Supplementary Appendix

A Proofs

A.1 Proof of Theorem 1

In this proof, we use $\overline{\mathbf{C}}$ and $\overline{\mathbf{C}}^P$ to denote $\overline{\mathbf{C}}_{it}$ and $\overline{\mathbf{C}}_{it}^P$ for notational simplicity.

A.1.1 Setup

Given that adjustment set $\overline{\mathbf{C}}$ are defined to be pre-treatment (i.e., variables not affected by the treatment), theoretical results on causal DAGs (Pearl, 1995; Shpitser *et al.*, 2012) imply that $Y_{it}(d) \perp\!\!\!\perp \mathbf{Y}_{\mathcal{N}_i,t-1} \mid \overline{\mathbf{C}}$ is equivalent to no unblocked back-door paths from $\mathbf{Y}_{\mathcal{N}_i,t-1}$ to Y_{it} with respect to $\overline{\mathbf{C}}$ in causal DAG $\mathcal{G}$ (see Lemma 1). Additionally, $Y_{i,t-1} \perp\!\!\!\perp \mathbf{Y}_{\mathcal{N}_i,t-1} \mid \overline{\mathbf{C}}^P$ is equivalent to no unblocked back-door paths from $\mathbf{Y}_{\mathcal{N}_i,t-1}$ to $Y_{i,t-1}$ with respect to $\overline{\mathbf{C}}^P$ in causal DAG $\mathcal{G}$.

The theorem requires one regularity condition – the violation of the no omitted confounders assumption, if any, is *proper*. Intuitively, 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. Importantly, contextual confounding and homophily bias are proper, and hence within the scope of this theorem.

Definition 1 (Proper Bias)

Suppose adjustment set $\overline{\mathbf{C}}$ does not satisfy Assumption 1. This violation (bias) is defined to be proper when it satisfies the following condition: If control set $\overline{\mathbf{C}}_{it}$ cannot block all back-door paths from $\mathbf{Y}_{\mathcal{N}_i,t-1}$ to Y_{it}, there is at least one back-door path that any subset of the following set cannot block.

$$\{\overline{\mathbf{C}}_{it}, \overline{\mathbf{C}}_{it}^{(-1)}, \overline{\mathbf{C}}_{it}^{(+1)}, \mathbf{Y}_{\mathcal{N}_i,t-2}\},$$

where $\overline{\mathbf{C}}_{it}^{(-1)}$ and $\overline{\mathbf{C}}_{it}^{(+1)}$ are a lag and a lead of the time-dependent variables in $\overline{\mathbf{C}}_{it}$.

A.1.2 Bias → Dependence in Placebo Test

Here, we show that when set $\overline{\mathbf{C}}$ cannot block all back-door paths from $\mathbf{Y}_{\mathcal{N}_i,t-1}$ to Y_{it}, set $\overline{\mathbf{C}}^P$ cannot block all back-door paths from $\mathbf{Y}_{\mathcal{N}_i,t-1}$ to $Y_{i,t-1}$.

Step 1 (Proper Bias): Given the assumption that the set $\overline{\mathbf{C}}$ is proper, set $\overline{\mathbf{C}}^P$ cannot block all back-door paths from $\mathbf{Y}_{\mathcal{N}_i,t-1}$ to Y_{it} because $\overline{\mathbf{C}}^P$ is a subset of $\{\overline{\mathbf{C}}, \overline{\mathbf{C}}^{(-1)}, \overline{\mathbf{C}}^{(+1)}, \mathbf{Y}_{\mathcal{N}_i,t-2}\}$.

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

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_{it} and time-independent. Let (Z, Y_{it}) 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_{it}. 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 $\overline{\mathbf{C}}^P$ and there is an unblocked back-door path from $Y_{k,t-1}$ to Z. Since Z is time-independent, there is a directed arrow from Z to $Y_{i,t-1}$ by the structural stationarity (Assumption 2). Therefore, set $\overline{\mathbf{C}}^P$ cannot block this back-door path from $Y_{k,t-1}$ to $Y_{i,t-1}$.

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_{it} and time-dependent. Let (B, X_t, Y_{it}) denote the last three node path segment on π where X_t is a time-dependent direct cause of Y_{it}. Note that we do not put any individual index to X_t because the proof holds for any index. $X_{t-1}, X_t \notin \overline{\mathbf{C}}^P$ because $X_t \notin \overline{\mathbf{C}}$ (see Lemma 2 in Section A.2).

