assumptions do not hold. A key advantage of the proposed approach is that the equivalence theorem (Theorem 1) is nonparametric. The theorem implies that when there exist no omitted confounders, the placebo outcome and the treatment variable are conditionally independent in any parametric and nonparametric tests. Therefore, in practice, researchers can verify the conditional independence of the placebo outcome and the treatment variable using additional nonparametric tests (e.g., Zhang *et al.*, 2012) or several other parametric tests as described in Section 5.1.



(a) A Stationary Causal DAG

	C	C^P	Placebo Test
No Bias	Y_{21}, U_2, G_2	$Y_{20}, Y_{10}, U_2, G_2, G_1$	Accept
Contextual Confounding	Y_{21}, U_2	Y_{20}, Y_{10}, U_2	Reject
Homophily Bias	Y_{21}, G_2, G_1	$Y_{20}, Y_{10}, G_2, G_1, G_0$	Reject
Both	Y_{21}, Y_{20}	Y_{20}, Y_{10}	Reject

(b) Control and Placebo Sets

Figure 2: Placebo Tests with A Stationary Causal DAG. Note: The causal DAG has twelve nodes in total; six nodes Y_{it} representing outcome variables for two individuals $i \in \{1, 2\}$ over three time periods $t \in \{0, 1, 2\}$, three nodes G_t representing contextual variables for $t \in \{0, 1, 2\}$, two nodes U_i representing individual-level characteristics for $i \in \{1, 2\}$, and finally variable W indicating the connection of two individuals, taking 1 if they are connected and 0 otherwise. I focus on the ACDE of Y_{11} on Y_{22} where Y_{11} is the treatment variable (blue), Y_{22} is the outcome variable (red), and the causal arrow of interest $Y_{11} \rightarrow Y_{22}$ is colored blue. The placebo outcome Y_{21} is colored orange.

This use of the SAR model as a placebo test differs from existing approaches in the spatial econometrics literature that are designed to capture spatial correlations. While researchers conventionally estimate and interpret the spatial autoregressive coefficient as the strength of spatial correlation, the proposed placebo test uses the spatial autoregressive coefficient to detect biases rather than to estimate diffusion effects. When the assumption of no omitted confounders holds, this spatial autoregressive coefficient in Equation (6) should be zero. For the estimation of the ACDE, causal diffusion analysis estimates the conditional expectation $\widehat{\mathbb{E}}[Y_{i,t+1} \mid D_{it}, \mathbf{C}]$ and then uses the identification formula in Result 1.

Illustration with Causal DAG

Here, I use a stationary causal DAG to illustrate how the proposed placebo test works in a simple case. Although the proposed test is applicable to a wide range of empirical settings, I introduce a specific causal DAG (Figure 2 (a)) as one concrete example. This causal DAG combines two causal DAGs in Figure 1 (b) and (c), and thus it can represent both contextual confounding and homophily bias. The causal DAG has twelve nodes in total and definitions of each variable are the same as Section 2.4 (also described in the Note of Figure 2). My main focus is on the ACDE of Y_{11} on Y_{22} where Y_{11} is the treatment variable (blue), Y_{22} is the outcome variable (red), and the causal arrow of interest $Y_{11} \to Y_{22}$ is colored blue. The placebo outcome Y_{21} is colored orange.

Based on this causal DAG in Figure 2 (a), Table in Figure 2 (b) shows four different scenarios: no bias, contextual confounding, homophily bias, and both types of biases. These scenarios show how the proposed placebo test detects biases by exploiting the stationarity of the causal DAG. For each set of control variables, the placebo test checks conditional independence, $Y_{11} \perp\!\!\!\perp Y_{21} \mid \mathbf{C}^P$ where we derive a placebo set $\mathbf{C}^P$ from a chosen control set $\mathbf{C}$ using Equation (5).

First, when we control for three variables $\{Y_{21}, U_2, G_2\}$, the ACDE of interest is identified ("No Bias"). Without knowledge of the entire causal DAG, we can assess the absence of bias by implementing the placebo test. Following Equation (5), we can derive a placebo set $\mathbf{C}^P = \{Y_{20}, Y_{10}, U_2, G_2, G_1\}$ and then the placebo test checks $Y_{11} \perp\!\!\!\perp Y_{21} | \mathbf{C}^P$. In Figure 2 (a), we can verify that there is no unblocked back-door path between Y_{11} and Y_{21}. In this first scenario, the conditional independence holds, and thus the selected control set correctly satisfies the null hypothesis of the placebo test.

Second, I consider a typical form of contextual confounding. When we control for two variables $\{Y_{21}, U_2\}$, the ACDE of interest is not identified due to a back-door path ($Y_{11} \leftarrow G_1 \to G_2 \to Y_{22}$). We now verify that the placebo test correctly detects this bias. In the first step, we derive a placebo set as $\mathbf{C}^P = \{Y_{20}, Y_{10}, U_2\}$. Then, we assess whether there is any unblocked back-door path between Y_{11} and Y_{21}. In fact, we can correctly reject the placebo test; $Y_{11} \not\perp\!\!\!\perp Y_{21} | \mathbf{C}^P$ due to a back-door path ($Y_{11} \leftarrow G_1 \to Y_{22}$). The placebo test shows that the selected control set fails to adjust for all confounders.

Third, I investigate homophily bias. When we control for three variables $\{Y_{21}, G_2, G_1\}$, the ACDE of interest is not identified due to a back-door path ($Y_{11} \leftarrow U_1 \to \boxed{W} \leftarrow U_2 \to Y_{22}$). Recall that W is always conditioned in causal diffusion analysis (indicated by the square box around W; see Section 2.4). For this case, a placebo set is $\mathbf{C}^P = \{Y_{20}, Y_{10}, G_2, G_1, G_0\}$ and we can verify that $Y_{11} \not\perp\!\!\!\perp Y_{21} | \mathbf{C}^P$ due to a back-door path ($Y_{11} \leftarrow U_1 \to \boxed{W} \leftarrow U_2 \to Y_{21}$). The placebo test correctly detects homophily bias. Finally, if we follow the same logic, it is straightforward to verify that the proposed placebo test can also detect biases even when both contextual confounding and homophily bias coexist.

3.3 Extensions

In this subsection, I introduce two extensions of the placebo test by relying on the equivalence theorem (Theorem 1). First, I show that the proposed placebo test can be viewed as a joint test of the sequential consistency assumption (Assumption 1) and the no omitted confounders assumption (Assumption 2). Second, by reverse-engineering the placebo test, I develop a data-assisted covariate selection algorithm to choose a valid set of control variables from pre-treatment covariates.

3.3.1 A Placebo Test as A Joint Test

I have mostly focused on how the proposed placebo test assesses the assumption of no omitted confounders given the sequential consistency assumption (Assumption 1). However, the proposed placebo test is helpful even for assessing the sequential consistency assumption. The proposed placebo test jointly assesses the two identification assumptions: both the sequential consistency assumption as well as the assumption of no omitted confounders.

Lemma 1 (Equivalence between Identification Assumptions and Conditional Independence of Simultaneous Outcomes) For every causal model faithful to stationary causal DAGs,

$$\begin{cases} \text{Sequential Consistency (Assumption 1)} \\ Y_{i,t+1}(d) \perp\!\!\!\perp \{Y_{jt}\}_{j \in \mathcal{N}_i} \mid \mathbf{C} \end{cases} \iff Y_{it} \perp\!\!\!\perp \{Y_{jt}\}_{j \in \mathcal{N}_i} \mid \mathbf{C}^P. \tag{7}$$

This lemma is powerful. It shows that researchers can assess not only the assumption of no omitted confounders (Assumption 2) but also the sequential consistency assumption (Assumption 1) together. That is, researchers can jointly detect simultaneity bias and omitted variable bias. When the conditional independence of simultaneous outcomes holds, it provides strong statistical evidence for both identification assumptions, i.e., the absence of simultaneity bias and omitted variable bias. In contrast, when we reject the null hypothesis of the placebo test, we cannot tell which assumption is violated. When the sequential consistency assumption is violated, the problem is more severe than omitted variable bias – causal diffusion effects are not well defined. Therefore, the first step of causal diffusion analysis, even before considering omitted variable bias, is to measure outcomes frequently enough to satisfy the sequential consistency assumption and have well-defined causal estimands.

Proof of this lemma is essentially the same as the one for Theorem 1. The additional idea is that when the sequential consistency assumption is violated, there is no set of variables that can make simultaneous outcomes conditionally independent – the null hypothesis of the placebo test is always rejected.

3.3.2 Data-Assisted Covariate Selection

By further exploiting the equivalence theorem (Theorem 1), I show that we can directly select a valid set of control variables from pre-treatment covariates, if any, under additional parametric assumptions. A proposed data-assisted covariate selection algorithm is based on the following idea. The equivalence theorem implies that, if we can find a set of covariates that makes simultaneous outcomes conditionally independent, we can reverse-engineer a valid control set using Equation (5). An important step is to estimate conditional independence relationships among observed covariates. I show how to apply Markov random fields to do so and then select a valid control set.[1]

Markov Random Fields: Review The equivalence theorem demonstrates that the assumption of no omitted confounders not only implies but is also implied by the conditional independence of observed simultaneous outcomes. As long as a placebo set $\mathbf{C}^P$ satisfies the conditional independence condition (the right-hand side of Equation (4)), its corresponding control set $\mathbf{C}$ should satisfy the assumption of no omitted confounders.

To find such a placebo set, Markov random fields, also known as undirected graphical models, can serve as the main basis. A Markov random field is a statistical model designed to encode conditional independence relationships over multiple variables. Formally, a Markov random field is specified by an undirected graph $\mathcal{G} = (V, Ed)$ with vertex set $V = \{1, \ldots, p\}$ and edge set $Ed \subset V \times V$. Each vertex represents a random variable, and an edge exists between two vertices A and B if and only if the two random variables are dependent conditional on all remaining variables. This property is known as the pairwise Markov property (Lauritzen, 1996). For graphs with positive distributions, this pairwise Markov property is equivalent to the global Markov property: if every path between two vertices A and B intersects a vertex in set $\mathbf{S}$, two random variables are independent conditional on $\mathbf{S}$, i.e., $A \perp\!\!\!\perp B \mid \mathbf{S}$. Set $\mathbf{S}$ is said to *separate* two random variables A and B in a Markov graph. According to this global Markov property, the selection of the valid placebo set can be recast as the problem of finding a set of covariates separating the placebo outcome and the treatment variable in a given Markov graph. Egami and Hartman (2018) use a similar idea in a different context of generalizing experimental estimates. Introductions to Markov random fields can be found in Lauritzen (1996) and Murphy (2012).

To select a control set from the data, I first need to estimate a Markov random field over the placebo outcome, the treatment variable, and all potential confounders. To respect both continuous and categorical variables, I rely on a pairwise mixed Markov random field (Yang *et al.*, 2015). In particular, I assume that each covariate Z_ℓ is drawn from the following

[1] Markov random fields are used in this paper to simply capture associations between outcomes and potential confounders. It is important to note that it is not used to directly estimate the underlying causal DAG.

exponential family distribution conditional on the remaining variable $\mathbf{Z}_{V\setminus\ell}$.

$$\Pr(Z_\ell \mid \mathbf{Z}_{V\setminus\ell}) = \exp\left\{\alpha_\ell Z_\ell + \sum_{m\neq\ell} \eta_{\ell,m} Z_\ell Z_m + \varphi(Z_\ell) - \Phi(\mathbf{Z}_{V\setminus\ell})\right\}, \tag{8}$$

where $\varphi(Z_\ell)$ is a base measure given by the chosen exponential family, and $\Phi(\mathbf{Z}_{V\setminus\ell})$ is the normalization constant. For example, for the normal distribution, the conditional distribution can be seen as a linear regression model.

$$Z_\ell = \alpha_\ell + \sum_{m\neq\ell} \eta_{\ell,m} Z_m + \epsilon_\ell, \tag{9}$$

where ϵ_ℓ is drawn from the normal distribution with mean 0. In general, each covariate is assumed to follow a generalized linear model conditional on the remaining variables, and hence $\mathbb{E}[Z_\ell \mid \mathbf{Z}_{V\setminus\ell}] = \texttt{link}(\alpha_\ell + \sum_{m\neq\ell} \eta_{\ell,m} Z_m)$ where `link` depends on types of outcomes.

Given this setup, the problem of graph estimation can be reduced to the estimation of parameters $\{\eta_{\ell,m}\}_{m\neq\ell}$; $\eta_{\ell,m} \neq 0$ for variable Z_m in the neighbors of variable Z_ℓ and $\eta_{\ell,m} = 0$ for all other variables. Following Meinshausen and Bühlmann (2006), each generalized linear model is estimated with ℓ_1 (lasso) penalty to encourage sparsity. Using the AND rule, $\widehat{Ed}_{\ell,m} = 1$ when $\eta_{\ell,m} \neq 0$ `and` $\eta_{m,\ell} \neq 0$. The same consistency result holds even when researchers use an alternative OR rule ($\widehat{Ed}_{\ell,m} = 1$ when $\eta_{\ell,m} \neq 0$ `or` $\eta_{m,\ell} \neq 0$).

Selecting A Valid Control Set Given an estimated Markov graph, we can find a desirable placebo set by using graph separation rules. Due to the global Markov property, neighbors of the placebo outcome in the estimated graph is guaranteed to make the placebo outcome and the treatment conditionally independent, as long as there is no edge between the placebo outcome and the treatment. It is important to note that even when this proposed algorithm cannot find any set that makes the placebo outcome and the treatment conditionally independent, it does not imply the absence of valid sets. In other words, while the discovered set is valid asymptotically, the failure of the algorithm does not imply that no sets of observed covariates satisfy the desirable conditional independence. This is because Markov random fields encode only a subset of conditional independence relationships.

Finally, we can reverse-engineer a valid control set from the estimated placebo set. In a simple case, an estimated control set comprises all time-independent variables and forward-lags of all time-dependent variables in the estimated placebo set. After estimating a valid set of control variables using the estimated Markov random field, it is important in practice to run the proposed placebo test and ensure that the selected control set actually satisfies conditions implied by the placebo test. Because Markov random fields are used only as an intermediate step to achieve this conditional independence, this additional placebo test is essential in verifying that the proposed algorithm selects a valid control set successfully.

4 A Bias-Corrected Estimator

In the previous section, I show how the proposed placebo test can detect a wide class of biases under stationary causal DAGs. In practice, if the placebo test detects bias, one may want to collect more data and improve the selection of control variables. This strategy might, however, be infeasible in many applied settings. To help researchers in such common situations, this section considers how to correct biases by introducing an additional parametric assumption. In particular, I propose a bias-corrected estimator – it subtracts the bias detected by the placebo test from a biased estimator. I show that this estimator can remove biases under an assumption that the effect and imbalance of unobserved confounders are stable over time. I also demonstrate that this proposed estimator is closely connected to the widely-used difference-in-difference estimator (Card and Krueger, 1994; Angrist and Pischke, 2008). Section 4.1 describes the bias-corrected estimator with a simple example of linear models, and Section 4.2 introduces a general theory of bias correction.

4.1 An Example with Linear Models

To develop some intuition for a theory of a bias-corrected estimator, I first consider a simple example with linear models. I assume here that a selected set of control variables is time-independent and the same as its corresponding placebo set. A general result is provided in the next subsection.

Suppose we fit a linear model in which we regress the outcome at time $t+1$ on the treatment variable and the selected control set.

$$Y_{i,t+1} \quad = \quad \alpha + \beta D_{it} + \gamma^\top \mathbf{C} + \tilde{\epsilon}_{i,t+1}, \tag{10}$$

where D_{it} is the treatment variable, $\mathbf{C}$ is the selected control set, and $\tilde{\epsilon}_{i,t+1}$ is an error term. As presented in Result 1, if the assumption of no omitted confounders (Assumption 2) holds, $\hat{\beta} \times (d^H - d^L)$ is an unbiased estimator of the ACDE given that the linear model specification is correct. In contrast, when the no omitted confounders assumption is violated, this estimator is biased. We would like to assess whether the assumption of no omitted confounders holds and also correct biases if any.

To assess the assumption of no omitted confounders, suppose we run a parametric placebo test using the following linear spatial autoregressive model as in Equation (6).

$$Y_{it} \quad = \quad \alpha_0 + \delta D_{it} + \gamma_0^\top \mathbf{C}^P + \epsilon_{it},$$

where $\mathbf{C}^P$ is a placebo set and ϵ_{it} is an error term. If the assumption of no omitted confounders holds, the spatial autoregressive coefficient δ should be zero (Theorem 1). In contrast, if the assumption of no omitted confounders does not hold, an estimated coefficient $\hat{\delta}$ then serves as a bias-correction term.

In this simple example, a proposed bias-corrected estimator is given by subtracting the bias-correction term $\hat{\delta}$ from an original biased estimator $\hat{\beta}$.

$$\hat{\tau}_{BC}(d^H, d^L) \equiv (\hat{\beta} - \hat{\delta}) \times (d^H - d^L). \tag{11}$$

This bias-corrected estimator is unbiased for the ACDE for the treated (Theorem 2 in the next subsection). Note that when the assumption of no omitted confounders holds, the expected value of $\hat{\delta}$ is zero, meaning no bias correction.

To understand an assumption necessary for this bias-corrected estimator, I rely on the causal DAG in Figure 1 (b). As one concrete example, consider the study of hate crime diffusion and suppose we fail to observe the unemployment rate. Variable G in the causal DAG can represent this unemployment rate.

To correct bias due to this omitted unemployment rate, we need two parametric assumptions. First, the effect of the unemployment rate on the incidence of hate crimes is the same at time t and $t+1$. In the causal DAG, this assumption requires that the effect of G_2 on Y_{22} be the same as the effect of G_1 on Y_{21}. Although this assumption is stronger than the time-invariant causal structure required for stationary DAGs (Definition 2), social scientists often assume this type of time-invariant effects, especially when time intervals are short.

Second, we need to assume that association between the incidence of hate crimes in neighborhoods (treatment) and the unemployment rate at time t is the same as the one between the treatment and the unemployment rate at time $t+1$. In the causal DAG, this assumption implies that the association between G_2 and Y_{11} is the same as the one between G_1 and Y_{11}. This assumption substantively means the stability of omitted confounder G. When the unemployment rate is stable over two time periods, $G_1 = G_2$, this second assumption holds. When we measure the incidence of hate crimes every month rather than every year, these necessary assumptions might be more tenable.

4.2 Identification with A Bias-Corrected Estimator

In this subsection, I provide a general theoretical result underlying the proposed bias-corrected estimator. I emphasize the connection between the proposed bias-corrected estimator and the difference-in-difference estimator (Card and Krueger, 1994).

I begin by defining the average causal diffusion effect for the treated (ACDT). I will show in Theorem 2 that the proposed bias-corrected estimator is unbiased for the ACDT. The formal definition is given as follows.

$$\tau_{t+1}^{d^H}(d^H, d^L) \equiv \mathbb{E}[Y_{i,t+1}(d^H) - Y_{i,t+1}(d^L) \mid D_{it} = d^H]. \tag{12}$$

This is the average causal diffusion effect for units who received the higher level of the treatment. This quantity could represent the causal diffusion effect of hate crimes for counties in a higher risk neighborhood, i.e., d^H% of neighboring counties had hate crimes in month t.