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

Step 4.2 (sub-Case: the second last node is time-dependent): Next, assume B is time-dependent and therefore we use B_t. First, we show that whenever B is time-dependent, then the directed arrow is always from X_t to B_t. Suppose there is a directed arrow from B_t

to X_t. If B_t in $\overline{\mathbf{C}}^P$, then this back-door is blocked (therefore, choose another π). So, B_t is not in $\overline{\mathbf{C}}^P$. Therefore, we can collapse B_t into X_t, meaning that if B is time dependent, then the directed arrow is always from X_t to B_t.

Now, suppose there is a directed arrow from X_t to B_t. We know there exists an unblocked path from $Y_{k,t-1}$ to X_t through B_t. Now, because $Y_{i,t-1} \leftarrow X_{t-1} \rightarrow X_t \rightarrow B_t$, there is an unblocked back-door path from $Y_{k,t-1}$ to $Y_{i,t-1}$ because the underlying causal DAG satisfies structural stationarity. $\square$

A.1.3 No Bias $\rightarrow$ Independence in Placebo Test

Next, we prove that when set $\overline{\mathbf{C}}$ can block all back-door paths from $\mathbf{Y}_{\mathcal{N}_i,t-1}$ to Y_{it}, set $\overline{\mathbf{C}}^P$ can block all back-door paths from $\mathbf{Y}_{\mathcal{N}_i,t-1}$ to $Y_{i,t-1}$. We show the contraposition: when there is a back-door path from $\mathbf{Y}_{\mathcal{N}_i,t-1}$ to $Y_{i,t-1}$ that set $\overline{\mathbf{C}}^P$ cannot block, set $\overline{\mathbf{C}}$ cannot block all back-door paths from $\mathbf{Y}_{\mathcal{N}_i,t-1}$ to Y_{it}. Since $\overline{\mathbf{C}}$ does not include any $\text{Des}(Y_{k,t-1})$, we know $\overline{\mathbf{C}}^P$ also does not include any $\text{Des}(Y_{k,t-1})$. Also, by definition, $\overline{\mathbf{C}}^P$ does not include any $\text{Des}(Y_{i,t-1})$. Therefore, without loss of generality, we can focus on unblocked back-door paths that start with an arrow pointing to $Y_{k,t-1}$ where $k \in \mathcal{N}_i$ and end with an arrow pointing to $Y_{i,t-1}$.

Step 1 (Control Set cannot block all back-door paths to the Placebo outcome): First, we show that when there is a back-door path from $Y_{k,t-1}$ to $Y_{i,t-1}$ that set $\overline{\mathbf{C}}^P$ cannot block, set $\overline{\mathbf{C}}$ cannot block all back-door paths from $Y_{k,t-1}$ to $Y_{i,t-1}$. From set $\overline{\mathbf{C}}^P$ to set $\overline{\mathbf{C}}$, we need to (1) add $\text{Des}(Y_{i,t-1})$ and (2) remove $\overline{\mathbf{C}}^{(-1)}$ and $\mathbf{Y}_{\mathcal{N}_i,t-2}$. We show here that this process cannot block a back-door path that set $\overline{\mathbf{C}}^P$ cannot block. The step (1) cannot block the back-door path because adding $\text{Des}(Y_{i,t-1})$ cannot block a back-door path from $Y_{k,t-1}$ to $Y_{i,t-1}$ unblocked by set $\overline{\mathbf{C}}^P$ (see Lemma 3 in Section A.2). For (2), we first check whether removing $X_{t-1} \in \overline{\mathbf{C}}^{(-1)}$ can block a back-door path that set $\overline{\mathbf{C}}^P$ cannot block. To begin with, we can remove X_{t-1} because $X_t \in \overline{\mathbf{C}}$. Removing variables X_{t-1} can be helpful if X_{t-1} is a collider or a descendant of a collider for a back-door path. However, if so, X_t is a descendant of a collider and it is in set $\overline{\mathbf{C}}$ and therefore, removing X_{t-1} cannot block any additional paths. Next, we need to check whether removing a variable $B \in \mathbf{Y}_{\mathcal{N}_i,t-2}$ can block the back-door path that the set $\overline{\mathbf{C}}^P$ 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 $\overline{\mathbf{C}}^P$) that starts with an arrow pointing to B and ends with an arrow pointing to $Y_{i,t-1}$, i.e., $B \leftarrow \ldots \rightarrow Y_{i,t-1}$. Since B has a directed arrow pointing to $Y_{k,t-1}$, removing B unblock a new back-door path from $Y_{k,t-1}$ through B, which points to $Y_{i,t-1}$. Although this unblocked back-door path with respect to $\overline{\mathbf{C}}$ is different from the unblocked back-door path with respect to $\overline{\mathbf{C}}^P$, 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 $Y_{k,t-1}$ to $Y_{i,t-1}$ that both sets $\overline{\mathbf{C}}$ and $\overline{\mathbf{C}}^P$ 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 $(Z \rightarrow Y_{i,t-1})$ and Z is time-independent. Then, since $Z \rightarrow Y_{it}$ from structural stationarity (Assumption 2), set $\overline{\mathbf{C}}$ 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 $(X_{t-1} \rightarrow Y_{i,t-1})$. Since $X_{t-1} \notin \overline{\mathbf{C}}^P$ and $X_{t-1} \notin \text{Des}(Y_{i,t-1})$, $X_{t-1}, X_t \notin \overline{\mathbf{C}}$. Therefore, set $\overline{\mathbf{C}}$ cannot block $Y_{k,t-1} \leftarrow \cdots X_{t-1} \rightarrow X_t \rightarrow Y_{it}$. □

A.2 Proof of Lemmas used for Theorem 1

Here, we prove all the lemmas used to prove Theorem 1.

Lemma 1 (Equivalence between Back-Door Criteria and No Omitted Confounder Assumption (Shpitser *et al.*, 2012)) For a pretreatment adjustment set $\overline{\mathbf{C}}$ (i.e., variables not affected by the treatment), the following two statements hold.

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

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

Lemma 2 $X_t \notin \overline{\mathbf{C}} \rightarrow X_{t-1}, X_t \notin \overline{\mathbf{C}}^P$.

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

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

Lemma 3 Adding $\text{Des}(Y_{i,t-1})$ cannot block a back-door path from $Y_{k,t-1}$ to $Y_{i,t-1}$ unblocked by set $\overline{\mathbf{C}}^P$.

Proof Suppose controlling for $\text{Des}(Y_{i,t-1})$ can block a back-door path from $Y_{k,t-1}$ to $Y_{i,t-1}$ that the original set $\overline{\mathbf{C}}^P$ cannot block. Since $\overline{\mathbf{C}}^P$ does not include any $\text{Des}(Y_{k,t-1})$ or $\text{Des}(Y_{i,t-1})$, this unblocked back-door path contains an arrow pointing to $Y_{i,t-1}$.

Step 1 (Set up the main node B): At least one of $\text{Des}(Y_{i,t-1})$ is a non-collider on this path given that controlling for $\text{Des}(Y_{i,t-1})$ 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 $Y_{i,t-1}$. Then, since B is a $\text{Des}(Y_{i,t-1})$, a directed path from node B to $Y_{i,t-1}$ cannot exist and therefore, there must be a collider on this direction of the path. Since this collider is also in $\text{Des}(Y_{i,t-1})$ and therefore not controlled in the original $\overline{\mathbf{C}}^P$, this back-door is blocked by set $\overline{\mathbf{C}}^P$.

Step 3 (Case II. Consider the other side of the main node B): Next, consider the direction that leads to $Y_{k,t-1}$. Then, since $Y_{i,t-1}$ is not a cause of $Y_{k,t-1}$, a directed path from node B to $Y_{k,t-1}$ cannot exist and therefore, there must be a collider on this direction of the path. Since this collider is also in $\text{Des}(Y_{i,t-1})$ and therefore not controlled in the original $\overline{\mathbf{C}}^P$, this back-door is blocked by set $\overline{\mathbf{C}}^P$. Hence, this is contradiction. This proves that controlling for $\text{Des}(Y_{i,t-1})$ cannot block a back-door path from $Y_{k,t-1}$ to $Y_{i,t-1}$ that set $\overline{\mathbf{C}}^P$ cannot block. □

A.3 Proof of Theorem 2

Below, we describe two lemmas useful for proving Theorem 2. For completeness, their proofs follow.