Without loss of generality, I divide a control set into three types of variables,

$$\mathbf{C} \ \equiv \ \{\mathbf{X}_{i,t+1}, \mathbf{V}_{i,t+1}, \mathbf{Z}_i\},$$

where (1) $\mathbf{X}_{i,t+1}$, the time-dependent variables that are descendants of Y_{it}, (2) $\mathbf{V}_{i,t+1}$, the time-dependent variables that are not descendants of Y_{it}, and (3) $\mathbf{Z}_i$, the time-independent variables. Then, I can write a corresponding placebo set as

$$\mathbf{C}^P \ \equiv \ \{\mathbf{X}_{it}, \mathbf{V}_{i,t+1}, \mathbf{V}_{it}, \mathbf{Z}_i, \{Y_{j,t-1}\}_{j \in \mathcal{N}_i}\}.$$

Using this notation, I introduce a general bias-corrected estimator. It subtracts bias detected by the proposed placebo test from an estimator that we would use under the no omitted confounders assumption.

Definition 3 (A Bias-Corrected Estimator)
A bias-corrected estimator $\hat{\tau}_{\text{BC}}$ is defined as the difference between two estimators $\hat{\tau}_{\text{Main}}$ and $\hat{\delta}_{\text{Placebo}}$.

$$\hat{\tau}_{\text{BC}} \equiv \hat{\tau}_{\text{Main}} - \hat{\delta}_{\text{Placebo}} \tag{13}$$

where

$$\hat{\tau}_{\text{Main}} \equiv \int \Big\{\widehat{\mathbb{E}}[Y_{i,t+1} \mid D_{it} = d^H, \mathbf{X}_{i,t+1}, \mathbf{C}^B] - \widehat{\mathbb{E}}[Y_{i,t+1} \mid D_{it} = d^L, \mathbf{X}_{i,t+1}, \mathbf{C}^B]\Big\} dF_{\mathbf{X}_{i,t+1}, \mathbf{C}^B | D_{it} = d^H}(\mathbf{x}, \mathbf{c}),$$
$$\hat{\delta}_{\text{Placebo}} \equiv \int \Big\{\widehat{\mathbb{E}}[Y_{it} \mid D_{it} = d^H, \mathbf{X}_{it}, \mathbf{C}^B] - \widehat{\mathbb{E}}[Y_{it} \mid D_{it} = d^L, \mathbf{X}_{it}, \mathbf{C}^B]\Big\} dF_{\mathbf{X}_{i,t+1}, \mathbf{C}^B | D_{it} = d^H}(\mathbf{x}, \mathbf{c}),$$

with $\mathbf{C}^B \ \equiv \ \{\mathbf{V}_{i,t+1}, \mathbf{V}_{it}, \mathbf{Z}_i, \{Y_{j,t-1}\}_{j \in \mathcal{N}_i}\}$. $\widehat{\mathbb{E}}[\cdot]$ is any unbiased estimator of $\mathbb{E}[\cdot]$, and researchers can use regression, weighting, matching or other techniques to obtain such an unbiased estimator. Note that both estimators are marginalized over the same conditional distribution $F_{\mathbf{X}_{i,t+1}, \mathbf{C}^B | D_{it} = d^H}(\mathbf{x}, \mathbf{c})$.

This bias-corrected estimator consists of two parts, $\hat{\tau}_{\text{Main}}$ and $\hat{\delta}_{\text{Placebo}}$. The first part is an estimator unbiased for the ACDT under the no omitted confounders assumption. However, $\hat{\tau}_{\text{Main}}$ suffers from bias when this identification assumption is violated. The purpose of the second part $\hat{\delta}_{\text{Placebo}}$ is to correct this bias. It is closely connected to the proposed placebo test; when the assumption of no omitted confounders holds, $\mathbb{E}[\hat{\delta}_{\text{Placebo}}] = 0$ and there is no bias correction. When the assumption is instead violated, $\hat{\delta}_{\text{Placebo}}$ serves as an estimator of the bias. I rely on $\widehat{\text{Var}}(\hat{\tau}_{\text{Main}}) + \widehat{\text{Var}}(\hat{\delta}_{\text{Placebo}})$ as a conservative variance estimator of the bias-corrected estimator given that $\hat{\tau}_{\text{Main}}$ and $\hat{\delta}_{\text{Placebo}}$ are often positively correlated. In the next subsection, I investigate under what assumptions $\hat{\delta}_{\text{Placebo}}$ can correct bias for $\hat{\tau}_{\text{Main}}$.

Assumption and Identification

Here, I propose a simple parametric assumption under which the proposed bias-corrected estimator is unbiased for the ACDT. I begin by defining an unobserved confounder U such that the no omitted confounder assumption holds conditional on $U_{i,t+1}$ and the original control set $\mathbf{C}$, i.e., $Y_{i,t+1}(d^L) \perp\!\!\!\perp \{Y_{jt}\}_{j\in\mathcal{N}_i} \mid U_{i,t+1}, \mathbf{C}$. For simpler illustrations, I assume here that this $U_{i,t+1}$ is a descendant of Y_{it} (general results are in Appendix D). Theorem 1 then implies that observed simultaneous outcomes are independent conditional on U_{it} and $\mathbf{C}^P$, i.e., $Y_{it} \perp\!\!\!\perp \{Y_{jt}\}_{j\in\mathcal{N}_i} \mid U_{it}, \mathbf{C}^P$.

With this setup, I can show that the bias correction requires a parametric assumption that the effect and imbalance of unobserved confounders are constant over time. This assumption is an extension of the stationary DAGs (Definition 2): while the stationary DAGs only require that the existence of causal relationships among outcomes and confounders be time-invariant, this additional parametric assumption requires that some of such causal relationships should have the same effect size over time. The formal statement is given below.

Assumption 3 (Time-Invariant Effect and Imbalance of Unobserved Confounder)

1. Time-invariant effect of unobserved confounder U: For all $u_1, u_0, \mathbf{x}$ and $\mathbf{c}$,

$$\begin{aligned} & \mathbb{E}[Y_{i,t+1}(d^L)|U_{i,t+1} = u_1, \mathbf{X}_{i,t+1} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}] - \mathbb{E}[Y_{i,t+1}(d^L)|U_{i,t+1} = u_0, \mathbf{X}_{i,t+1} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}] \\ = \; & \mathbb{E}[Y_{it}(d^L)|U_{it} = u_1, \mathbf{X}_{it} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}] - \mathbb{E}[Y_{it}(d^L)|U_{it} = u_0, \mathbf{X}_{it} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}]. \end{aligned}$$

2. Time-invariant imbalance of unobserved confounder U: For all $u, \mathbf{x}$ and $\mathbf{c}$,

$$\begin{aligned} & \Pr(U_{i,t+1} \leq u \mid D_{it} = d^H, \mathbf{X}_{i,t+1} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}) - \Pr(U_{i,t+1} \leq u \mid D_{it} = d^L, \mathbf{X}_{i,t+1} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}) \\ = \; & \Pr(U_{it} \leq u \mid D_{it} = d^H, \mathbf{X}_{it} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}) - \Pr(U_{it} \leq u \mid D_{it} = d^L, \mathbf{X}_{it} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}). \end{aligned}$$

Assumption 3.1 requires that the effect of unobserved confounders on the potential outcomes be stable over time. This assumption is more plausible when we can control for a variety of observed time-varying confounders $\mathbf{X}_{i,t+1}$ and $\mathbf{X}_{it}$. However, this assumption might be violated when the change in the effect of U is quick and cannot be explained by observed covariates $\mathbf{X}$. Assumption 3.2 requires that the imbalance of unobserved confounders be stable over time. In other words, the strength of association between the treatment variable and unobserved confounders is the same at time t and $t+1$. Importantly, it does not require that the distribution of confounders is the same across different treatment groups. Instead, it requires that the difference between treatment groups be stable over time.

In practice, both assumptions are more likely to hold when the interval between time t and $t+1$ is shorter because $U_{i,t+1} \approx U_{it}$ and $\mathbf{X}_{i,t+1} \approx \mathbf{X}_{it}$. In particular, when all confounders are time-invariant between time t and $t+1$, Assumption 3.2 holds exactly. Even when confounders are time-varying, researchers can make these assumptions more plausible by adjusting for observed time-varying confounders $\mathbf{X}_{i,t+1}$ and $\mathbf{X}_{it}$.

In a special case where there is no descendant of Y_{it} in the control set, i.e., $\mathbf{X}_{i,t+1} = \mathbf{X}_{it} = \emptyset$, Assumption 3 is equivalent to the parallel trend assumption and the proposed bias-corrected estimator in Equation (13) is equal to the difference-in-difference estimator (Card and Krueger, 1994; Angrist and Pischke, 2008). By allowing for time-varying confounders, Assumption 3 extends the parallel trend assumption. It is also closely connected to the change-in-change method (Athey and Imbens, 2006; Sofer *et al.*, 2016). Specifically, the assumption of the time-invariant imbalance (Assumption 3.2 in this paper) is a simple extension of Assumption 3.3 in Athey and Imbens (2006).

The theorem below shows that under Assumption 3, the bias-corrected estimator is unbiased for the ACDT.

Theorem 2 (Identification with A Bias-Corrected Estimator) Under Assumptions 1 and 3, a bias-corrected estimator in Equation (13) is unbiased for the ACDT.

$$\mathbb{E}[\hat{\tau}_{\mathrm{BC}}] \;\; = \;\; \tau_{t+1}^{d^H}(d^H, d^L).$$

Proof is in Appendix D. It is also true that this estimator is unbiased for the ACDT when the no omitted confounders assumption holds.

5 Empirical Analysis

In this section, using the proposed methods, I analyze two empirical questions that have served as running examples throughout the paper. Although the two applications differ in important ways – the hate crime study is about the spatial diffusion and the human rights norms study is about the network diffusion, I show that researchers can apply the proposed placebo test and the bias-corrected estimator in the same way to both problems.

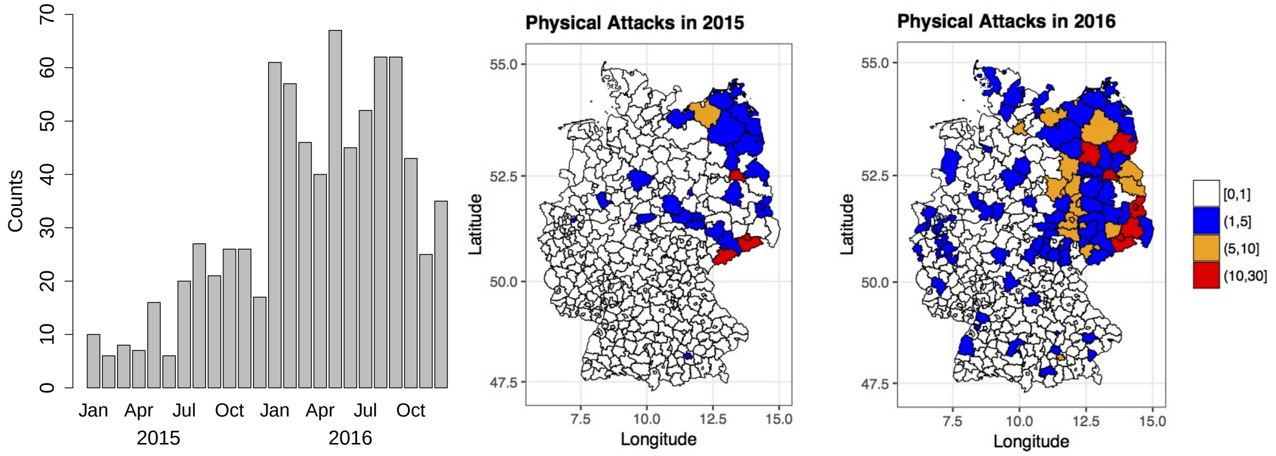
5.1 Spatial Diffusion of Hate Crimes against Refugees

Research across the social sciences has shown that many types of violence are contagious (Wilson and Kelling, 1982; Skogan, 1990; Myers, 2000). In political science, the spatial diffusion of conflicts has received great attention in particular (Hill and Rothchild, 1986; Lake and Rothchild, 1998; Buhaug and Gleditsch, 2008). The central argument in these studies is that one small act of violence can trigger another act of violence, which again induces another, and can lead to waves of violence. Without taking into account how violent behaviors spread across space, it is difficult to explain when, where and why some areas experience violence. Hate crime is not an exception (e.g., Koopmans and Olzak, 2004). For example, Braun (2011) finds that racist violence spread across space in the Netherlands. He argues that the diffusion dynamics might turn "violence from local deviance into a supra-local phenomenon" and "seemingly tolerant regions can suddenly turn into xenophobic hotbeds" (Braun, 2011, p. 753).

In this paper, I investigate the spatial diffusion of hate crimes against refugees, using a data set from Germany documented in Benček and Strasheim (2016) and expanded by Dancygier, Egami, Jamal, and Rischke (2018). Over the last few years, Germany has experienced a record influx of refugees, and at the same time, the number of hate crimes against refugees has increased dramatically. Figure 3 (a) reports the total number of physical attacks against refugees in each month, from the beginning of 2015 to the end of 2016. While there were about 15 hate crimes on average in each month of 2015, this number rose to more than 40 in 2016, a close to 200% increase. Where do we see this massive increase and why? Figure 3 (b) presents the spatial patterns of physical attacks over the two years.

Two empirical patterns are worth noting. First, hate crimes are spatially clustered in East Germany. Second, the number of counties that experience hate crimes grows over time. This dynamic spatial pattern is consistent with the spatial diffusion theory which argues that hate crimes diffuse from one county to another spatially proximate county over time (Myers, 2000; Braun, 2011). Indeed, Jäckle and König (2016) recently found that the incidence of hate crimes in one county predicts that of hate crimes in its spatially proximate counties using the data from Germany in 2015.

However, it is challenging to estimate the causal impact of this spatial diffusion process because there exist potential concerns of contextual confounding: many unobserved confounders



(a) Time Trend

(b) Spatial Patterns

Figure 3: Temporal and Spatial Patterns of Hate Crimes in Germany. Note: The left figure shows the number of physical attacks in each month from the beginning of 2015 to the end of 2016. In the middle and right figures, I show the number of physical attacks in each county in 2015 and 2016, respectively. Each of 402 counties is colored in white, blue, orange, or red if the number of hate crimes in a given year is less than or equal to 1, 5, 10, or greater than 10, respectively.

can be spatially correlated. For example, the number of refugees in each county increased substantially during this period and is also spatially correlated. Even if researchers collect a long list of covariates, it is difficult to assess whether a selected set of control variables is sufficient for removing contextual confounding. To address this concern over potential bias, I rely on the proposed methods: the placebo test and the bias-corrected estimator.

First, I estimate the average causal diffusion effect (ACDE) on the incidence of hate crimes. In contrast to some existing studies (Braun, 2011; Jäckle and König, 2016), I find that the spatial diffusion effect is small when averaging over all counties. By removing contextual confounding that previous studies have suffered from, the analysis in this paper avoids the overestimation of the causal diffusion effect. Then, I extend this analysis by considering types of counties that are more susceptible to the diffusion of hate crimes. This further investigation of heterogeneous causal effects shows that the spatial diffusion effect is large only for counties that have a higher proportion of school dropouts. This finding suggests that the spatial diffusion of hate crimes is concentrated in areas with low educational performance. This is consistent with rich qualitative evidence that hate crime is often a problem of young people (Green *et al.*, 2001).

Setup

Data on hate crimes come from a project, Mut gegen rechte Gewalt (courage against right-wing violence), by the Amadeu Antonio Foundation and the weekly magazine *Stern*, which has been documenting anti-refugee violence in Germany since the beginning of 2014. The data collection is based on a wide range of sources, including newspaper articles, press releases by the German police, and parliamentary interpellations that can retrieve specific categories of the official hate crime data from the Federal Criminal Police Office in Germany (Bundeskriminalamt, BKA). In particular, this project documents four different types of hate crimes against refugees and refugee housing: physical attacks, arson attacks, other attacks on refugee housing, and demonstrations. This data source has been recently analyzed by several papers (e.g., Benček and Strasheim, 2016; Jäckle and König, 2016). The dataset I analyze in this empirical analysis, compiled by Dancygier, Egami, Jamal, and Rischke (2018), extends this data source on hate crimes by merging in other variables, such as the number of refugees, the population size, a proportion of school dropouts and unemployment rates, collected from the Federal Statistical Office in Germany.

As one of the most well-studied outcomes in the literature of ethnic violence, I focus on physical attacks as the main dependent variable. Note that a definition of physical attacks in this data set focuses on attacks against refugees and it does not include attacks on supporters of refugees or other racially motivated attacks. Formally, I define the outcome variable Y_{it} to be binary, taking the value 1 if there exists any physical attack against refugees at county i in month t, and taking the value 0 otherwise. The outcomes are defined for 402 counties in Germany every month from the beginning of 2015 to the end of 2016. Averaging over all counties in Germany during this period, the sample mean of the outcome variable is 6.4%. This means that 6.4% of counties experienced at least one physical attack in a typical month. In Saxony, a state with the largest number of hate crimes, the sample mean of the outcome variable is 34%. As robustness checks, I also investigate ordinal outcomes and weekly data in Appendix E.3. Results are similar to those presented below.

I use a distance matrix to encode the physical proximity between counties. In particular, I construct an initial distance matrix $\widetilde{\mathbf{W}}$ using an inverse of the straight distance between counties i and j as $\widetilde{W}_{ij}$. I then row-standardize the initial matrix $\widetilde{\mathbf{W}}$ and obtain a final distance matrix $\mathbf{W}$. For the outcome variable in month $t+1$, the treatment variable is defined to be $D_{it} \equiv \mathbf{W}_i^\top \mathbf{Y}_t$, the weighted proportion of neighboring counties that experience the incidence of physical attacks in month t. The first causal quantity of interest is the ACDE, which quantifies how much the probability of having hate crimes changes due to the increase in the proportion of neighboring counties that have experienced hate crimes last month.

To illustrate the use of a placebo test and a bias-corrected estimator, I consider five different sets of control variables in order. Analyzing these five sets step by step, I show how researchers can remove biases in each step. As the first set of control variables, I include one-month lagged dependent and treatment variables. I also adjust for basic summary statistics of $\mathbf{W}_i$, i.e., the number of neighbors and variance of $\mathbf{W}_i$, in order to compare observations with similar spatial characteristics. As discussed in Figure 1, these lagged variables and basic summary statistics of the spatial distance are sufficient for the identification if the spatial diffusion is the only mechanism through which neighboring counties exhibit similar outcomes. Then, as the second set of control variables, I add two-month lagged dependent variables to see whether adjusting for a longer history of past outcomes can reduce bias (e.g., Christakis and Fowler, 2013; Eckles and Bakshy, 2017). The third set of control variables add state fixed effects. Although the state fixed effects are often excluded from existing studies of hate crimes in Germany (e.g., Braun and Koopmans, 2009; Jäckle and König, 2016), I will show how much these fixed effects help remove biases. Then, the fourth set adds a list of contextual variables related to the number of refugees, demographics, education, general crimes, economic indicators, and politics. Finally, the fifth set controls for the time trend using third-order polynomials.

I conduct a placebo test by deriving a placebo set for each of the specified five control sets. Following Equation (5), I add lags of time-dependent control variables and the treatment variable and then remove those affected by the placebo outcome. I provide details of the five control sets and the corresponding placebo sets in Appendix E.1.

Estimation of the Average Causal Diffusion Effect

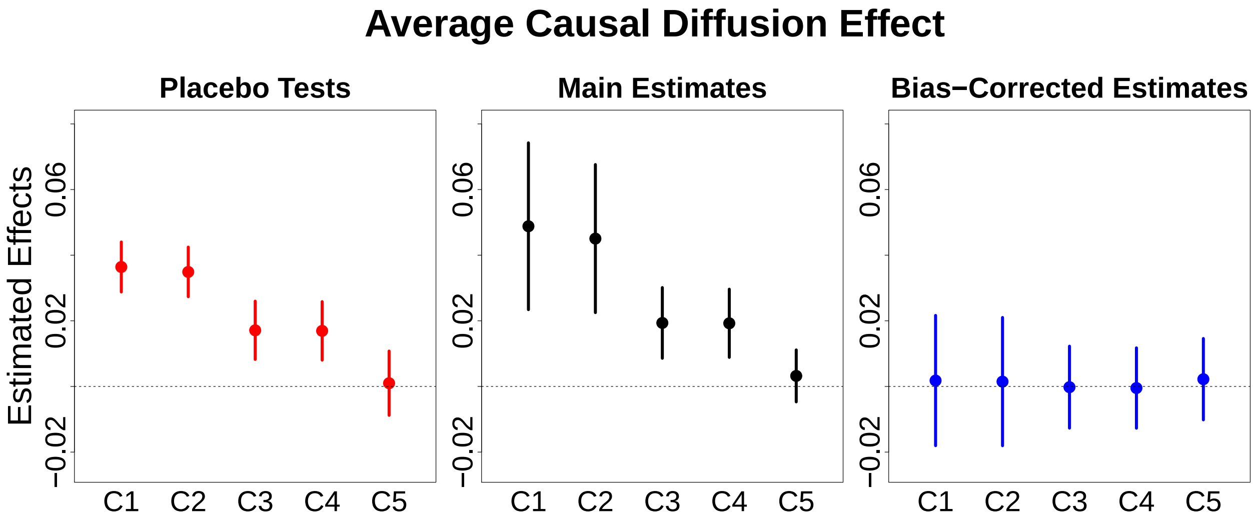
To estimate the ACDE, I rely on a simple parametric model. I use the following logistic regression to model the main outcome variable $Y_{i,t+1}$ with the treatment variable and each of the five control sets.

$$\text{logit}(\Pr(Y_{i,t+1} = 1 \mid D_{it}, \mathbf{C})) \ = \alpha + \beta D_{it} + \gamma^\top \mathbf{C}, \tag{14}$$

where D_{it} is the treatment variable and $\mathbf{C}$ is a specified set of control variables. Under the assumption of no omitted confounders, the difference in the predicted probabilities of $Y_{i,t+1}$ under $D_{it} = d^H$ and $D_{it} = d^L$ serves as an estimator for the ACDE. In particular, I estimate the ACDE that compares the following two treatment values; $d^H = 27\%$, the treatment received by the average counties in Saxony (a state with the largest number of hate crimes) and $d^L = 0\%$, none of the neighbors experiencing hate crimes (common for safe areas in West Germany). Formally, $\hat{\tau} \equiv \int \{\widehat{\Pr}(Y_{i,t+1} = 1 \mid D_{it} = 0.27, \mathbf{C}) - \widehat{\Pr}(Y_{i,t+1} = 1 \mid D_{it} = 0, \mathbf{C})\} dF_{\mathbf{C}}(\mathbf{c})$.

To assess the no omitted confounders assumption, I estimate the following logistic regression.

$$\text{logit}(\Pr(Y_{it} = 1 \mid D_{it}, \mathbf{C}^P)) \ = \alpha_0 + \rho D_{it} + \gamma_0^\top \mathbf{C}^P, \tag{15}$$



(a) (b) (c)

Figure 4: Placebo Tests, Main Estimates, and Bias-Corrected Estimates of the ACDE.
Note: Figures (a), (b) and (c) present results from the placebo tests, estimates of the ACDE under the no omitted confounders assumption, and estimates from bias-corrected estimators with 95% confidence intervals, respectively. C1, C2, C3, C4, and C5 refer to the five different control sets. Figure (a) shows that while the first four sets of control variables are not sufficient, the fifth set successfully adjusts for confounders. Focusing on the fifth control set, which produces a placebo estimate close to zero, a point estimate of the ACDE in Figure (b) is smaller than 1 percentage point and its 95% confidence interval covers zero. Figure (c) shows that bias-corrected estimates are similar regardless of the selection of control variables and they all cover a point estimate from the most credible fifth control set.

where Y_{it} is the placebo outcome and $\mathbf{C}^P$ is a placebo set corresponding to the control set $\mathbf{C}$. When the no omitted confounders assumption holds, Theorem 1 implies that $\rho = 0$. I use the difference in the predicted probabilities of Y_{it} under $D_{it} = d^H$ and $D_{it} = d^L$ as a test statistic of the placebo test. Formally, $\hat{\delta} \equiv \int \{\widehat{\Pr}(Y_{it} = 1 \mid D_{it} = 0.27, \mathbf{C}^P) - \widehat{\Pr}(Y_{it} = 1 \mid D_{it} = 0, \mathbf{C}^P)\} dF_{\mathbf{C}^P}(\mathbf{c}^P)$.

Figures 4 (a) and (b) present results from the placebo tests (Equation (15)) as Placebo Tests and estimates from the main model (Equation (14)) as Main Estimates with 95% confidence intervals, respectively. All standard errors are clustered at the state level. C1, C2, C3, C4, and C5 refer to the five different control sets I introduced before. When a given set of control variables satisfies the no omitted confounders assumption, estimates from the placebo tests $\hat{\delta}$ should be close to zero. Figure 4 (a) shows that while the first four sets of control variables are not sufficient, the fifth set (C5) successfully adjusts for confounders; a point estimate is close to zero and its 95% confidence interval covers zero. It is not enough to control for lagged dependent variables and contextual variables. It turns out to be critical to control for the time trend.

On the basis of these results from the placebo tests, I can now investigate estimates of the ACDE from the main model in Figure 4 (b). For the first two cases (C1 and C2), main estimates are as large as 5 percentage points, but the placebo tests suggest that these estimates are heavily biased. Similarly, while the next two cases show point estimates of around 2 percentage points, they are also likely to be biased. When we focus on the fifth control set, which produces a placebo estimate close to zero, a point estimate of the ACDE is smaller than 1 percentage point, and its 95% confidence interval covers zero. The comparison between this more credible estimate and the one from the fourth set shows that an estimate of the ACDE can suffer from 100% bias by missing one variable. This demonstrates the importance of bias detection in causal diffusion analysis.

Although the proposed placebo tests suggest that the fifth control successfully adjusts for relevant confounders in this example, it is often infeasible to find such control sets in many other applications. To address these common scenarios, I now examine whether researchers could obtain similar results using a bias-corrected estimator even with control sets that reject the null hypothesis of the placebo test.

Figure 4 (c) shows that bias-corrected estimates are similar regardless of the selection of control variables and they all cover the most credible point estimate from the fifth control set. Even though the proposed placebo test detected a large amount of bias, researchers can obtain credible estimates by correcting the biases in this example.

What do these results tell us? In contrast to some existing studies (Braun, 2011; Jäckle and König, 2016), this analysis shows that the ACDE on the incidence of hate crimes is small when averaging over all counties in Germany. In the next subsection, I show that the spatial diffusion of hate crimes is concentrated among a small subset of counties that have a higher proportion of school dropouts.

Heterogeneous Diffusion Effects by Education

Now, I extend the previous analysis by considering types of counties that are more susceptible to the diffusion of hate crimes. In particular, I examine the role of education. Given rich qualitative and quantitative evidence that hate crime is often a problem of young people, it is critical to take into account one of the most important institutional contexts around them, i.e., schooling. There are at least three mechanisms through which education can reduce the risk of hate crimes. First, education increases economic returns to current and future legitimate work, thereby raising the opportunity cost of committing hate crimes (e.g., Lochner and Moretti, 2004). Second, education may change psychological costs associated with hate crimes (e.g., Frindte *et al.*, 1996). More educated people tend to have lower levels of ethnocentrism and place more emphasis on cultural diversity (Hainmueller and Hiscox, 2007). Finally, some scholars emphasize that schooling has incapacitation effects – keeping adolescents busy and off the street, thereby directly reducing the chances of committing crimes (Jacob and Lefgren, 2003).

Building on the literature above, I investigate whether local educational contexts condition the spatial diffusion dynamics of hate crimes. In this paper, I use a proportion of school dropouts without a secondary school diploma as a measure of local educational performance. To better disentangle the education explanation, I analyze East Germany and West Germany separately because they have substantially different distributions of proportions of school dropouts (counties in East Germany have higher proportions of school dropouts; see Appendix E.2 for details). Here I report results from East Germany and provide those for West Germany in Appendix E.2. In particular, I estimate the conditional average causal diffusion effects (conditional ACDEs) for counties that have high and low proportions of school dropouts without a secondary school diploma. I use 9% as a cutoff for high and low proportions of school dropouts, which is approximately the median value in East Germany. I add an interaction term between the treatment variable and this indicator variable to the original model in Equation (14) and to the original placebo model in Equation (15). Although this investigation of heterogeneous effects is observational in nature as in a typical subgroup analysis, results are consistent with the education explanation.

Figure 5 presents results for the conditional ACDE for counties that have a higher proportion of school dropouts. Similar to the case of the ACDE estimation, Figure 5 (a) shows strong concerns of biases in the first four sets of control variables. Even though a 95% confidence interval of the fourth estimate covers zero, its point estimate is far from zero (around 4 percentage points). In contrast, the placebo test suggests that the fifth control set adjusts for relevant confounders where a placebo estimate is close to zero.

Based on results from the placebo tests, I examine estimates from the main model in Figure 5 (b). The first four sets, likely to be biased, exhibit large point estimates, larger than 10 percentage points. More interestingly, even with the most credible fifth control set, a point estimate is as large as 6 percentage points and is statistically significant. This effect size is substantively important given that it is about one-fourth of the sample average outcome in this subset (26%). Bias-corrected estimates in Figure 5 (c) confirm that the conditional ACDE for counties with a higher proportion of school dropouts is large and similar regardless of the selection of control sets.

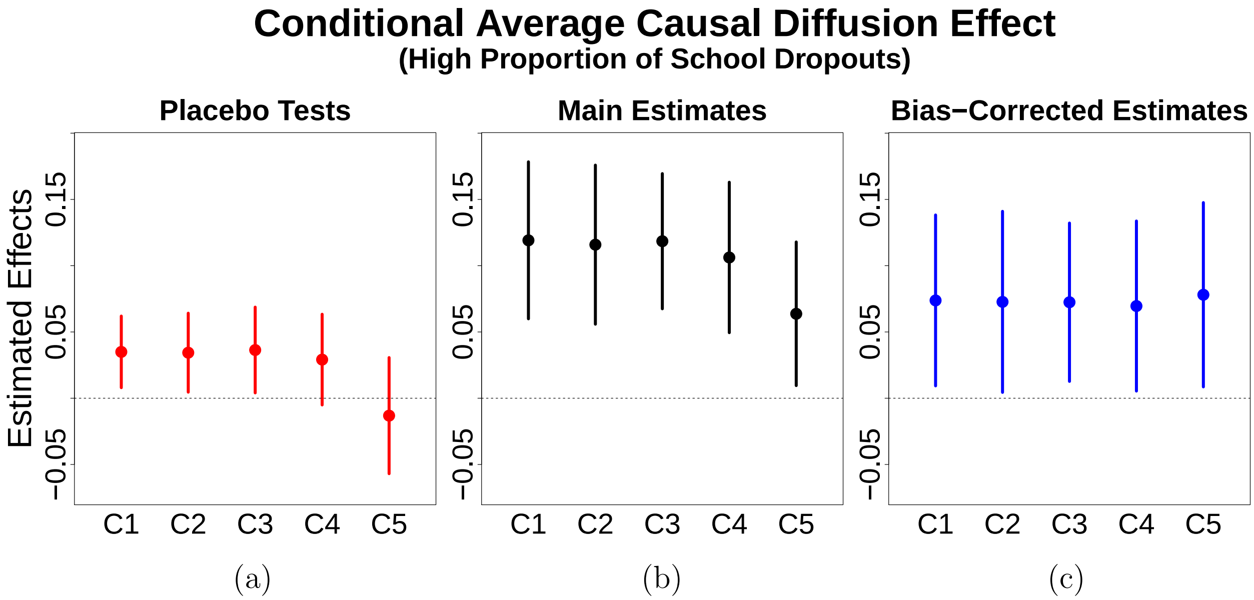


Figure 5: Placebo Tests, Main Estimates, and Bias-Corrected Estimates of the conditional ACDE for counties with a high proportion of school dropouts. Note: Figures (a), (b) and (c) present results from the placebo tests, estimates of the conditional ACDE under the no omitted confounders assumption, and estimates from bias-corrected estimators with 95% confidence intervals, respectively. C1, C2, C3, C4, and C5 refer to the five different control sets. Figure (a) shows that while the first four sets of control variables are not sufficient, the fifth set successfully adjusts for confounders. Focusing on the most credible fifth control set, a point estimate of the conditional ACDE in Figure (b) is as large as 6 percentage points and is statistically significant at the 0.05 level. Figure (c) shows that bias-corrected estimates are similar regardless of the selection of control variables.

When I estimate the conditional ACDE for counties that have a lower proportion of school dropouts, effects are close to zero and their 95% confidence intervals cover zero, as the education hypothesis expects (see Appendix E.2). Causal diffusion effects are also precisely estimated to be zero in West Germany, where proportions of school dropouts are much lower than East Germany. This additional analysis suggests that the spatial diffusion dynamics of hate crimes operate only in places with low educational performance.

5.2 Network Diffusion of Human Rights Norms

In political science, diffusion theories have played a central role in explaining a wide range of phenomena. Topics include regime transitions (Huntington, 1991; Starr, 1991; Pevehouse, 2002; Gleditsch and Ward, 2006), conflicts (Lake and Rothchild, 1998; Buhaug and Gleditsch, 2008; Gleditsch *et al.*, 2008), foreign economic policies (Simmons and Elkins, 2004; Simmons *et al.*, 2006), election monitoring (Hyde, 2011) and capital taxation (Cao, 2010). In particular, scholars of the human rights politics have examined the international diffusion of human rights norms from a variety of aspects (e.g., Keck and Sikkink, 1998).

Recently, by extending an influential work by Johnston (2001), Greenhill (2010, 2016) takes this human rights literature in a new direction by considering the socialization effect of intergovernmental organizations (IGOs). He emphasizes that the IGOs offer social forums in which high-level policymakers interact with each other and transmit human rights norms to fellow IGO members. Consistent with this main theory, he finds that human rights performance, measured by the Personal Integrity Rights (PIR) score, is strongly associated with those of the IGO partners even after controlling for a number of domestic and international-level variables. However, as in a typical network diffusion analysis, it is essential to address potential concerns over homophily bias in order to establish causal claims. Because similar types of states are more likely to join similar types of IGOs, it is possible to observe similar levels of human rights performance among IGO partners even without any network diffusion. Unless researchers take this homophily story into account, they could overestimate the network diffusion effect.

Recognizing this potential problem, Greenhill (2016) adjusts for both spatial proximity and cultural similarity between states in addition to a list of conventional control variables, such as population size and GDP. In this section, I apply the proposed methods and evaluate the confounder selection of the original study. Although including spatial and cultural connections reduces bias, the proposed placebo test reveals that a large amount of confounding remains. When adjusting for time trends and enough lags, an estimate of the ACDE is close to zero, in contrast to the original findings. The proposed bias-corrected estimator also confirms this null finding. By employing the proposed methods, we can avoid the overestimation of causal diffusion effects. I emphasize that this subsection focuses on only one of the main models in the important work (Greenhill, 2016) and hence, it is plausible that human rights norms diffuse in a way, not captured by the model discussed here. One advantage of the proposed approach is that it can also help researchers who might incorporate different network channels or investigate heterogeneous treatment effects.

Setup

In this section, I use the replication data (Greenhill, 2015), which covers the period from 1985 to 2005. Using the Correlates of War 2 International Governmental Organizations Data Set, version 2.3 (Pevehouse *et al.*, 2004), Greenhill (2016) measures an original IGO-based network between states. In particular, Greenhill (2016) views the IGO network as a bipartite network in which there are two types of nodes (states and IGOs) and edges exist only between states and IGOs. This new approach explicitly incorporates his theory that states are interacting with each other through IGO memberships. Based on this bipartite network, Greenhill (2016) defines the strength of a direct tie from state j to state i as follows. First, define $\mathcal{I}_{ijt}$ to be a set of IGOs that both states i and j belong to in year t, and for each IGO k in this set $\mathcal{I}_{ijt}$, compute the total number of members N_{kt}. Then, the strength of the direct tie from state j to state i in year t is defined as

$$W_{ijt} \equiv \frac{\sum_{k \in \mathcal{I}_{ijt}} 1/(N_{kt} - 1)}{\text{Total number of IGOs that state } i \text{ belongs to in year } t}. \tag{16}$$

This measure encodes several important intuitions. First, when two states share more memberships, their tie is stronger because more states are in set $\mathcal{I}_{ijt}$. Second, when two states share memberships in smaller IGOs (therefore, officials from two states potentially have more interactions), their tie is stronger because N_{kt} is smaller. Finally, when state i belongs to many IGOs, interactions in each IGO are less important and hence, the tie is weaker.

Greenhill (2016) argues that human rights performance can diffuse through this IGO-based network. The main outcome, the PIR score, is denoted by Y_{it} for state i in year t. For the outcome variable in year $t + 1$, the treatment variable is defined to be $D_{it} \equiv \mathbf{W}_{it}^{\top}\mathbf{Y}_t$, the weighted average of the PIR scores of state i's fellow IGO members. This quantity is defined as *IGO Context* in the original study. The main causal quantity of interest is the ACDE of the PIR score over one year (see Table 3 in Chapter 3 of Greenhill (2016)). It quantifies how much the level of human rights performance diffuse from IGO partners over time. More precisely, the study asks: how much would the PIR score of a given state in year $t + 1$ change if its fellow IGO members have higher PIR scores in year t?

I consider three different sets of control variables. As the first control set, I follow the original analysis (Table 3 in Chapter 3 of Greenhill (2016)) and includes logged GDP per capita, regime durability, population density, democracy, trade dependence, FDI dependence, conflict, and a lagged dependent variable. As the second control set, I again follow Greenhill (2016) and add two additional network measures, i.e., spatial proximity and cultural similarity between states, to adjust for potential homophily bias. Finally, the third set controls for year fixed effects and longer periods of the diffusion history, i.e., two- and three-year lagged dependent and treatment variables, in addition to the original one-year lagged dependent variable. I also include basic network characteristics, i.e., the number of neighbors and variance of $\mathbf{W}_{it}$. I provide details of the three control sets and the corresponding placebo sets in Appendix F.

Models and Results

To estimate the ACDE, I use the following linear regression to model the main outcome variable $Y_{i,t+1}$ with the treatment variable and each of the three control sets.

$$\mathbb{E}[Y_{i,t+1} \mid D_{it}, \mathbf{C}] \; = \alpha + \beta D_{it} + \gamma^\top \mathbf{C}, \tag{17}$$

where D_{it} is the treatment variable and $\mathbf{C}$ is a specified set of control variables. Under the assumption of no omitted confounders, $\hat{\beta}$ is an unbiased estimator for the ACDE. Then, to assess the no omitted confounders assumption, I estimate the linear regression model.