Lemma 4

$$Y_{it}(d^L) \perp\!\!\!\perp \mathbf{Y}_{\mathcal{N}_i,t-1} \mid U_{it}, \overline{\mathbf{C}}_{it} \Longrightarrow Y_{it}(d^L) \perp\!\!\!\perp \mathbf{Y}_{\mathcal{N}_i,t-1} \mid U_{it}, \overline{\mathbf{X}}_{it}, \overline{\mathbf{C}}_{it}^B$$

Lemma 5 Under Assumption 3,

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

Proof of the theorem Based on Lemma 5 and Assumption 3,

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

Therefore,

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

$$= \int \{\mathbb{E}[Y_{it} \mid D_{it} = d^H, \overline{\mathbf{X}}_{it}, \overline{\mathbf{C}}^B_{it}] - \mathbb{E}[Y_{it} \mid D_{it} = d^L, \overline{\mathbf{X}}_{it}, \overline{\mathbf{C}}^B_{it}]\} dF_{\overline{\mathbf{X}}_{it}, \overline{\mathbf{C}}^B_{it} | D_{it}=d^H}(\overline{\mathbf{x}}, \overline{\mathbf{c}})$$
$$- \int \{\mathbb{E}[Y_{i,t-1} \mid D_{it} = d^H, \overline{\mathbf{X}}_{i,t-1}, \overline{\mathbf{C}}^B_{it}] - \mathbb{E}[Y_{i,t-1} \mid D_{it} = d^L, \overline{\mathbf{X}}_{i,t-1}, \overline{\mathbf{C}}^B_{it}]\} dF_{\overline{\mathbf{X}}_{it}, \overline{\mathbf{C}}^B_{it} | D_{it}=d^H}(\overline{\mathbf{x}}, \overline{\mathbf{c}}).$$

This completes the proof of Theorem 2 in cases where U_{it} is time-dependent and affected by the outcome at time t. In Section A.3.3, we extend results to two other cases (1) when U_{it} is time-dependent but is not affected by the outcome at time t and (2) when unobserved confounder is time-independent Z_i. □

A.3.1 Proof of Lemma 4

If we write out control set $\overline{\mathbf{C}}$, the lemma can be rewritten as

$$Y_{it}(d^L) \perp\!\!\!\perp \mathbf{Y}_{\mathcal{N}_i,t-1} \mid U_{it}, \overline{\mathbf{X}}_{it}, \overline{\mathbf{X}^*}_{it}, \widetilde{\mathbf{X}}_i$$
$$\implies Y_{it}(d^L) \perp\!\!\!\perp \mathbf{Y}_{\mathcal{N}_i,t-1} \mid U_{it}, \overline{\mathbf{X}}_{it}, \overline{\mathbf{X}^*}_{it}, \overline{\mathbf{X}^*}_{i,t-1}, \widetilde{\mathbf{X}}_i, \mathbf{Y}_{\mathcal{N}_i,t-2}.$$

First, note that all variables in set $\{U_{it}, \overline{\mathbf{X}}_{it}, \overline{\mathbf{X}^*}_{it}, \overline{\mathbf{X}^*}_{i,t-1}, \widetilde{\mathbf{X}}_i, \mathbf{Y}_{\mathcal{N}_i,t-2}\}$ are neither affected by the potential outcome, $Y_{it}(d^L)$, nor affected by the treatment $\mathbf{Y}_{\mathcal{N}_i,t-1}$. The difference between the conditioning sets in the right- and left-hand sides is $\overline{\mathbf{X}^*}_{i,t-1}$ and $\mathbf{Y}_{\mathcal{N}_i,t-2}$. 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 $\overline{\mathbf{X}^*}_{i,t-1}$, $\overline{\mathbf{X}^*}_{it}$, is in the conditioning set, it is contradictory if conditioning on $\overline{\mathbf{X}^*}_{i,t-1}$ can open a new back-door path. Additionally, because $\mathbf{Y}_{\mathcal{N}_i,t-2}$ is a parent of the treatment $\mathbf{Y}_{\mathcal{N}_i,t-1}$, it is contradictory if conditioning on $\mathbf{Y}_{\mathcal{N}_i,t-2}$ can open a new back-door path. Therefore, including $\overline{\mathbf{X}^*}_{i,t-1}$ and $\mathbf{Y}_{\mathcal{N}_i,t-2}$ don't open any back-door path, which completes the proof. □

A.3.2 Proof of Lemma 5

Under Assumption 3,

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

Now we analyze each side of the equation.

$$\int_{\mathcal{C}} \{\mathbb{E}[Y_{it}(d^L)|U_{it} = u_1, \overline{\mathbf{X}}_{it} = \overline{\mathbf{x}}, \overline{\mathbf{C}}^B_{it} = \overline{\mathbf{c}}] - \mathbb{E}[Y_{it}(d^L)|U_{it} = u_0, \overline{\mathbf{X}}_{it} = \overline{\mathbf{x}}, \overline{\mathbf{C}}^B_{it} = \overline{\mathbf{c}}]\}$$

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

where the first equality follows from the fact that $\mathbb{E}[Y_{it}(d^L)|U_{it}=u_0,\overline{\mathbf{X}}_{it}=\overline{\mathbf{x}},\overline{\mathbf{C}}^B_{it}=\overline{\mathbf{c}}]$ does not include u_1, the second equality comes from Lemma 4, and the final from the rule of conditional expectations. Similarly,

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

Taken together,

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

□

A.3.3 Other cases

In Theorem 2, we consider cases in which U_{it} is time-dependent and affected by the outcome at time t. Now we study two other cases (1) when U_{it} 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 $\widetilde{U}_i \equiv (U_{it}, U_{i,t-1})$ 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 $u_1, u_0, \overline{\mathbf{x}}$ and $\overline{\mathbf{c}}$,

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

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

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

A.4 Extensions

A.4.1 Sensitivity Analysis

As Lemma 5 shows, Assumption 3 is equivalent to the following equality.

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

which substantively means the time-invariant bias. However, this assumption might hold only approximately in applied settings. To assess the robustness of the bias-corrected estimates, we consider a sensitivity analysis. In particular, we introduce sensitivity parameter λ as follows.

$$\frac{\mathrm{B}_t(\overline{\mathbf{x}}, \overline{\mathbf{c}})}{\mathrm{B}_{t-1}(\overline{\mathbf{x}}, \overline{\mathbf{c}})} = \lambda$$

where

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

The time-invariance assumption (Assumption 3) corresponds to $\lambda = 1$. Using this sensitivity parameter, we can re-define the bias-corrected estimator as follows.

$$\hat{\tau}_{\text{Main}} - \lambda \times \hat{\delta}_{\text{Placebo}}$$

Therefore, a sensitivity analysis is to compute the bias-corrected estimator for a range of plausible values of λ and investigate whether substantive conclusions vary according to the choice of the sensitivity parameter.

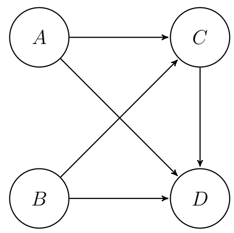
B Causal Directed Acyclic Graphs: Review

In the paper, we use a causal directed acyclic graph and nonparametric structural equations to represent causal relationships. Here, we review basic definitions and results. See Pearl (2000) for a comprehensive review. Following Pearl (1995), we 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 A1 (a), four random variables (A, B, C, D) are given by nonparametric structural equations in Figure A1 (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, 2000). 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.

C Example of Structural Stationarity

Structural stationarity is satisified in a more general NPSEM than the example in the main text. First, variables can be affected not only by one-time lag but also by longer-time lags. For example, outcome Y_{it} can be affected not only by the neighbors' outcomes at the last period $\mathbf{Y}_{\mathcal{N}_i,t-1}$ but also by the neighbors' outcomes at two periods before $\mathbf{Y}_{\mathcal{N}_i,t-2}$. Second,



(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 A1: An Example of Causal DAGs and SEMs

each variable can be not only affected by other variables within each unit but also by other variables of neighbors. For example, outcome Y_{it} can be affected by $\mathbf{L}_{\mathcal{N}_i,t-1}$ and $\mathbf{U}_{\mathcal{N}_i,t-1}$.