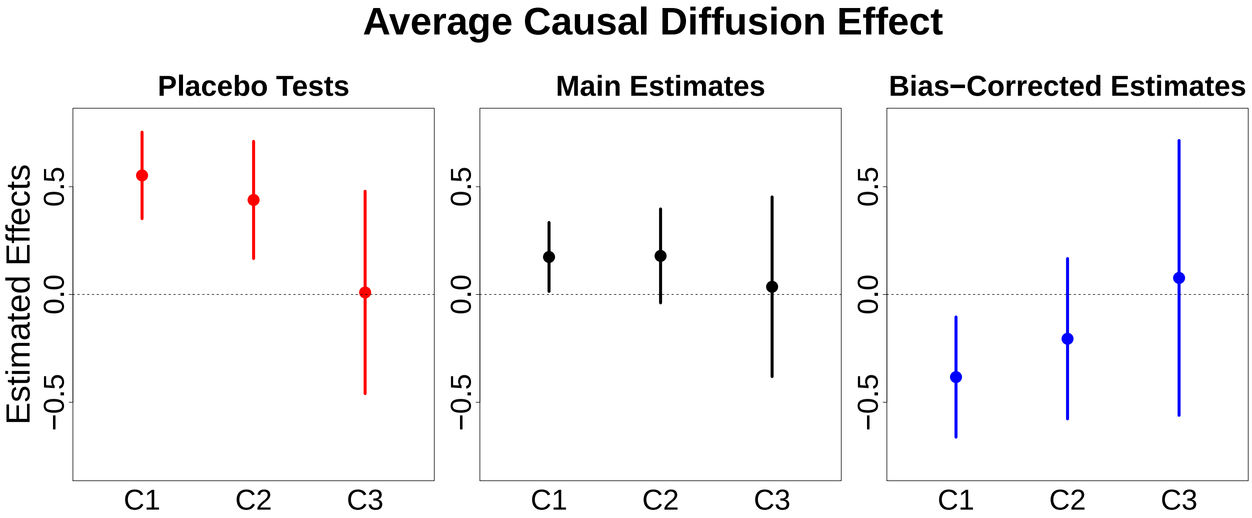
$$\mathbb{E}[Y_{it} \mid D_{it}, \mathbf{C}^P] \; = \alpha_0 + \rho D_{it} + \gamma_0^\top \mathbf{C}^P, \tag{18}$$

where Y_{it} is the placebo outcome and $\mathbf{C}^P$ is a placebo set corresponding to the control set $\mathbf{C}$. When the no omitted confounders assumption holds, Theorem 1 implies that $\rho = 0$. I use $\hat{\rho}$ as a test statistic of the placebo test. I also utilize the proposed bias-corrected estimator by combining two regression models (Equations (17) and (18)). All standard errors are clustered at the country level.

Figures 6 (a) and (b) present results from the placebo tests (Placebo Tests) and estimates under the no omitted confounders assumption (Main Estimates) with 95% confidence intervals, respectively. C1, C2, and C3 refer to the three control sets I explained above. Figure 6 (a) suggests that while the first two sets of control variables are not sufficient for removing confounding, the last third set (C3) adjusts for relevant confounders. Although adjusting for spatial proximity and cultural similarity (C2) slightly reduces bias compared to the first control set (C1), it still suffers from a large amount of bias. As in the first application of the hate crime diffusion, it is critical to adjust for the temporal dynamics (the third set, C3).

According to these results of the placebo tests, estimates from the first two control sets (C1 and C2 in Figure 6 (b)) are likely to be biased. When we focus on the most credible third set, a point estimate of the ACDE of the PIR score is close to zero with its 95% confidence interval covering zero. This suggests that after removing biases of the first two estimates, there is no statistical evidence for the ACDE of the PIR score.

Looking at estimates from the bias-corrected estimators in Figure 6 (c), two points are worth noting. First, 95% confidence intervals for the second and third estimates cover zero, confirming the null finding from main estimates. Second, and more importantly, bias-corrected estimates are not as stable as those reported in the first application of the hate crime diffusion. The bias-corrected point estimate from C1 differs from the one from C3 by about 0.4. This instability is likely because the first control set C1 violates the assumption that the effect



(a) (b) (c)

Figure 6: Placebo Tests, Main Estimates, and Bias-Corrected Estimates for the ACDE on the Personal Integrity Rights (PIR) score. Note: Figures (a), (b) and (c) present results from the placebo tests, estimates of the ACDE under the no omitted confounders assumption, and estimates from bias-corrected estimators with 95% confidence intervals, respectively. C1, C2, and C3 refer to the three different control sets. Figure (a) shows that while the first two sets of control variables are not sufficient, the last third set successfully adjusts for confounders. Focusing on the most credible third control set, a point estimate of the ACDE in Figure (b) is close to zero. Figure (c) shows that the bias-corrected estimate from the first control set differs from the one from the most credible third set. This is likely because the first control set violates the assumption that the effect and imbalance of unobserved confounders are stable over time.

and imbalance of unobserved confounders are stable over time (Assumption 3). This result suggests that scholars need to carefully evaluate the plausibility of Assumption 3 especially when analyzing annual data.

6 Concluding Remarks

Spatial and network diffusion dynamics have been an integral part of many social science theories. Given that spatial and network panel data have become increasingly common, it is essential to develop methodologies to draw causal inference for diffusion effects. To address pervasive concerns over contextual confounding and homophily bias, this paper introduces a new class of stationary causal DAGs. By making use of the time-invariant structure of stationary causal DAGs, I develop two statistical tools to facilitate credible causal diffusion analysis. First, I propose a new statistical test that can detect a wide class of biases, including contextual confounding and homophily bias. Then, I develop a difference-in-difference style estimator that can directly correct biases under an additional parametric assumption. The proposed approach offers a simple way to assess and adjust for omitted variable bias, the central challenge for causal diffusion analysis.

Using the proposed methods, I examined the spatial diffusion of hate crimes in Germany. After removing upward bias in previous studies, I found that the average effect of spatial diffusion is small, in contrast to recent quantitative analyses (Braun, 2011; Jäckle and König, 2016). The investigation of heterogeneous effects, however, revealed that the spatial diffusion effect of hate crimes is large only in areas that have a high proportion of school dropouts, which is consistent with qualitative evidence from Germany (e.g., Hagan *et al.*, 1995). My reanalysis of the norm diffusion study (Greenhill, 2016) detected a large amount of bias and found little evidence of the network diffusion effect after correcting the bias. Both applications demonstrate the large differences in substantive conclusions that can result from contextual confounding and homophily bias. By directly accounting for these biases, the proposed placebo test and bias-corrected estimator help researchers make more credible causal inference for diffusion studies.

There are a number of possible future extensions. First, whereas I propose an extension of the difference-in-difference estimator to causal diffusion analysis, future research should also investigate how to incorporate into causal diffusion analysis other popular tools developed for estimating the average treatment effect in panel data settings, such as synthetic control methods (Abadie and Gardeazabal, 2003; Abadie *et al.*, 2010), latent factor models (Bai, 2009; Stewart, 2014; Xu, 2017), and matrix completion methods (Athey *et al.*, 2017). Second, to further disentangle different channels of diffusion effects, it is essential to incorporate multiple networks into the proposed framework. With this extension, researchers can analyze, for example, micromechanisms of hate crime diffusion by estimating causal diffusion effects through offline face-to-face networks and online social networks. Finally, although this paper focuses on the causal diffusion effect of the first order lag, which is the main focus in many applications, researchers might be interested in the longer term effect of diffusion, for example, the network diffusion effect of international norms over decades. To analyze such long-term diffusion processes, it is of interest to extend the proposed methods to a framework of the marginal structural model (Robins *et al.*, 2000).

References

Abadie, A. and Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review*, **93**(1), 113–132.

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

Alt, J. E., Jensen, A., Larreguy, H. A., Lassen, D. D., and Marshall, J. (2018). Contagious Political Concerns: Identifying Unemployment Shock Information Transmission Using the Danish Population Network. *Working paper*.

An, W. (2015). Instrumental Variables Estimates of Peer Effects in Social Networks. *Social Science Research*, **50**, 382–394.

Anagnostopoulos, A., Kumar, R., and Mahdian, M. (2008). Influence and Correlation in Social Networks. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 7–15. ACM.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ.

Anselin, L. (2013). *Spatial Econometrics: Methods and Models*. Springer Science & Business Media.

Aral, S., Muchnik, L., and Sundararajan, A. (2009). Distinguishing Influence-based Contagion from Homophily-driven Diffusion in Dynamic Networks. *Proceedings of the National Academy of Sciences*, **106**(51), 21544–21549.

Athey, S. and Imbens, G. W. (2006). Identification and Inference in Nonlinear Difference-in-Differences Models. *Econometrica*, **74**(2), 431–497.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2017). Matrix Completion Methods for Causal Panel Data Models. *arXiv preprint arXiv:1710.10251*.

Bai, J. (2009). Panel Data Models with Interactive Fixed Effects. *Econometrica*, **77**(4), 1229–1279.

Beck, N., Gleditsch, K. S., and Beardsley, K. (2006). Space is More Than Geography: Using Spatial Econometrics in the Study of Political Economy. *International Studies Quarterly*, **50**(1), 27–44.

Beissinger, M. R. (2007). Structure and Example in Modular Political Phenomena: The Diffusion of Bulldozer/Rose/Orange/Tulip Revolutions. *Perspectives on Politics*, **5**(02), 259–276.

Benček, D. and Strasheim, J. (2016). Refugees Welcome? A Dataset on Anti-Refugee Violence in Germany. *Research & Politics*, **3**(4).

Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of Peer Effects through Social Networks. *Journal of Econometrics*, **150**(1), 41–55.

Braun, R. (2011). The Diffusion of Racist Violence in the Netherlands: Discourse and Distance. *Journal of Peace Research*, **48**(6), 753–766.

Braun, R. and Koopmans, R. (2009). The Diffusion of Ethnic Violence in Germany: The Role of Social Similarity. *European Sociological Review*, **26**(1), 111–123.

Buhaug, H. and Gleditsch, K. S. (2008). Contagion or Confusion? Why Conflicts Cluster in Space. *International Studies Quarterly*, **52**(2), 215–233.

Cao, X. (2010). Networks as Channels of Policy Diffusion: Explaining Worldwide Changes in Capital Taxation, 1998–2006. *International Studies Quarterly*, **54**(3), 823–854.

Card, D. and Krueger, A. B. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *American Economic Review*, **84**(4), 772–793.

Christakis, N. A. and Fowler, J. H. (2007). The Spread of Obesity in A Large Social Network Over 32 years. *New England Journal of Medicine*, **357**(4), 370–379.

Christakis, N. A. and Fowler, J. H. (2013). Social Contagion Theory: Examining Dynamic Social Networks and Human Behavior. *Statistics in Medicine*, **32**(4), 556–577.

Cinelli, C. and Hazlett, C. (2018). Making Sense of Sensitivity: Extending Omitted Variable Bias. *Working Paper*.

Cressie, N. (2015). *Statistics for Spatial Data*. John Wiley & Sons.

Dancygier, R. M., Egami, N., Jamal, A. A., and Rischke, R. (2018). Hate Crimes in Germany. *Work in Progress*.

Danks, D. and Plis, S. (2013). Learning Causal Structure from Undersampled Time Series. In *NIPS 2013 Workshop on Causality*.

Duflo, E. and Saez, E. (2003). The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment. *The Quarterly Journal of Economics*, **118**(3), 815–842.

Eckles, D. and Bakshy, E. (2017). Bias and High-Dimensional Adjustment in Observational Studies of Peer Effects. *arXiv preprint arXiv:1706.04692*.

Eckles, D., Kizilcec, R. F., and Bakshy, E. (2016). Estimating Peer Effects in Networks with Peer Encouragement Designs. *Proceedings of the National Academy of Sciences*, **113**(27), 7316–7322.

Egami, N. and Hartman, E. (2018). Covariate Selection for Generalizing Experimental Results. *Working Paper*.

Elkins, Z. and Simmons, B. (2005). On Waves, Clusters, and Diffusion: A Conceptual Framework. *The Annals of the American Academy of Political and Social Science*, **598**(1), 33–51.

Flanders, W. D., Strickland, M. J., and Klein, M. (2017). A New Method for Partial Correction of Residual Confounding in Time-Series and Other Observational Studies. *American Journal of Epidemiology*, **185**(10), 941–949.

Fowler, J. H. and Christakis, N. A. (2010). Cooperative Behavior Cascades in Human Social Networks. *Proceedings of the National Academy of Sciences*, **107**(12), 5334–5338.

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

Frindte, W., Funke, F., and Waldzus, S. (1996). Xenophobia and Right-Wing-Extremism in German Youth Groups—Some Evidence against Unidimensional Misinterpretations. *International Journal of Intercultural Relations*, **20**(3-4), 463–478.

Galton, F. (1889). 'Comments' on E.B. Tylor 'On a Method of Investigating the Development of Institutions: Applied to Laws of Marriage and Descent'. *Journal of the Royal Anthropological Institute*, **18**, 245–256, 261–269.

Gibbons, S. and Overman, H. G. (2012). Mostly Pointless Spatial Econometrics? *Journal of Regional Science*, **52**(2), 172–191.

Gilardi, F. (2010). Who Learns from What in Policy Diffusion Processes? *American Journal of Political Science*, **54**(3), 650–666.

Gleditsch, K. S. and Ward, M. D. (2006). Diffusion and the International Context of Democratization. *International Organization*, **60**(04), 911–933.

Gleditsch, K. S., Salehyan, I., and Schultz, K. (2008). Fighting at Home, Fighting Abroad: How Civil Wars Lead to International Disputes. *Journal of Conflict Resolution*, **52**(4), 479–506.

Goldsmith-Pinkham, P. and Imbens, G. W. (2013). Social Networks and the Identification of Peer Effects. *Journal of Business & Economic Statistics*, **31**(3), 253–264.

Graham, E. R., Shipan, C. R., and Volden, C. (2013). The Diffusion of Policy Diffusion Research in Political Science. *British Journal of Political Science*, **43**(03), 673–701.

Granger, C. W. (1988). Some Recent Development in A Concept of Causality. *Journal of Econometrics*, **39**(1-2), 199–211.

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, **78**(6), 1360–1380.

Green, D. P., McFalls, L. H., and Smith, J. K. (2001). Hate Crime: An Emergent Research Agenda. *Annual Review of Sociology*, **27**(1), 479–504.

Greenhill, B. (2010). The Company You Keep: International Socialization and the Diffusion of Human Rights Norms. *International Studies Quarterly*, **54**(1), 127–145.

Greenhill, B. (2015). Replication Data for: Transmitting Rights: International Organizations and the Diffusion of Human Rights Practices. *Harvard Dataverse, DOI: 10.7910/DVN/29298*.

Greenhill, B. (2016). *Transmitting rights: International Organizations and the Diffusion of Human Rights Practices*. Oxford University Press.

Hagan, J., Merkens, H., and Boehnke, K. (1995). Delinquency and Disdain: Social Capital and the Control of Right-Wing Extremism among East and West Berlin Youth. *American Journal of Sociology*, **100**(4), 1028–1052.

Hainmueller, J. and Hiscox, M. J. (2007). Educated Preferences: Explaining Attitudes toward Immigration in Europe. *International Organization*, **61**(2), 399–442.

Halloran, M. E. and Struchiner, C. J. (1995). Causal Inference in Infectious Diseases. *Epidemiology*, **6**(2), 142–151.

Hill, S. and Rothchild, D. (1986). The Contagion of Political Conflict in Africa and the World. *Journal of Conflict Resolution*, **30**(4), 716–735.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis*, **15**(3), 199–236.

Houle, C., Kayser, M. A., and Xiang, J. (2016). Diffusion or Confusion? Clustered Shocks and the Conditional Diffusion of Democracy. *International Organization*, **70**(4), 687.

Huntington, S. P. (1991). *The Third Wave: Democratization in the Late Twentieth Century*. University of Oklahoma Press.

Hyde, S. D. (2011). Catch Us If You Can: Election Monitoring and International Norm Diffusion. *American Journal of Political Science*, **55**(2), 356–369.

Hyttinen, A., Plis, S., Järvisalo, M., Eberhardt, F., and Danks, D. (2016). Causal Discovery from Subsampled Time Series Data by Constraint Optimization. In *Proceedings of the 8th International Conference on Probabilistic Graphical Models (PGM)*, pages 216–227.

Imbens, G. W. (2004). Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review. *Review of Economics and Statistics*, **86**(1), 4–29.

Imbens, G. W. and Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, **142**(2), 615–635.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Jäckle, S. and König, P. D. (2016). The Dark Side of the German 'Welcome Culture': Investigating the Causes behind Attacks on Refugees in 2015. *West European Politics*, **40**(2), 223–251.

Jacob, B. A. and Lefgren, L. (2003). Are Idle Hands the Devil's Workshop? Incapacitation, Concentration, and Juvenile Crime. *American Economic Review*, **93**(5), 1560–1577.

Joffe, M. M. and Robins, J. M. (2009). Controlling the Future: Revised Assumptions and Methods for Causal Inference with Repeated Measures Outcomes. *Working Paper*.

Johnston, A. I. (2001). Treating International Institutions as Social Environments. *International Studies Quarterly*, **45**(4), 487–515.

Keck, M. E. and Sikkink, K. (1998). *Activists beyond Borders: Advocacy Networks in International Politics*. Cornell University Press.

Koopmans, R. and Olzak, S. (2004). Discursive Opportunities and the Evolution of Right-Wing Violence in Germany. *American Journal of Sociology*, **110**(1), 198–230.

Lake, D. A. and Rothchild, D. S. (1998). *The International Spread of Ethnic Conflict: Fear, Diffusion, and Escalation*. Princeton University Press.

Lauritzen, S. L. (1996). *Graphical Models*. Clarendon Press, Oxford.

Lipsitch, M., Tchetgen Tchetgen, E. J., and Cohen, T. (2010). Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies. *Epidemiology*, **21**(3), 383.

Lochner, L. and Moretti, E. (2004). The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports. *American Economic Review*, **94**(1), 155–189.

Lyons, R. (2011). The Spread of Evidence-Poor Medicine via Flawed Social-Network Analysis. *Statistics, Politics, and Policy*, **2**(1).

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, **60**(3), 531–542.

Meinshausen, N. and Bühlmann, P. (2006). High-Dimensional Graphs and Variable Selection with the Lasso. *The Annals of Statistics*, **34**(3), 1436–1462.

Miao, W. and Tchetgen Tchetgen, E. J. (2017). Invited Commentary: Bias Attenuation and Identification of Causal Effects with Multiple Negative Controls. *American Journal of Epidemiology*, **185**(10), 950–953.

Morozova, O., Cohen, T., and Crawford, F. W. (2018). Risk Ratios for Contagious Outcomes. *Journal of The Royal Society Interface*, **15**(138), 20170696.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. The MIT Press, Cambridge, MA.

Myers, D. J. (2000). The Diffusion of Collective Violence: Infectiousness, Susceptibility, and Mass Media Networks. *American Journal of Sociology*, **106**(1), 173–208.

Neyman, J. (1923). On the Application of Probability Theory to Agricultural Experiments. Essay on Principles (with discussion). Section 9 (translated). *Statistical Science*, **5**(4), 465–472.

Nickerson, D. W. (2008). Is Voting Contagious? Evidence from Two Field Experiments. *American Political Science Review*, **102**(01), 49–57.

Ogburn, E. L. and VanderWeele, T. J. (2014). Causal Diagrams for Interference. *Statistical Science*, **29**(4), 559–578.

Ogburn, E. L., Sofrygin, O., Diaz, I., and van der Laan, M. J. (2017). Causal Inference for Social Network Data. *arXiv preprint arXiv:1705.08527*.

O'Malley, A. J., Elwert, F., Rosenquist, J. N., Zaslavsky, A. M., and Christakis, N. A. (2014). Estimating Peer Effects in Longitudinal Dyadic Data Using Instrumental Variables. *Biometrics*, **70**(3), 506–515.

Pearl, J. (1995). Causal Diagrams for Empirical Research. *Biometrika*, **82**(4), 669–688.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. Cambridge University Press, Cambridge.

Pevehouse, J., Nordstrom, T., and Warnke, K. (2004). The Correlates of War 2 International Governmental Organizations Data. *Conflict Management and Peace Science*, **21**(2), 101–119.

Pevehouse, J. C. (2002). Democracy from the Outside-In? International Organizations and Democratization. *International Organization*, **56**(3), 515–549.

Robins, J. M., Hernan, M. A., and Brumback, B. (2000). Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology*, **11**(5), 550–560.

Rogers, E. M. (1962). *Diffusion of Innovations*. Simon and Schuster.

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, **66**(5), 688.

Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics*, **116**(2), 681–704.

Shalizi, C. R. and McFowland III, E. (2016). Controlling for Latent Homophily in Social Networks through Inferring Latent Locations. *arXiv preprint arXiv:1607.06565*.

Shalizi, C. R. and Thomas, A. C. (2011). Homophily and Contagion are Generically Confounded in Observational Social Network Studies. *Sociological Methods & Research*, **40**(2), 211–239.

Shpitser, I., VanderWeele, T., and Robins, J. M. (2012). On the Validity of Covariate Adjustment for Estimating Causal Effects. In *Proceedings of the 26th Conference on Uncertainty and Artificial Intelligence*, pages 527–536, Corvallis, WA. AUAI Press.

Simmons, B. A. and Elkins, Z. (2004). The Globalization of Liberalization: Policy Diffusion in the International Political Economy. *American Political Science Review*, pages 171–189.

Simmons, B. A., Dobbin, F., and Garrett, G. (2006). Introduction: The International Diffusion of Liberalism. *International Organization*, **60**(4), 781–810.

Sinclair, B. (2012). *The Social Citizen: Peer Networks and Political Behavior*. University of Chicago Press.

Skogan, W. G. (1990). *Disorder and Decline: Crime and the Spiral of Decay in American Neighborhoods*. University of California Press.

Sofer, T., Richardson, D. B., Colicino, E., Schwartz, J., and Tchetgen Tchetgen, E. J. (2016). On Negative Outcome Control of Unobserved Confounding as a Generalization of Difference-in-Differences. *Statistical Science*, **31**(3), 348.

Starr, H. (1991). Democratic Dominoes: Diffusion Approaches to the Spread of Democracy in the International System. *Journal of Conflict Resolution*, **35**(2), 356–381.

Stewart, B. (2014). Latent Factor Regressions for the Social Sciences. *Working Paper*.

Strang, D. (1991). Adding Social Structure to Diffusion Models: An Event History Framework. *Sociological Methods & Research*, **19**(3), 324–353.

Tarrow, S. G. (1994). *Power in Movement: Social Movements and Contentious Politics*. Cambridge University Press.

Tchetgen Tchetgen, E. (2013). The Control Outcome Calibration Approach for Causal Inference with Unobserved Confounding. *American Journal of Epidemiology*, **179**(5), 633–640.

Tchetgen Tchetgen, E. J., Fulcher, I., and Shpitser, I. (2017). Auto-G-Computation of Causal Effects on a Network. *arXiv preprint arXiv:1709.01577*.

van der Laan, M. J. (2014). Causal Inference for A Population of Causally Connected Units. *Journal of Causal Inference*, **2**(1), 13–74.

VanderWeele, T. J. (2011). Sensitivity Analysis for Contagion Effects in Social Networks. *Sociological Methods & Research*, **40**(2), 240–255.

VanderWeele, T. J. and An, W. (2013). Social Networks and Causal Inference. In *Handbook of Causal Analysis for Social Research*, pages 353–374. Springer.

VanderWeele, T. J., Ogburn, E. L., and Tchetgen Tchetgen, E. J. (2012). Why and When "Flawed" Social Network Analyses Still Yield Valid Tests of No Contagion. *Statistics, Politics and Policy*, **3**(1).

Ver Steeg, G. and Galstyan, A. (2010). Ruling out Latent Homophily in Social Networks. *NIPS Workshop on Social Computing*.

Ver Steeg, G. and Galstyan, A. (2013). Statistical Tests for Contagion in Observational Social Network Studies. In *the 16th International Conference on Artificial Intelligence and Statistics*, pages 563–571.

Wilson, J. Q. and Kelling, G. L. (1982). Broken Windows. *Atlantic Monthly*, **249**(3), 29–38.

Xu, Y. (2017). Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models. *Political Analysis*, **25**(1), 57–76.

Yang, E., Ravikumar, P., Allen, G. I., and Liu, Z. (2015). Graphical Models via Univariate Exponential Family Distributions. *Journal of Machine Learning Research*, **16**(1), 3813–3847.

Zhang, K., Peters, J., Janzing, D., and Schölkopf, B. (2012). Kernel-based Conditional Independence Test and Application in Causal Discovery. *arXiv preprint arXiv:1202.3775*.

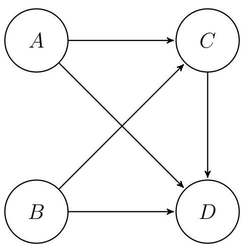
Zhang, M., Joffe, M. M., and Small, D. S. (2011). Causal Inference for Continuous-Time Processes When Covariates are Observed Only at Discrete Times. *Annals of Statistics*, **39**(1), 131 – 173.

Appendix

A Causal Directed Acyclic Graphs: Review

In this paper, I use a causal directed acyclic graph and nonparametric structural equations to represent causal relationships. Here, I review basic definitions and results. See Pearl (2009) for a comprehensive review. Following Pearl (1995), I define a causal directed acyclic graph (causal DAG) to be a set of nodes and directed edges among nodes such that the graph has no cycles and each node corresponds to a univariate random variable. Each random variable is given by its nonparametric structural equation. When there is a directed edge from one variable to another variable, the latter variable is a function of the former variable. For example, in a causal DAG in Figure 7 (a), four random variables (A, B, C, D) are given by nonparametric structural equations in Figure 7 (b); $A = f_A(\epsilon_A), B = f_B(\epsilon_B), C = f_C(A, B, \epsilon_C)$, and $D = f_D(A, B, C, \epsilon_D)$, where f_A, f_B, f_C and f_D are unknown nonparametric structural equations and $(\epsilon_A, \epsilon_B, \epsilon_C, \epsilon_D)$ are mutually independent errors. The node that a directed edge starts from is called the *parent* of the node that the edge goes into. The node that the edge goes into is the *child* of the node it comes from. If two nodes are connected by a directed path, the first node is the *ancestor* of every node on the path, and every node on the path is the *descendant* of the first node (Pearl, 2009). For example, node A is a parent of node C, and nodes C and D are descendants of node B. The requirement that the errors be mutually independent essentially means that there is no variable absent from the graph which, if included on the graph, would be a parent of two or more variables.

The nonparametric structural equations are general – random variables may depend on any function of their parents and variable-specific errors. They encode counterfactual relationships between the variables on the graph by recursively representing one-step-ahead counterfactuals. Under a hypothetical intervention setting A to a, the distribution of the variables B, C, and D are then recursively given by the nonparametric structural equations with $A = f_A(\epsilon_A)$ replaced by $A = a$. Specifically, $B = f_B(\epsilon_B),\ C = C(a) = f_C(A = a, B, \epsilon_C)$, and $D = D(a) = f_D(A = a, B, C = C(a), \epsilon_D)$ where $C(a), D(a)$ are the counterfactual values of C and D when A is set to a.



(a) A causal directed acyclic graph

$$\begin{aligned} A &= f_A(\epsilon_A) \\ B &= f_B(\epsilon_B) \\ C &= f_C(A, B, \epsilon_C) \\ D &= f_D(A, B, C, \epsilon_D) \end{aligned}$$

(b) A structural equation model

Figure 7: An Example of Causal DAGs and SEMs

B Proof of Result 1

Under Assumption 2, $Y_{i,t+1}(d) \perp\!\!\!\perp D_{it} \mid \mathbf{C}$ because $D_{it} \equiv \mathbf{W}_i^\top \mathbf{Y}_t$ and $\mathbf{C}$ includes $\mathbf{W}_i$. Hence,

$$\tau_{t+1}(d^H, d^L) = \int_{\mathcal{C}} \Big\{ \mathbb{E}[Y_{i,t+1}(d^H) \mid \mathbf{C} = \mathbf{c}] - \mathbb{E}[Y_{i,t+1}(d^L) \mid \mathbf{C} = \mathbf{c}] \Big\} dF_{\mathbf{C}}(\mathbf{c})$$

$$= \int_{\mathcal{C}} \left\{ \mathbb{E}[Y_{i,t+1} \mid D_{it} = d^H, \mathbf{C} = \mathbf{c}] - \mathbb{E}[Y_{i,t+1} \mid D_{it} = d^L, \mathbf{C} = \mathbf{c}] \right\} dF_{\mathbf{C}}(\mathbf{c}),$$

where the first equality follows from the linearity of expectation and the rule of conditional expectations, and the second from Assumptions 1 and 2.

C A Placebo Test

C.1 Proper Bias

Here, I define bias (i.e., the violation of the no omitted confounders assumption) to be *proper* if the bias cannot be removed by simply changing the lag structure of control variables. I provide a formal definition and then examples below.

Definition 4 (Proper Bias)
Suppose a control set $\mathbf{C}$ does not satisfy the no omitted confounders assumption (Assumption 2). This violation of the no omitted confounders assumption, i.e., bias, is defined to be proper if it satisfies the following condition.

If a control set $\mathbf{C}$ cannot block all back-door paths from $\{Y_{jt}\}_{j \in \mathcal{N}_i}$ to $Y_{i,t+1}$, there is at least one back-door path that any subset of the following set cannot block.

$$\{\mathbf{C}, \mathbf{C}^{(-1)}, \mathbf{C}^{(+1)}, \{Y_{j,t-1}\}_{j \in \mathcal{N}_i}\},$$

where $\mathbf{C}^{(-1)}$ and $\mathbf{C}^{(+1)}$ are a lag and a forward-lag of the time-dependent variables in $\mathbf{C}$. A control set $\mathbf{C}$ inducing no bias or proper bias is said to be proper.

Now, I examine Definition 4 in details. I consider two practical scenarios of proper bias. The first scenario is that a control set suffers from bias due to omitted time-independent confounders. Since the bias comes from the omitted time-independent variable, the change in the lag structure of existing control variables cannot adjust for it, i.e., this bias is proper. Homophily bias often fits into this category. For example, in the norm diffusion study (Greenhill, 2016), I discussed in Section 2.3 potential concerns about homophily bias due to unobserved political culture, which is often stable over time. Additionally, in causal DAGs in Shalizi and Thomas (2011) and O'Malley *et al.* (2014) as well as in Section 2.4 of this paper, the homophily bias results from time-independent omitted variables. In many applications, homophily bias is proper.

The second scenario is that a control set induces bias because it omits one type of time-dependent confounders not only at one time period but at every time period. Since this time-dependent confounder is not adjusted for at any time period, this bias is also proper. Contextual confounding often falls into this category. For example, in the study of hate crime diffusion, I examined in Section 2.3 possibilities of contextual confounding due to unobserved economic policies. Here, researchers are mainly concerned that they miss such economic policies entirely, not that they miss the variable at only one time period. As this example shows, in practice, contextual confounding is often proper.

Although many important types of biases are proper, when does a control set suffer from improper bias? It results from the misadjustment of the lag structure in control sets. For example, in the study of hate crime diffusion, suppose that bias exists because a control set omits a measure of economic conditions in July, but it includes the measure of economic conditions in August. In this case, because we can remove this bias by modifying the lag structure, i.e., controlling for the measure of economic conditions in July in addition to the one in August, this bias is not proper. When the lag-structure of time-dependent confounders is complex, it is more likely to suffer from improper bias.

As shown by these examples, Definition 4 encompasses a wide class of biases. For example, not only each of the two types but any mix of contextual confounding and homophily bias are often proper. Importantly, bias is proper unless it results from the misadjustment of the lag structure.

C.2 Proof of Theorem 1

C.2.1 Setup

Here, I provide some preliminary results useful for proving Theorem 1.

Lemma 2 (Equivalence between Back-Door Criteria and No Omitted Confounder Assumption) For a pretreatment control set $\mathbf{C}$, the following two statements hold.

1. If a set $\mathbf{C}$ satisfies the back-door criterion with respect to $(Y_{i,t+1}, \{Y_{jt}\}_{j \in \mathcal{N}_i})$ in causal DAG $\mathcal{G}$, then $Y_{i,t+1}(d) \perp\!\!\!\perp \{Y_{jt}\}_{j \in \mathcal{N}_i} \mid \mathbf{C}$ holds in every causal model inducing causal DAG $\mathcal{G}$ (Pearl, 1995).

2. If $Y_{i,t+1}(d) \perp\!\!\!\perp \{Y_{jt}\}_{j \in \mathcal{N}_i} \mid \mathbf{C}$ holds in every causal model inducing causal DAG $\mathcal{G}$, then a set $\mathbf{C}$ satisfies the back-door criterion with respect to $(Y_{i,t+1}, \{Y_{jt}\}_{j \in \mathcal{N}_i})$ in causal DAG $\mathcal{G}$ (Shpitser *et al.*, 2012).

Based on Lemma 2, $Y_{i,t+1}(d) \perp\!\!\!\perp \{Y_{jt}\}_{j \in \mathcal{N}_i} \mid \mathbf{C}$ is equivalent to no unblocked back-door paths from $\{Y_{jt}\}_{j \in \mathcal{N}_i}$ to $Y_{i,t+1}$ with respect to $\mathbf{C}$ in causal DAG $\mathcal{G}$. Additionally, based on Lemma 2, $Y_{it}(d) \perp\!\!\!\perp \{Y_{jt}\}_{j \in \mathcal{N}_i} \mid \mathbf{C}^P$ is equivalent to no unblocked back-door paths from $\{Y_{jt}\}_{j \in \mathcal{N}_i}$ to Y_{it} with respect to $\mathbf{C}^P$ in causal DAG $\mathcal{G}$. Under the sequential consistency assumption (Assumption 1), $Y_{it} = Y_{it}(d)$ for any d. Therefore, $Y_{it} \perp\!\!\!\perp \{Y_{jt}\}_{j \in \mathcal{N}_i} \mid \mathbf{C}^P$ is equivalent to no unblocked back-door paths from $\{Y_{jt}\}_{j \in \mathcal{N}_i}$ to Y_{it} with respect to $\mathbf{C}^P$ in causal DAG $\mathcal{G}$.

C.2.2 Proof of Theorem 1: Bias → Dependence in Placebo Test

In this proof, I show that when set $\mathbf{C}$ cannot block all back-door paths from $\{Y_{jt}\}_{j \in \mathcal{N}_i}$ to $Y_{i,t+1}$, set $\mathbf{C}^P$ cannot block all back-door paths from $\{Y_{jt}\}_{j \in \mathcal{N}_i}$ to Y_{it}.

Step 1 (Proper Bias): Given the assumption that the set $\mathbf{C}$ is proper, set $\mathbf{C}^P$ cannot block all back-door paths from $\{Y_{jt}\}_{j \in \mathcal{N}_i}$ to $Y_{i,t+1}$ because $\mathbf{C}^P$ is a subset of $\{\mathbf{C}, \mathbf{C}^{(-1)}, \mathbf{C}^{(+1)}, \{Y_{j,t-1}\}_{j \in \mathcal{N}_i}\}$.

Step 2 (Set up the main unblocked back-door path to investigate): Let π be a back-door path from $\{Y_{jt}\}_{j \in \mathcal{N}_i}$ to $Y_{i,t+1}$ that both $\mathbf{C}$ and $\mathbf{C}^P$ and any subset of $\{\mathbf{C}, \mathbf{C}^{(-1)}, \mathbf{C}^{(+1)}, \{Y_{j,t-1}\}_{j \in \mathcal{N}_i}\}$ cannot block. Without loss of generality, we assume that this unblocked back-door path starts with an arrow pointing to Y_{kt} where $k \in \mathcal{N}_i$ and it ends with an arrow pointing to $Y_{i,t+1}$.

Step 3 (Case I. the last node of the unblocked back-door path is time-independent): First, consider a case in which the last variable in an unblocked back-door path has a directed arrow pointing to $Y_{i,t+1}$ and time-independent. Let $(Z,\, Y_{i,t+1})$ denote the last two node path segment on π where Z is a time-independent variable and there exists a directed arrow from Z to $Y_{i,t+1}$. Note that we do not put any individual index to Z because the proof holds for any index. Since this is an unblocked path, Z is not in $\mathbf{C}^P$ and there is an unblocked back-door path from Y_{kt} to Z. Since Z is time-independent, there is a directed arrow from Z to Y_{it} by the definition of a stationary causal DAG (Definition 2). Therefore, set $\mathbf{C}^P$ cannot block this back-door path from Y_{kt} to Y_{it}.

Step 4 (Case II. the last node of the unblocked back-door path is time-dependent): Next, consider the case in which the last variable in an unblocked back-door path points to $Y_{i,t+1}$ and time-dependent. Let $(B, X_{t+1}, Y_{i,t+1})$ denote the last three node path segment on π where X_{t+1} is a time-dependent direct cause of $Y_{i,t+1}$. Note that we do not put any individual index to X_{t+1} because the proof holds for any index. Based on Lemma 3 presented below, $X_t, X_{t+1} \notin \mathbf{C}^P$ because $X_{t+1} \notin \mathbf{C}$.

Step 4.1 (sub-Case: the second last node is time-independent): First, assume B is time-independent. Then, because a causal DAG is stationary (Definition 2), X_t and B have the same relationship as the one between X_{t+1} and B. In addition, since there is an unblocked path from Y_{kt} to X_{t+1} to through B, there exists an unblocked path from Y_{kt} to X_t through B. Given that there exists a directed arrow from X_{t+1} to $Y_{i,t+1}$, there exists a directed arrow from X_t to Y_{it}. Therefore, there is an unblocked back-door path from Y_{kt} to Y_{it}.

Step 4.2 (sub-Case: the second last node is time-dependent): Next, assume B is time-dependent and therefore we use B_{t+1}. First, I show that whenever B is time-dependent, then the directed arrow is always from X_{t+1} to B_{t+1}. Suppose there is a directed arrow from B_{t+1} to X_{t+1}. If B_{t+1} in $\mathbf{C}^P$, then this back-door is blocked (therefore, choose another π). So, B_{t+1} is not in $\mathbf{C}^P$. Therefore, we can collapse B_{t+1} into X_{t+1}, meaning that if B is time dependent, then the directed arrow is always from X_{t+1} to B_{t+1}.