We now consider an example that incorporates more complex feedback between variables across time and neighbors. For $i \in \{1, \ldots, n\}$ and $t \in \{1, \ldots, T\}$, suppose the data are generated by sequentially evaluating the following set of equations:

$$
\begin{aligned}
&\text{(Outcome variable)}\\
&\quad Y_{it} = f_Y(\mathbf{Y}_{\mathcal{N}_i,t-1}, \mathbf{Y}_{\mathcal{N}_i,t-2}, Y_{i,t-1}, \mathbf{L}_{it}, \mathbf{L}_{\mathcal{N}_i,t-1}, \widetilde{\mathbf{L}}_i, \mathbf{U}_{it}, \mathbf{U}_{\mathcal{N}_i,t-1}, \epsilon_{it}^Y),\\
&\text{(Time-varying Observed variables)}\\
&\quad \mathbf{L}_{it} = f_L(\mathbf{L}_{i,t-1}, \mathbf{L}_{\mathcal{N}_i,t-1}, \widetilde{\mathbf{L}}_i, Y_{i,t-1}, \mathbf{Y}_{\mathcal{N}_i,t-2}, \mathbf{U}_{i,t-1}, \mathbf{U}_{\mathcal{N}_i,t-2}, \boldsymbol{\epsilon}_{it}^L),\\
&\text{(Time-invariant Observed variables)}\\
&\quad \widetilde{\mathbf{L}}_i = f_{\widetilde{L}}(\mathbf{L}_{i,0}, \mathbf{L}_{\mathcal{N}_i,0}, Y_{i,0}, \mathbf{Y}_{\mathcal{N}_i,0}, \mathbf{U}_{i,0}, \mathbf{U}_{\mathcal{N}_i,0}, \boldsymbol{\epsilon}_i^{\widetilde{L}}),\\
&\text{(Time-varying Unobserved variables)}\\
&\quad \mathbf{U}_{it} = f_U(\mathbf{U}_{i,t-1}, \mathbf{U}_{\mathcal{N}_i,t-1}, Y_{i,t-1}, \mathbf{Y}_{\mathcal{N}_i,t-2}, \mathbf{L}_{i,t-1}, \mathbf{L}_{\mathcal{N}_i,t-2}, \widetilde{\mathbf{L}}_i, \boldsymbol{\epsilon}_{it}^U).
\end{aligned}
\tag{A1}
$$

Several points are worth noting. First, variables can be affected not only by one-time lag but also by longer-time lags. For example, outcome Y_{it} is affected not only by the neighbors' outcomes at the last period $\mathbf{Y}_{\mathcal{N}_i,t-1}$ but also by the neighbors' outcomes at two periods before $\mathbf{Y}_{\mathcal{N}_i,t-2}$. While we do not restrict the number of time-lags and allow for higher-order temporal dependence, we keep our focus on the ACDE defined in equation (1) as our causal estimand. Second, each variable is not only affected by other variables within each unit but also by other variables of neighbors. For example, outcome Y_{it} is affected by $\mathbf{L}_{\mathcal{N}_i,t-1}$ and $\mathbf{U}_{\mathcal{N}_i,t-1}$. Time-vaying unmeasured variables $\mathbf{U}_{it}$ is affected by $\mathbf{U}_{\mathcal{N}_i,t-1}$, $\mathbf{Y}_{\mathcal{N}_i,t-2}$, and $\mathbf{L}_{\mathcal{N}_i,t-2}$. Even though the complexity of the NPSEMs are different in equations (5) and (A1), they both satisfy structural stationarity.

D Simulation Study

In this section, we consider the performance of the proposed placebo test and bias-corrected estimator in a simulation study calibrated to the real hate crime data. In Section D.1, we show that (1) a placebo estimator is consistent for zero under the no omitted confounders assumption as Theorem 1 implies and (2) the statistical power of the proposed placebo test is comparable to an "oracle" test — test whether an estimated ACDE is statistically distinguishable from the true ACDE, which is available only in simulations. In Section D.2, we demonstrate that the bias-corrected estimator reduces bias and root mean squared error (RMSE) even under a slight violation of the time-invariance assumption (Assumption 3).