Now, suppose there is a directed arrow from X_{t+1} to B_{t+1}. We know there exists an unblocked path from Y_{kt} to X_{t+1} through B_{t+1}. Now, because $Y_{it} \leftarrow X_t \rightarrow X_{t+1} \rightarrow B_{t+1}$, there is an unblocked back-door path from Y_{kt} to Y_{it} because the underlying causal DAG is stationary. □

Lemma 3 (Substep in Step 4 in Appendix C.2.2) Within the fourth step of the proof above (Appendix C.2.2), $X_{t+1} \notin \mathbf{C} \rightarrow X_t, X_{t+1} \notin \mathbf{C}^P$.

Proof First, I show that $X_t, X_{t+1}, X_{t+2} \notin \mathbf{C}$ because set $\mathbf{C}$ is proper. It is because if X_t or X_{t+2} are in $\mathbf{C}$, then the lag adjustment of the control set $\mathbf{C}$ can block this path. If this path is the only back-door path, then $\mathbf{C}$ is not proper. If there is another back-door path that any subset of $\{\mathbf{C}, \mathbf{C}^{(-1)}, \mathbf{C}^{(+1)}, \{Y_{j,t-1}\}_{j\in\mathcal{N}_i}\}$ cannot block, choose it as π.

Next, I show that $X_t, X_{t+1} \notin \mathbf{C}^P$. There are three ways for a variable to be in the placebo set $\mathbf{C}^P$. I discuss them in order. First, a variable can be in the placebo set because it was already in the control set. We know $X_t, X_{t+1} \notin \mathbf{C}$, so this option is not feasible. Second, a variable can be in the placebo set because it is a lag of the original control variables. Given that X_{t+1}, X_{t+2} are not in the control set, this option is also not feasible. Finally, a variable can be in the placebo set because it is a lag of the treatment variable. (a) It is important to notice that $X_t \notin \{Y_{j,t-1}\}_{j\in\mathcal{N}_i}$ because $X_{t+1} \notin \{Y_{jt}\}_{j\in\mathcal{N}_i}$ (i.e., the treatment cannot be the last node of the unblocked back-door path). (b) Now, I verify $X_{t+1} \notin \{Y_{j,t-1}\}_{j\in\mathcal{N}_i}$. First, this back-door path can be blocked by a subset of $\{\mathbf{C}, \mathbf{C}^{(-1)}, \mathbf{C}^{(+1)}, \{Y_{j,t-1}\}_{j\in\mathcal{N}_i}\}$. If this back-door is the only unblocked back-door, set $\mathbf{C}$ is not proper, therefore this is contradictory. If there is another back-door path that both $\mathbf{C}$ and $\mathbf{C}^P$ cannot block, choose it as π. □

C.2.3 Proof of Theorem 1: No Bias → Independence in Placebo Test

Next, I prove that when set $\mathbf{C}$ can block all back-door paths from $\{Y_{jt}\}_{j\in\mathcal{N}_i}$ to $Y_{i,t+1}$, set $\mathbf{C}^P$ can block all back-door paths from $\{Y_{jt}\}_{j\in\mathcal{N}_i}$ to Y_{it}. I show the contraposition: when there is a back-door path from $\{Y_{jt}\}_{j\in\mathcal{N}_i}$ to Y_{it} that set $\mathbf{C}^P$ cannot block, set $\mathbf{C}$ cannot block all back-door paths from $\{Y_{jt}\}_{j\in\mathcal{N}_i}$ to $Y_{i,t+1}$. Since $\mathbf{C}$ does not include any $\text{Des}(Y_{kt})$, we know $\mathbf{C}^P$ also does not include any $\text{Des}(Y_{kt})$. Also, by definition, $\mathbf{C}^P$ does not include any $\text{Des}(Y_{it})$.

Therefore, without loss of generality, we can focus on unblocked back-door paths that start with an arrow pointing to Y_{kt} where $k \in \mathcal{N}_i$ and end with an arrow pointing to Y_{it}.

Step 1 (Control Set cannot block all back-door paths to the Placebo outcome): First, I show that when there is a back-door path from Y_{kt} to Y_{it} that set $\mathbf{C}^P$ cannot block, set $\mathbf{C}$ cannot block all back-door paths from Y_{kt} to Y_{it}. From set $\mathbf{C}^P$ to set $\mathbf{C}$, we need to (1) add $\text{Des}(Y_{it})$ and (2) remove $\mathbf{C}^{(-1)}$ and $\{Y_{j,t-1}\}_{j\in\mathcal{N}_i}$. I show here that this process cannot block a back-door path that set $\mathbf{C}^P$ cannot block. For (1), see Lemma 4 presented below. For (2), we first check whether removing $X_t \in \mathbf{C}^{(-1)}$ can block a back-door path that set $\mathbf{C}^P$ cannot block. To begin with, we can remove X_t because $X_{t+1} \in \mathbf{C}$. Removing variables X_t can be helpful if X_t is a collider or a descendant of a collider for a back-door path. However, if so, X_{t+1} is a descendant of a collider and it is in set $\mathbf{C}$ and therefore, removing X_t cannot block any additional paths. Next, we need to check whether removing a variable $B \in \{Y_{j,t-1}\}_{j\in\mathcal{N}_i}$ can block the back-door path that the set $\mathbf{C}^P$ cannot block. Removing variable B can be helpful if B is a collider or a descendant of a collider for a back-door path. If so, there is an unblocked back-door path (with respect to $\mathbf{C}^P$) that starts with an arrow pointing to B and ends with an arrow pointing to Y_{it}, i.e., $B \leftarrow \ldots \rightarrow Y_{it}$. Since B has a directed arrow pointing to Y_{kt}, removing B unblock a new back-door path from Y_{kt} through B, which points to Y_{it}. Although this unblocked back-door path with respect to $\mathbf{C}$ is different from the unblocked back-door path with respect to $\mathbf{C}^P$, the paths are the same after node B and therefore at least the last three nodes are the same. Therefore, we can use π to be a back-door from Y_{kt} to Y_{it} that both sets $\mathbf{C}$ and $\mathbf{C}^P$ cannot block.

Step 2 (Case I: the last node of the unblocked back-door path is time-independent): Consider the case in which the last two nodes are $(Z \rightarrow Y_{it})$ and Z is time-independent. Then, since $Z \rightarrow Y_{i,t+1}$ from the stationarity (Definition 2), set $\mathbf{C}$ cannot block this back-door.

Step 3 (Case II: the last node of the unblocked back-door path is time-dependent): Next, consider the case in which the last two nodes are $(X_t \rightarrow Y_{it})$. Since $X_t \notin \mathbf{C}^P$ and $X_t \notin \text{Des}(Y_{it})$, $X_t, X_{t+1} \notin \mathbf{C}$. Therefore, set $\mathbf{C}$ cannot block $Y_{kt} \leftarrow \cdots X_t \rightarrow X_{t+1} \rightarrow Y_{i,t+1}$. □

Lemma 4 (Substep in Step 1 in Appendix C.2.3) Adding $\text{Des}(Y_{it})$ cannot block a back-door path from Y_{kt} to Y_{it} unblocked by set $\mathbf{C}^P$.

Proof Suppose controlling for $\text{Des}(Y_{it})$ can block a back-door path from Y_{kt} to Y_{it} that the original set $\mathbf{C}^P$ cannot block. Since $\mathbf{C}^P$ does not include any $\text{Des}(Y_{kt})$ or $\text{Des}(Y_{it})$, this unblocked back-door path contains an arrow pointing to Y_{it}.

Step 1 (Set up the main node B): At least one of $\text{Des}(Y_{it})$ is a non-collider on this path given that controlling for $\text{Des}(Y_{it})$ can block this path. Let B be such a variable and focus on one arrow pointing out from the node B.

Step 2 (Case I. Consider one side of the main node B): First, suppose this direction leads to Y_{it}. Then, since B is a $\text{Des}(Y_{it})$, a directed path from node B to Y_{it} cannot exist and therefore, there must be a collider on this direction of the path. Since this collider is also in $\text{Des}(Y_{it})$ and therefore not controlled in the original $\mathbf{C}^P$, this back-door is blocked by set $\mathbf{C}^P$.

Step 3 (Case II. Consider the other side of the main node B): Next, consider the direction that leads to Y_{kt}. Then, since Y_{it} is not a cause of Y_{kt}, a directed path from node B to Y_{kt} cannot exist and therefore, there must be a collider on this direction of the path. Since this collider is also in $\text{Des}(Y_{it})$ and therefore not controlled in the original $\mathbf{C}^P$, this back-door is blocked by set $\mathbf{C}^P$. Hence, this is contradiction. This proves that controlling for Des(Y_{it}) cannot block a back-door path from Y_{kt} to Y_{it} that set $\mathbf{C}^P$ cannot block. □

D A Bias-Corrected Estimator

D.1 Proof of Theorem 2

Before proving Theorem 2, I demonstrate two useful lemmas here. First, I show that when set $\{U_{i,t+1}, \mathbf{C}\}$ can block all back-door paths from $\{Y_{jt}\}_{j\in\mathcal{N}_i}$ to $Y_{i,t+1}$, $\{U_{i,t+1}, \mathbf{X}_{i,t+1}, \mathbf{C}^B\}$ can also block all back-door paths from $\{Y_{jt}\}_{j\in\mathcal{N}_i}$ to $Y_{i,t+1}$

Lemma 5

$$Y_{i,t+1}(d^L) \perp\!\!\!\perp \{Y_{jt}\}_{j\in\mathcal{N}_i} \mid U_{i,t+1}, \mathbf{C} \Longrightarrow Y_{i,t+1}(d^L) \perp\!\!\!\perp \{Y_{jt}\}_{j\in\mathcal{N}_i} \mid U_{i,t+1}, \mathbf{X}_{i,t+1}, \mathbf{C}^B \tag{19}$$

Proof If I write out control set $\mathbf{C}$ using notations introduced in Section 4.2, the lemma can be rewritten as

$$\begin{aligned} & Y_{i,t+1}(d^L) \perp\!\!\!\perp \{Y_{jt}\}_{j\in\mathcal{N}_i} \mid U_{i,t+1}, \mathbf{X}_{i,t+1}, \mathbf{V}_{i,t+1}, \mathbf{Z}_i \\ \Longrightarrow\ & Y_{i,t+1}(d^L) \perp\!\!\!\perp \{Y_{jt}\}_{j\in\mathcal{N}_i} \mid U_{i,t+1}, \mathbf{X}_{i,t+1}, \mathbf{V}_{i,t+1}, \mathbf{V}_{it}, \mathbf{Z}_i, \{Y_{j,t-1}\}_{j\in\mathcal{N}_i}. \end{aligned}$$

First, note that all variables in set $\{U_{i,t+1}, \mathbf{X}_{i,t+1}, \mathbf{V}_{i,t+1}, \mathbf{V}_{it}, \mathbf{Z}_i, \{Y_{j,t-1}\}_{j\in\mathcal{N}_i}\}$ are neither affected by the potential outcome at time t, $Y_{i,t+1}(d^L)$, nor affected by the treatment $\{Y_{jt}\}_{j\in\mathcal{N}_i}$. The difference between the conditioning sets in the right- and left-hand sides is $\mathbf{V}_{it}$ and $\{Y_{j,t-1}\}_{j\in\mathcal{N}_i}$. Including these variables can open back-door paths only when these variables are colliders for these new back-door paths. However, because a descendant of $\mathbf{V}_{it}$, $\mathbf{V}_{i,t+1}$, is in the conditioning set, it is contradictory if conditioning on $\mathbf{V}_{it}$ can open a new back-door path. Additionally, because $\{Y_{j,t-1}\}_{j\in\mathcal{N}_i}$ is a parent of the treatment $\{Y_{jt}\}_{j\in\mathcal{N}_i}$, it is contradictory if conditioning on $\{Y_{j,t-1}\}_{j\in\mathcal{N}_i}$ can open a new back-door path. Therefore, including $\mathbf{V}_{it}$ and $\{Y_{j,t-1}\}_{j\in\mathcal{N}_i}$ don't open any back-door path, which completes the proof. □

Next, I prove the key equality under Assumption 3.

Lemma 6

$$\begin{aligned} & \mathbb{E}[Y_{i,t+1}(d^L) \mid D_{it} = d^H, \mathbf{X}_{i,t+1} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}] - \mathbb{E}[Y_{i,t+1}(d^L) \mid D_{it} = d^L, \mathbf{X}_{i,t+1} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}] \\ = \ & \mathbb{E}[Y_{it}(d^L) \mid D_{it} = d^H, \mathbf{X}_{it} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}] - \mathbb{E}[Y_{it}(d^L) \mid D_{it} = d^L, \mathbf{X}_{it} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}]. \end{aligned}$$

Proof Under Assumption 3,

$$\begin{aligned} & \int_{\mathcal{C}} \{\mathbb{E}[Y_{i,t+1}(d^L)|U_{i,t+1} = u_1, \mathbf{X}_{i,t+1} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}] - \mathbb{E}[Y_{i,t+1}(d^L)|U_{i,t+1} = u_0, \mathbf{X}_{i,t+1} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}]\} \\ & \qquad \times \{dF_{U_{i,t+1}|D_{it}=d^H, \mathbf{X}_{i,t+1}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}}(u_1) - dF_{U_{i,t+1}|D_{it}=d^L, \mathbf{X}_{i,t+1}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}}(u_1)\} \\ = \ & \int_{\mathcal{C}} \{\mathbb{E}[Y_{it}(d^L)|U_{it} = u_1, \mathbf{X}_{it} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}] - \mathbb{E}[Y_{it}(d^L)|U_{it} = u_0, \mathbf{X}_{it} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}]\} \\ & \qquad \times \{dF_{U_{it}|D_{it}=d^H, \mathbf{X}_{it}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}}(u_1) - dF_{U_{it}|D_{it}=d^L, \mathbf{X}_{it}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}}(u_1)\}. \end{aligned}$$

Now I analyze each side of the equation.

$$\int_{\mathcal{C}} \{\mathbb{E}[Y_{i,t+1}(d^L)|U_{i,t+1} = u_1, \mathbf{X}_{i,t+1} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}] - \mathbb{E}[Y_{i,t+1}(d^L)|U_{i,t+1} = u_0, \mathbf{X}_{i,t+1} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}]\}$$

$$\begin{aligned}
&\quad \times\{dF_{U_{i,t+1}|D_{it}=d^H,\mathbf{X}_{i,t+1}=\mathbf{x},\mathbf{C}^B=\mathbf{c}}(u_1) - dF_{U_{i,t+1}|D_{it}=d^L,\mathbf{X}_{i,t+1}=\mathbf{x},\mathbf{C}^B=\mathbf{c}}(u_1)\} \\
= & \int_{\mathcal{C}} \mathbb{E}[Y_{i,t+1}(d^L)|U_{i,t+1}=u_1, \mathbf{X}_{i,t+1}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}] \\
&\quad \times\{dF_{U_{i,t+1}|D_{it}=d^H,\mathbf{X}_{i,t+1}=\mathbf{x},\mathbf{C}^B=\mathbf{c}}(u_1) - dF_{U_{i,t+1}|D_{it}=d^L,\mathbf{X}_{i,t+1}=\mathbf{x},\mathbf{C}^B=\mathbf{c}}(u_1)\} \\
= & \int_{\mathcal{C}} \mathbb{E}[Y_{i,t+1}(d^L)|D_{it}=d^H, U_{i,t+1}=u_1, \mathbf{X}_{i,t+1}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}]dF_{U_{i,t+1}|D_{it}=d^H,\mathbf{X}_{i,t+1}=\mathbf{x},\mathbf{C}^B=\mathbf{c}}(u_1) \\
&\quad -\int_{\mathcal{C}} \mathbb{E}[Y_{i,t+1}(d^L)|D_{it}=d^L, U_{i,t+1}=u_1, \mathbf{X}_{i,t+1}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}]dF_{U_{i,t+1}|D_{it}=d^L,\mathbf{X}_{i,t+1}=\mathbf{x},\mathbf{C}^B=\mathbf{c}}(u_1) \\
= & \ \mathbb{E}[Y_{i,t+1}(d^L)|D_{it}=d^H, \mathbf{X}_{i,t+1}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}] - \mathbb{E}[Y_{i,t+1}(d^L)|D_{it}=d^L, \mathbf{X}_{i,t+1}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}],
\end{aligned}$$

where the first equality follows from the fact that $\mathbb{E}[Y_{i,t+1}(d^L)|U_{i,t+1}=u_0, \mathbf{X}_{i,t+1}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}]$ does not include u_1, the second equality comes from Lemma 5, and the final from the rule of conditional expectations. Similarly,

$$\begin{aligned}
& \int_{\mathcal{C}} \{\mathbb{E}[Y_{it}(d^L)|U_{it}=u_1, \mathbf{X}_{it}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}] - \mathbb{E}[Y_{it}(d^L)|U_{it}=u_0, \mathbf{X}_{it}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}]\} \\
&\quad \times\{dF_{U_{it}|D_{it}=d^H,\mathbf{X}_{it}=\mathbf{x},\mathbf{C}^B=\mathbf{c}}(u_1) - dF_{U_{it}|D_{it}=d^L,\mathbf{X}_{it}=\mathbf{x},\mathbf{C}^B=\mathbf{c}}(u_1)\}. \\
= & \ \mathbb{E}[Y_{it}(d^L) \mid D_{it}=d^H, \mathbf{X}_{it}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}] - \mathbb{E}[Y_{it}(d^L) \mid D_{it}=d^L, \mathbf{X}_{it}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}].
\end{aligned}$$

Taken together,

$$\begin{aligned}
& \mathbb{E}[Y_{i,t+1}(d^L) \mid D_{it}=d^H, \mathbf{X}_{i,t+1}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}] - \mathbb{E}[Y_{i,t+1}(d^L) \mid D_{it}=d^L, \mathbf{X}_{i,t+1}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}] \\
= & \ \mathbb{E}[Y_{it}(d^L) \mid D_{it}=d^H, \mathbf{X}_{it}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}] - \mathbb{E}[Y_{it}(d^L) \mid D_{it}=d^L, \mathbf{X}_{it}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}].
\end{aligned}$$

□

Proof of the theorem Based on Lemma 6 and Assumption 1,

$$\begin{aligned}
& \mathbb{E}[Y_{i,t+1}(d^L) \mid D_{it}=d^H, \mathbf{X}_{i,t+1}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}] \\
= & \ \mathbb{E}[Y_{i,t+1}(d^L) \mid D_{it}=d^L, \mathbf{X}_{i,t+1}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}] \\
&\quad +\mathbb{E}[Y_{it}(d^L) \mid D_{it}=d^H, \mathbf{X}_{it}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}] - \mathbb{E}[Y_{it}(d^L) \mid D_{it}=d^L, \mathbf{X}_{it}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}] \\
= & \ \mathbb{E}[Y_{i,t+1} \mid D_{it}=d^L, \mathbf{X}_{i,t+1}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}] \\
&\quad +\mathbb{E}[Y_{it} \mid D_{it}=d^H, \mathbf{X}_{it}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}] - \mathbb{E}[Y_{it} \mid D_{it}=d^L, \mathbf{X}_{it}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}].
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \mathbb{E}[Y_{i,t+1}(d^H) - Y_{i,t+1}(d^L) \mid D_{it}=d^H] \\
= & \int \big\{\mathbb{E}[Y_{i,t+1}(d^H) \mid D_{it}=d^H, \mathbf{X}_{i,t+1}, \mathbf{C}^B] - \mathbb{E}[Y_{i,t+1}(d^L) \mid D_{it}=d^H, \mathbf{X}_{i,t+1}, \mathbf{C}^B]\big\} dF_{\mathbf{X}_{i,t+1},\mathbf{C}^B|D_{it}=d^H}(\mathbf{x},\mathbf{c}) \\
= & \int \mathbb{E}[Y_{i,t+1} \mid D_{it}=d^H, \mathbf{X}_{i,t+1}, \mathbf{C}^B] dF_{\mathbf{X}_{i,t+1},\mathbf{C}^B|D_{it}=d^H}(\mathbf{x},\mathbf{c}) \\
& -\big\{\mathbb{E}[Y_{i,t+1} \mid D_{it}=d^L, \mathbf{X}_{i,t+1}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}] \\
&\quad +\mathbb{E}[Y_{it} \mid D_{it}=d^H, \mathbf{X}_{it}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}] - \mathbb{E}[Y_{it} \mid D_{it}=d^L, \mathbf{X}_{it}=\mathbf{x}, \mathbf{C}^B=\mathbf{c}]\big\} dF_{\mathbf{X}_{i,t+1},\mathbf{C}^B|D_{it}=d^H}(\mathbf{x},\mathbf{c}) \\
= & \int \big\{\mathbb{E}[Y_{i,t+1} \mid D_{it}=d^H, \mathbf{X}_{i,t+1}, \mathbf{C}^B] - \mathbb{E}[Y_{i,t+1} \mid D_{it}=d^L, \mathbf{X}_{i,t+1}, \mathbf{C}^B]\big\} dF_{\mathbf{X}_{i,t+1},\mathbf{C}^B|D_{it}=d^H}(\mathbf{x},\mathbf{c}) \\
& -\int \big\{\mathbb{E}[Y_{it} \mid D_{it}=d^H, \mathbf{X}_{it}, \mathbf{C}^B] - \mathbb{E}[Y_{it} \mid D_{it}=d^L, \mathbf{X}_{it}, \mathbf{C}^B]\big\} dF_{\mathbf{X}_{i,t+1},\mathbf{C}^B|D_{it}=d^H}(\mathbf{x},\mathbf{c}).
\end{aligned}$$

□

D.2 Other Cases

In Theorem 2, I consider cases in which $U_{i,t+1}$ is time-dependent and affected by the outcome at time t. Now I study two other cases (1) when $U_{i,t+1}$ is time-dependent but is not affected by the outcome at time t and (2) when unobserved confounder is time-independent Z_i. For both cases, Assumption 3 needs to be modified accordingly, although their substantive meanings stay the same. The definition of the bias-corrected estimator is also the same. For case (1), define $\widetilde{U}_i \equiv (U_{i,t+1}, U_{it})$ and for case (2), define $\widetilde{U}_i \equiv Z_i$. Then, Assumption 3 is modified as follows.

1. Time-invariant effect of unobserved confounder $\widetilde{U}$: For all $u_1, u_0, \mathbf{x}$ and $\mathbf{c}$,

$$\begin{aligned} & \mathbb{E}[Y_{i,t+1}(d^L) \mid \widetilde{U}_i = u_1, \mathbf{X}_{i,t+1} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}] - \mathbb{E}[Y_{i,t+1}(d^L) \mid \widetilde{U}_i = u_0, \mathbf{X}_{i,t+1} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}] \\ = \; & \mathbb{E}[Y_{it}(d^L) \mid \widetilde{U}_i = u_1, \mathbf{X}_{it} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}] - \mathbb{E}[Y_{it}(d^L) \mid \widetilde{U}_i = u_0, \mathbf{X}_{it} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}]. \end{aligned}$$

2. Time-invariant imbalance of unobserved confounder $\widetilde{U}$: For all $u, \mathbf{x}$ and $\mathbf{c}$,

$$\begin{aligned} & \Pr(\widetilde{U}_i \leq u \mid D_{it} = d^H, \mathbf{X}_{i,t+1} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}) - \Pr(\widetilde{U}_i \leq u \mid D_{it} = d^L, \mathbf{X}_{i,t+1} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}) \\ = \; & \Pr(\widetilde{U}_i \leq u \mid D_{it} = d^H, \mathbf{X}_{it} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}) - \Pr(\widetilde{U}_i \leq u \mid D_{it} = d^L, \mathbf{X}_{it} = \mathbf{x}, \mathbf{C}^B = \mathbf{c}). \end{aligned}$$

As in Assumption 3, the first condition requires that the effect of unobserved confounders on the potential outcomes is stable over time. The second condition requires that the imbalance of unobserved confounders is stable even after changing time-dependent confounders $\mathbf{X}$. In a special case where there is no descendant of Y_{it} in the control set, i.e., $\mathbf{X}_{i,t+1} = \mathbf{X}_{it} = \emptyset$, the second condition always holds and the first condition is equivalent to the parallel trend assumption necessary for the difference-in-difference design.

D.3 Diagnostics with Observed Confounders

Since Assumption 3 is fundamental to a successful application of the bias-corrected estimator, it is critical to assess its plausibility. Although I cannot directly test the assumption about unobserved confounders, I propose to inspect whether the effect and imbalance of *observed* confounders are stable over time. If not, it would call into question the assumption that the effect and imbalance of *unobserved* confounders are time-invariant. This is similar to diagnostics in the regression discontinuity design where researchers investigate the continuity of observed confounders to assess the plausibility of the continuity in the conditional expectation of potential outcomes (Imbens and Lemieux, 2008). See Cinelli and Hazlett (2018) for limitations of this type of diagnostics with observed confounders.

To investigate the assumption of the time-invariant effect of U, I propose to test whether the effect of each observed confounder is stable over time, conditional on other control variables. For the k th variable X^k, we can test

$$\begin{aligned}
&\mathbb{E}[Y_{i,t+1}|X^k_{i,t+1} = x_1, \mathbf{X}^{-k}_{i,t+1} = \bar{\mathbf{x}}, D_{it} = d^L, \mathbf{C}^B = \mathbf{c}] - \mathbb{E}[Y_{i,t+1}|X^k_{i,t+1} = x_0, \mathbf{X}^{-k}_{i,t+1} = \bar{\mathbf{x}}, D_{it} = d^L, \mathbf{C}^B = \mathbf{c}]\\
= \;& \mathbb{E}[Y_{it}|X^k_{it} = x_1, \mathbf{X}^{-k}_{it} = \bar{\mathbf{x}}, D_{it} = d^L, \mathbf{C}^B = \mathbf{c}] - \mathbb{E}[Y_{it}|X^k_{it} = x_0, \mathbf{X}^{-k}_{it} = \bar{\mathbf{x}}, D_{it} = d^L, \mathbf{C}^B = \mathbf{c}],
\end{aligned}$$

where $\mathbf{X}^{-k}$ is the observed confounders without the k th variable. One simple parametric approach is to run two separate linear additive regressions, (1) $Y_{i,t+1}$ on $(X^k_{i,t+1}, \mathbf{X}^{-k}_{i,t+1}, D_{it}, \mathbf{C}^B)$ and (2) Y_{it} on $(X^k_{it}, \mathbf{X}^{-k}_{it}, D_{it}, \mathbf{C}^B)$. Then, check whether a coefficient of $X^k_{i,t+1}$ from the first regression is equal to that of X^k_{it} from the second one. When these two coefficients are similar for all observed time-varying confounders, it provides suggestive evidence for the assumption of the time-invariant effect of unobserved confounders (Assumption 3.1).

To inspect the assumption of the time-invariant imbalance of U, it is helpful to test whether the imbalance of each observed confounder is stable over time. For the k th variable X^k, we can test

$$\begin{aligned}
&\Pr(X^k_{i,t+1} \leq x \mid D_{it} = d^H, \mathbf{X}^{-k}_{i,t+1} = \bar{\mathbf{x}}, \mathbf{C}^B = \mathbf{c}) - \Pr(X^k_{i,t+1} \leq x \mid D_{it} = d^L, \mathbf{X}^{-k}_{i,t+1} = \bar{\mathbf{x}}, \mathbf{C}^B = \mathbf{c})\\
= \;& \Pr(X^k_{it} \leq x \mid D_{it} = d^H, \mathbf{X}^{-k}_{it} = \bar{\mathbf{x}}, \mathbf{C}^B = \mathbf{c}) - \Pr(X^k_{it} \leq x \mid D_{it} = d^L, \mathbf{X}^{-k}_{it} = \bar{\mathbf{x}}, \mathbf{C}^B = \mathbf{c}).
\end{aligned}$$

In practice, researchers can estimate two parametric models, one for $X^k_{i,t+1}$ given $(D_{it}, \mathbf{X}^{-k}_{i,t+1}, \mathbf{C}^B)$ and the other for X^k_{it} given $(D_{it}, \mathbf{X}^{-k}_{it}, \mathbf{C}^B)$. Then, assess whether a coefficient of D_{it} from the first model is equal to the one from the second model, i.e., whether association between the treatment variable and confounders is stable over time. If we find that these two coefficients are similar for all observed time-varying confounders, it suggests that the assumption of the time-invariant imbalance of unobserved confounders (Assumption 3.2) is more likely to hold.

E Empirical Analysis in Section 5.1

E.1 Control Sets and Placebo Sets

I investigate five different control sets to illustrate how to use the proposed placebo test and bias-corrected estimator. Table 1 describes types of variables I use for those five control sets and their corresponding placebo sets. The column of "Main model" indicates variables used for control sets and the column of "Placebo model" indicates corresponding variables in placebo sets.

Type	**Main Model**	**Placebo Model**
Outcome	Physical Attack$_{t+1}$	Physical Attack$_t$
Treatment	Physical Attack$_t$ in Neighbors	Physical Attack$_t$ in Neighbors
A Control Set/A Placebo Set		
Basic Variables	Physical Attack$_t$	Physical Attack$_{t-1}$
	Physical Attack$_{t-1}$ in Neighbors	Physical Attack$_{t-1,t-2}$ in Neighbors
	the number of neighbors	the number of neighbors
	variance of $\mathbf{W}_i$	variance of $\mathbf{W}_i$
Two-month Lags	Physical Attack$_{t-1}$	Physical Attack$_{t-2}$
Contextual Variables (annual)		
Refugee variables	Total number of refugees	Total number of refugees
	Total number of foreign born	Total number of foreign born
Population variables	Population size	Population size
	Share of male inhabitants	Share of male inhabitants
Crime variables	Number of general crimes per 100,000 inhabitants	Number of general crimes per 100,000 inhabitants
	Percent of general crimes solved	Percept of general crimes solved
Economic variables	Number of newly registered business	Number of newly registered business
	Number of newly deregistered business	Number of newly deregistered business
	Number of insolvency	Number of insolvency
	per capita income	per capita income
	Number of employees with social security	Number of employees with social security
	Unemployment rate	Unemployment rate
Education variables	Share of school leavers without lower secondary education graduation	Share of school leavers without lower secondary education graduation
Political variables	Turnout rate in 2013	Turnout rate in 2013
	Vote share of extreme right and populist right-wing parties in 2013	Vote share of extreme right and populist right-wing parties in 2013

Table 1: Five Control Sets and Placebo Sets: Spatial Diffusion of Hate Crimes.

The first control set (C1) includes variables from "Basic Variables". The second control set (C2) adds variables from "Two-month Lags" to the first control set. The third control set adds state fixed effects to the second control set. The fourth control set adds all the variables from "Contextual Variables", which include variables on refugees, demographics, general crimes, economic indicators, education, and politics. Note that these contextual variables are measured only annually. The final fifth set adds the time trend variable as third-order polynomials to the fourth set.

E.2 Conditional ACDEs by Education

Here, I present the distribution of proportions of school dropouts without a secondary school diploma, separately for East Germany and West Germany. Because these distributions are substantially different between the East and the West (as evident in Figure 8), I estimate the conditional ACDE by proportions of school dropouts, separately for the East and the West. Although it is important to note that this investigation of heterogeneous effects is observational in nature, results are consistent with the education explanation.

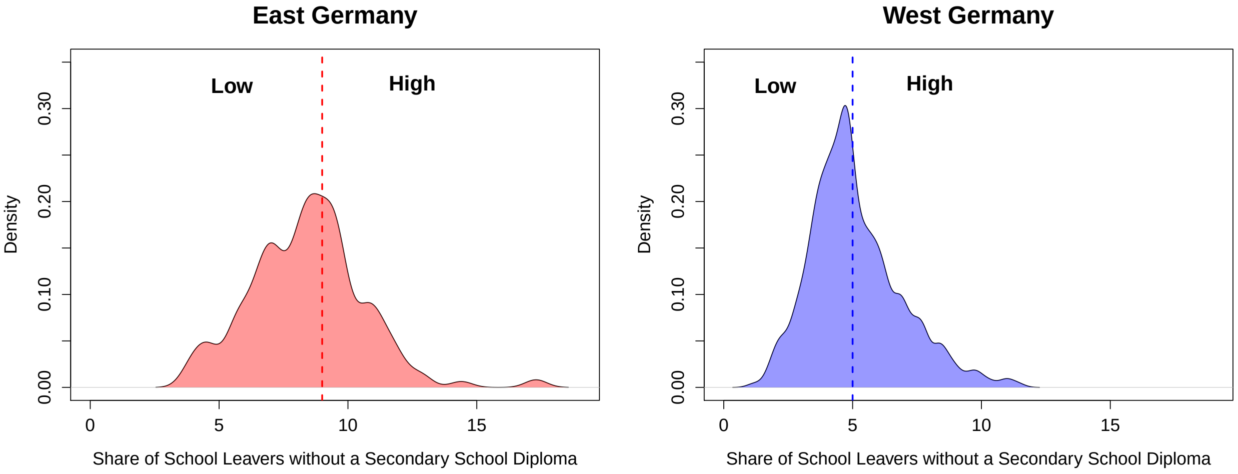


Figure 8: Distribution of Proportions of School Dropouts in East Germany and West Germany. Note: For East Germany, I use 9% as a cutoff for high and low proportions of school dropouts, which is approximately the median value in East Germany. For West Germany, I use 5% as a cutoff for high and low proportions of school dropouts, which is approximately the median value in West Germany.

Next, I present the conditional ACDE for counties in East Germany with low proportions of school dropouts. In contrast to Figure 5, estimates of the conditional ACDE are small.

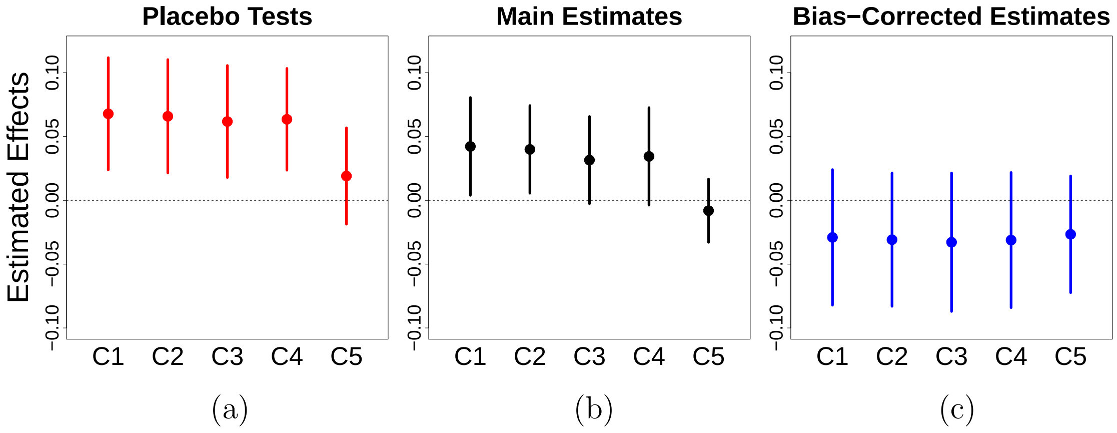
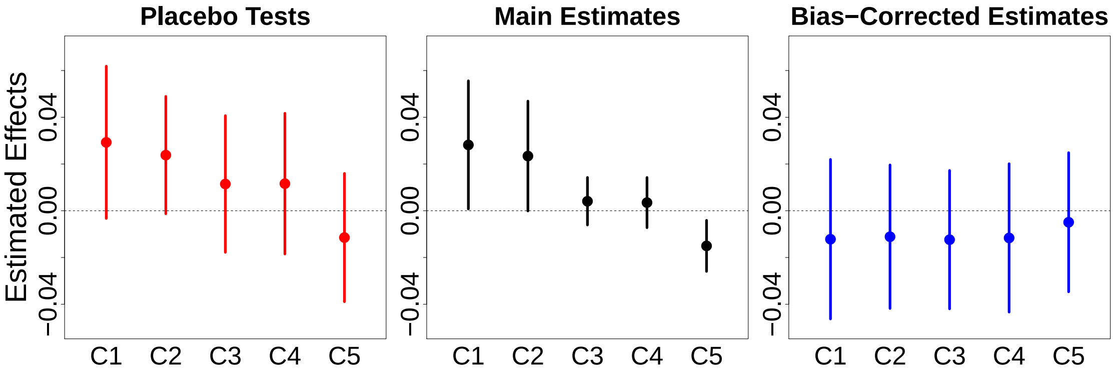


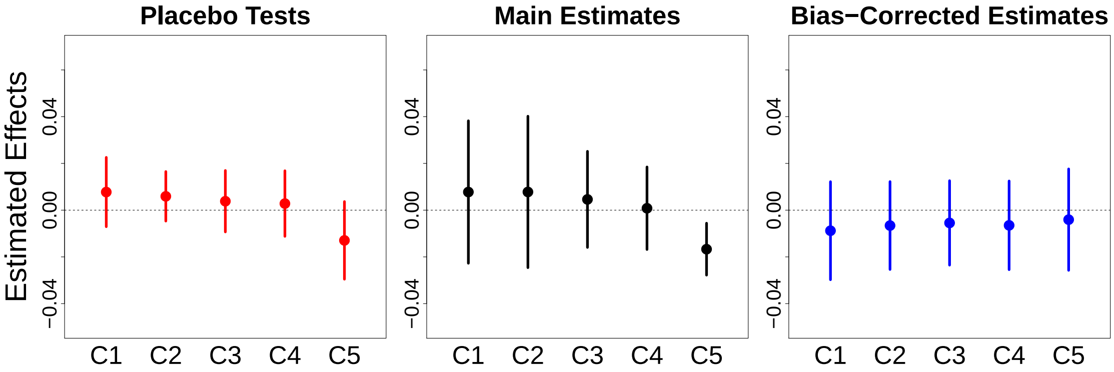
Figure 9: Results of the conditional ACDE (Low Proportion of School Dropouts, East). Note: Figure (a) shows that the last fifth set produces the smallest placebo estimate. Focusing on this fifth control set, a point estimate of the ACDE in Figure (b) is close to zero and its 95% confidence interval covers zero. Figure (c) shows that bias-corrected estimates are similar regardless of the selection of control variables and all of their 95% confidence intervals cover zero.

Now, I present the conditional ACDEs for counties in West Germany with high and low proportions of school dropouts. Given that proportions of school dropouts are lower in West Germany, estimates of the conditional ACDEs are small, in contrast to Figure 5.



(a) (b) (c)

Figure 10: Results of the conditional ACDE (High Proportion of School Dropouts, West). Note: Figure (a) shows that the third, fourth and fifth sets produce small placebo estimates. Focusing on these sets, point estimates of the ACDE in Figure (b) are close to zero and sometimes negative. Figure (c) shows that bias-corrected estimates are similar regardless of the selection of control variables and all of their 95% confidence intervals cover zero.



(a) (b) (c)

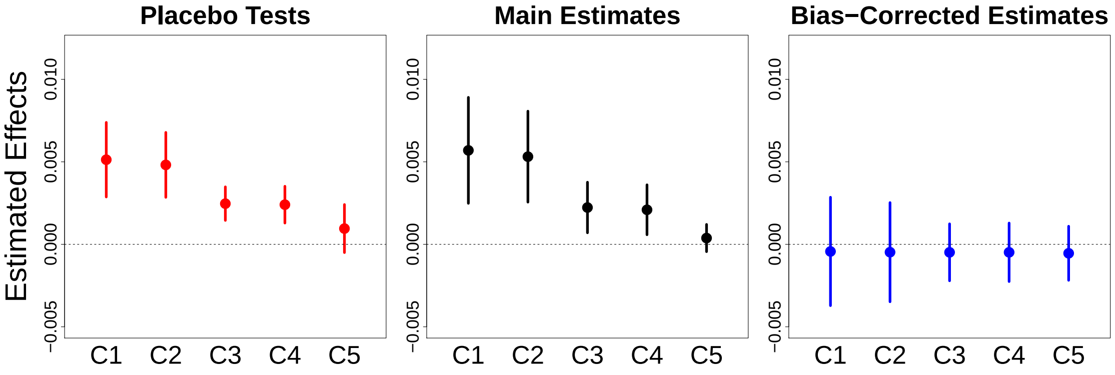
Figure 11: Results of the conditional ACDE (Low Proportion of School Dropouts, West). Note: Figure (a) shows that all the sets produce small placebo estimates. This is partly because there are few hate crimes in this area and hence, there is no variation in outcomes and treatments. In addition, point estimates of the ACDE in Figure (b) are close to zero and sometimes negative. Figure (c) shows that bias-corrected estimates are similar regardless of the selection of control variables and all of their 95% confidence intervals cover zero.

E.3 Robustness Checks

Here, I offer robustness checks for the main results in Section 5.1. I show the robustness with respect to different time-scale (weekly) and a different outcome type (ordinal outcomes).

E.3.1 Weekly Data

I estimate the ACDE and the conditional ACDEs with weekly data. Two findings are worth noting. First, the size of estimated effects is much smaller, which justifies the main analysis with the monthly data because the sequential consistency assumption (Assumption 1) holds approximately with monthly measures. Second, although estimates are smaller, substantive findings stay the same, which suggests that the main findings are robust to different time-scale.

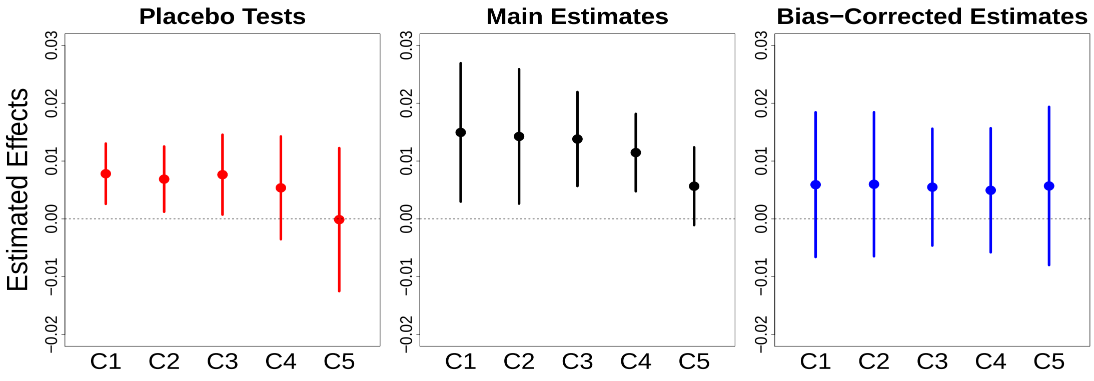


(a)

(b)

(c)

Figure 12: Results of the ACDE with Weekly data. Note: Point estimates are smaller than those presented in Figure 4, but the overall patterns stay the same. Figure (a) shows that the last fifth set produces the smallest placebo estimate. Focusing on the fifth control set, a point estimate of the ACDE in Figure (b) is close to zero and its 95% confidence interval covers zero. Figure (c) shows that bias-corrected estimates are similar regardless of the selection of control variables.

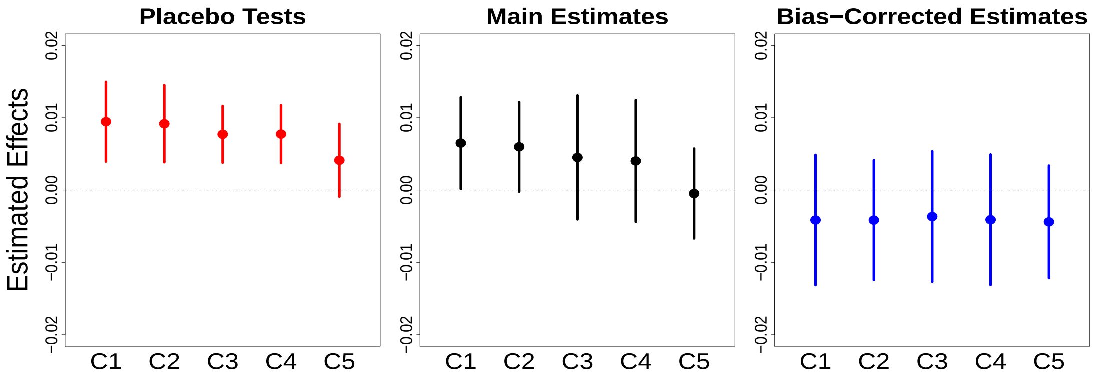


(a)

(b)

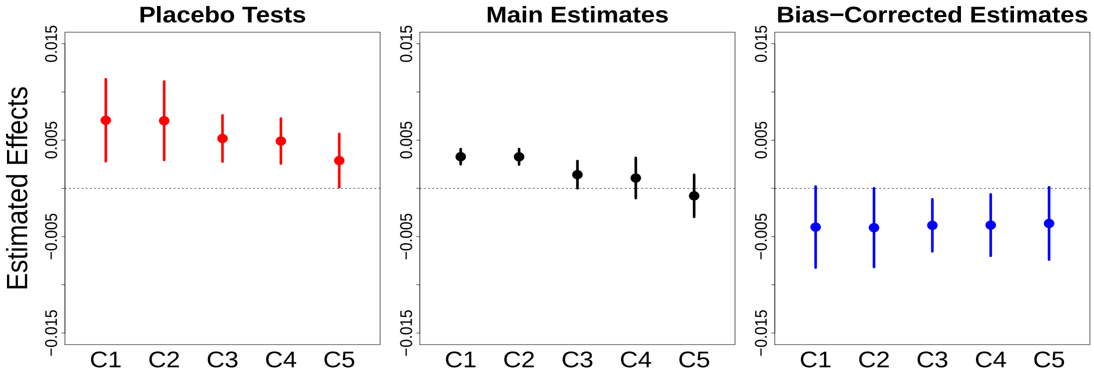
(c)

Figure 13: Results of the conditional ACDE (High Proportion of School Dropouts, East Germany) with Weekly data. Note: Point estimates are smaller than those presented in Figure 5, but the overall patterns stay the same. Figure (a) suggests that the last fifth set successfully adjusts for relevant confounders. Focusing on the most credible fifth control set, a point estimate of the conditional ACDE in Figure (b) is as large as 0.5 percentage points. Figure (c) shows that bias-corrected estimates are similar regardless of the selection of control variables. Unlike the analysis with the monthly data, their 95% confidence intervals cover zero.



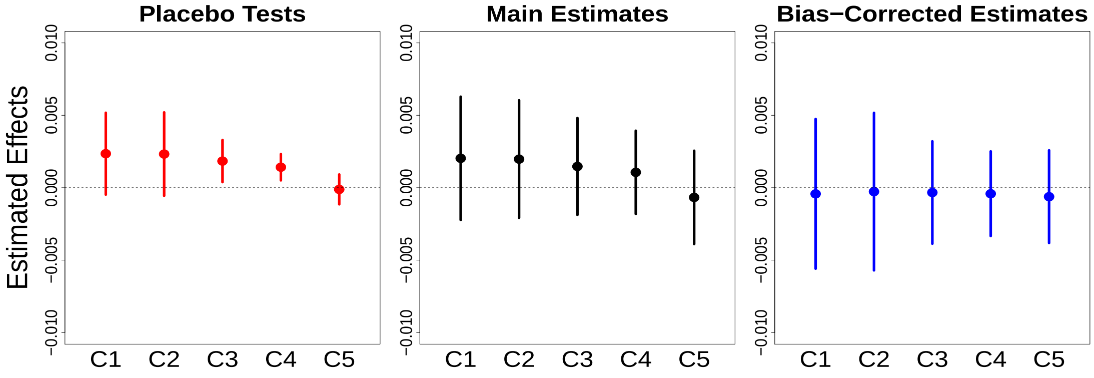
(a) (b) (c)

Figure 14: Results of the conditional ACDE (Low Proportion of School Dropouts, East Germany) with Weekly data. Note: Point estimates are smaller than those presented in Figure 9, but the overall patterns stay the same.



(a) (b) (c)

Figure 15: Results of the conditional ACDE (High Proportion of School Dropouts, West Germany) with Weekly data. Note: Point estimates are smaller than those presented in Figure 10, but the overall patterns stay the same.

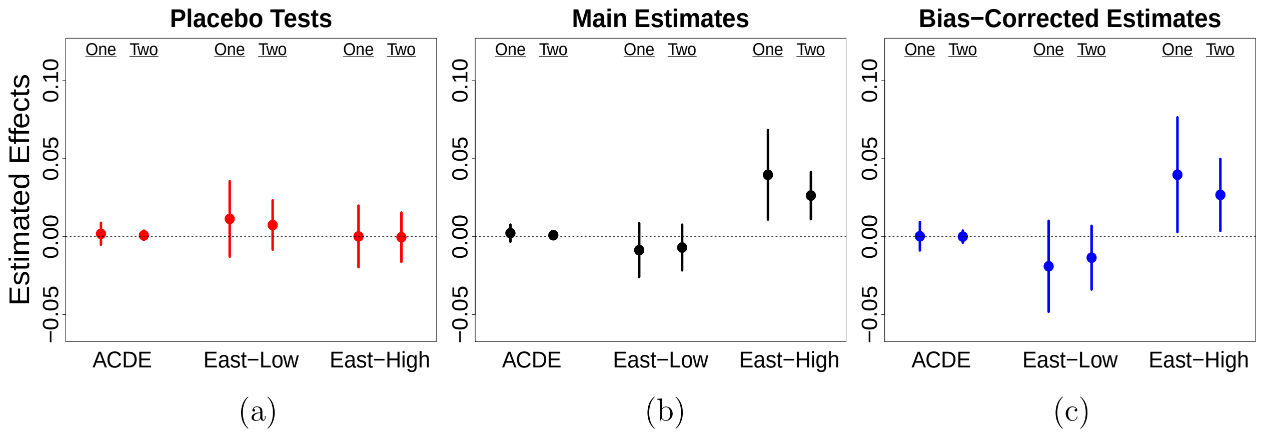


(a) (b) (c)

Figure 16: Results of the conditional ACDE (Low Proportion of School Dropouts, West Germany) with Weekly data. Note: Point estimates are smaller than those presented in Figure 11, but the overall patterns stay the same.

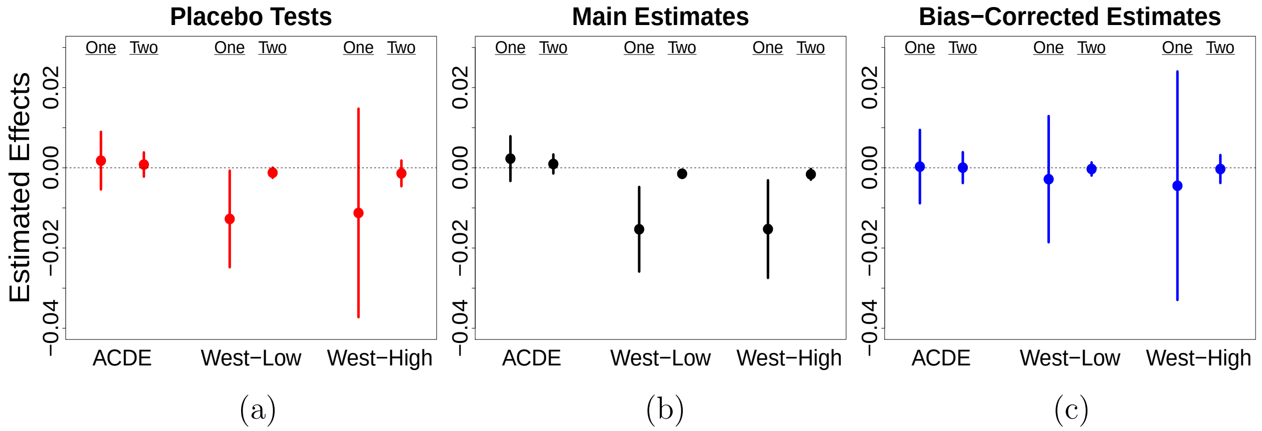
E.3.2 Count Data

I estimate the ACDE and the conditional ACDEs with ordinal outcomes. In Section 5.1, I analyze the binary outcome, indicating whether there is at least one physical attack in a given month. Here, I show that results are robust to a different outcome type. Because only 1.5% of counties experience more than one and only 0.3% of counties experience more than two, I focus on the three-level ordinal outcome: (1) no physical attack, (2) one physical attack, and (3) more than or equal to two physical attacks. I report the ACDE on the probability of having one physical attack relative to no physical attack ("One") and on the probability of having more than or equal to two physical attacks relative to no physical attack ("Two"). I adjust for lagged outcomes, lagged treatments, the number of neighbors, variance of $\mathbf{W}_i$, state fixed effects, and the time trend variable as third-order polynomials.



(a) (b) (c)

Figure 17: Results of the ACDE and the conditional ACDEs in East Germany with Count data. Note: As presented in Figure 4, estimates for the ACDE are small. However, estimates of the conditional ACDEs differ according to proportions of school dropouts. As presented in Figure 5, estimates of the conditional ACDE for counties in East Germany with high proportions of school dropouts are large. On the other hand, estimates of the conditional ACDE for counties in East Germany with low proportions of school dropouts are small.



(a) (b) (c)

Figure 18: Results of the ACDE and the conditional ACDEs in West Germany with Count data. Note: For comparison, I again present estimates for the ACDE. As in Figures 10 and 11, estimates of the conditional ACDEs are small in West Germany given that proportions of school dropouts are substantially lower than in East Germany (see Figure 8).

F Empirical Analysis in Section 5.2

F.1 Control Sets and Placebo Sets

I investigate three different control sets to illustrate how to use the proposed placebo test and bias-corrected estimator. Table 2 describes what types of variables I use for those three control sets and their corresponding placebo sets. The column of "Main model" indicates variables used for control sets and the column of "Placebo model" indicates corresponding variables in placebo sets.

The first control set (C1) includes variables from the original analysis. It includes logged GDP per capita, regime durability, population density, democracy, trade dependence, FDI dependence, conflict, and a lagged dependent variable. The second control set (C2) adds two network measures, i.e., spatial proximity and cultural similarity between states. Finally, the third set (C3) adjusts for year fixed effects and longer periods of the diffusion history, i.e., two- and three-year lagged dependent and treatment variables, in addition to basic network characteristics.

Type	**Main Model**	**Placebo Model**
Outcome	PIR score $_{t+1}$	PIR score $_{t}$
Treatment	PIR score $_{t}$ of IGO partners	PIR score $_{t}$ of IGO partners
A Control Set/A Placebo Set		
Original Variables	PIR score$_t$	PIR score$_{t-1}$
	logged GDP per capita$_t$	logged GDP per capita$_{t-1}$
	regime durability$_t$	regime durability$_{t-1}$
	population density$_t$	population density$_{t-1}$
	democracy$_t$	democracy$_{t-1}$
	trade dependence$_t$	trade dependence$_{t-1}$
	FDI dependence$_t$	FDI dependence$_{t-1}$
	conflict$_t$	conflict$_{t-1}$
Alternative Network Measures	spatial proximity$_t$	spatial proximity$_{t,t-1}$
	cultural similarity$_t$	cultural similarity$_{t,t-1}$
Time-trend Variables	year fixed effects	year fixed effects
	PIR score$_{t-1}$	PIR score$_{t-2}$
	PIR score$_{t-2}$	PIR score$_{t-3}$
	PIR score $_{t-1,t-2}$ of IGO partners	PIR score $_{t-1,t-2,t-3}$ of IGO partners
Network characteristics	the number of IGO partners$_t$	the number of IGO partners$_{t,t-1}$
	variance of $\mathbf{W}_{it}$	variance of $\mathbf{W}_{it}$ and $\mathbf{W}_{i,t-1}$

Table 2: Three Control Sets and Placebo Sets: Network Diffusion of Human Rights Norms.

F.2 Descriptive Statistics

Here, I present basic descriptive statistics: the distribution of the treatment variable and association between the treatment variable and the outcome variable.

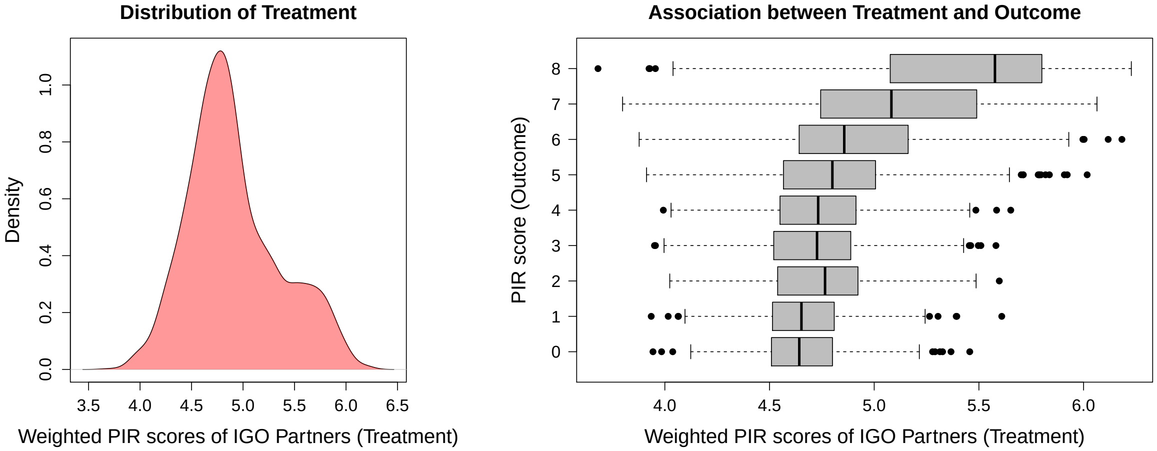


Figure 19: Distribution of the Treatment (Left) and Association between the Treatment and the Outcome (Right). Note: The outcome ranges from 0 to 8, and the treatment from 3.68 to 6.23.