Setup. To approximate realistic data generating processes, we use the same hate crime data as in the main application but focus on another important outcome, the number of attacks against refugee housing, which is also an important aspect of hate crimes studied in the literature. As for observed covariates, we include five major contextual variables; the number of refugees, the number of crimes per 100,000 inhabitants, per capita income, the unemployment rate, and the share of school leavers without lower secondary education graduation. We fit a linear regression with these five covariates, as in equation (9), to estimate the basic parameters of the data generating process.

We simulate a distance matrix $\mathbf{W}$ based on the stochastic block model (Holland *et al.*, 1983) for each of the sample size $n \in \{100, 500, 1000, 2000\}$. Each group consists of ten units and there exist $K = n/10$ groups. K groups are divided into $L = K/5$ blocks. If units i and j are within the same group, $\Pr(W_{ij} = 1) = 0.8$. If units i and j are within the same block but not in the same group, $\Pr(W_{ij} = 1) = 0.2$. If units i and j are in different blocks, $\Pr(W_{ij} = 1) = 0$. This setup is designed to ensure that the network dependency does not keep growing as the sample size grows. See Sävje *et al.* (2017) and Ogburn *et al.* (2017) for general discussions on network asymptotics.

We then simulate an unobserved contextual variable U_{it}. In particular, we consider two scenarios; (1) time-invariant confounding where assumptions for both the placebo test and the bias-corrected estimator hold, and (2) structural stationarity where assumptions hold for the placebo test but the time-invariance assumption required for the bias-correction is violated. For the first scenario, we set unobserved contextual variable U to be time-invariant where $U_i = \widetilde{U}_{k[i]}$ where $\tilde{U}_k \sim \mathcal{N}(0, 0.5)$ and $k[i]$ is a group indicator for unit i. For the second scenario, we draw unobserved contextual variable U as follows. $U_{it} = \widetilde{U}_{k[i],t}$ where $U_{k,t} = 0.9U_{k,t-1} + \mathcal{N}(0, 0.1)$ where $U_{k0} \sim \mathcal{N}(0, 0.5)$.

Given this setup, we sample potential outcomes using the following data generating process.

$$Y_{i,t+1}(D_{it}) = \alpha + \tau D_{it} + \overline{\mathbf{X}}_{i,t+1}^{\top}\beta + \gamma U_{i,t+1} + \epsilon_{i,t+1}, \tag{A2}$$

for sample size in each time period $n \in \{100, 500, 1000, 2000\}$ and the total number of time periods $T = 20$. $D_{it} \equiv \mathbf{W}_i^{\top}\mathbf{Y}_t$ indicates the treatment variable, five-dimensional vector $\overline{\mathbf{X}}_{i,t+1}$ represents five observed covariates from the real hate crime data, $U_{i,t+1}$ is the unobserved contextual confounder affecting multiple units, and the error term $\epsilon_{i,t+1}$ follows the normal distribution, $\epsilon_{i,t+1} \sim \mathcal{N}(0, 0.1)$. Coefficients $\{\alpha = 0.59, \tau = 0.74, \beta = (0.75, -0.11, -0.28, -3.38, 3.90)\}$ are based on estimated parameters from the real hate crime data. For the effect of unobserved contextual confounder U, we consider two different values $\gamma \in \{0.05, 0.1\}$ in Appendix D.1, and we set it to larger unmeasured confounding $\gamma = 0.1$ in Appendix D.2. Based on this data generating process, we conduct 5000 independent Monte Carlo simulations.

D.1 Placebo Test

First, we consider the consistency of the proposed placebo test under the no omitted confounders assumption. Theorem 1 implies that when the no omitted confounders assumption holds, the treatment variable and the lagged dependent variable are conditionally independent. In particular, we fit a placebo regression:

$$Y_{it} = \alpha_0 + \delta D_{it} + \tau_0 D_{i,t-1} + \overline{\mathbf{X}}_{it}^{\top}\beta_0 + \gamma_0 U_{it} + \epsilon_{it}. \tag{A3}$$

We expect that a test statistic $\widehat{\delta}$ is consistent for zero under the no omitted confounders assumption. The first row in Figure A2 presents the results. As Theorem 1 shows, under the no omitted confounders assumption, the placebo estimator $\widehat{\delta}$ converges to zero as the sample size grows. Because Theorem 1 only requires structural stationarity, the placebo test is consistent under both scenarios.

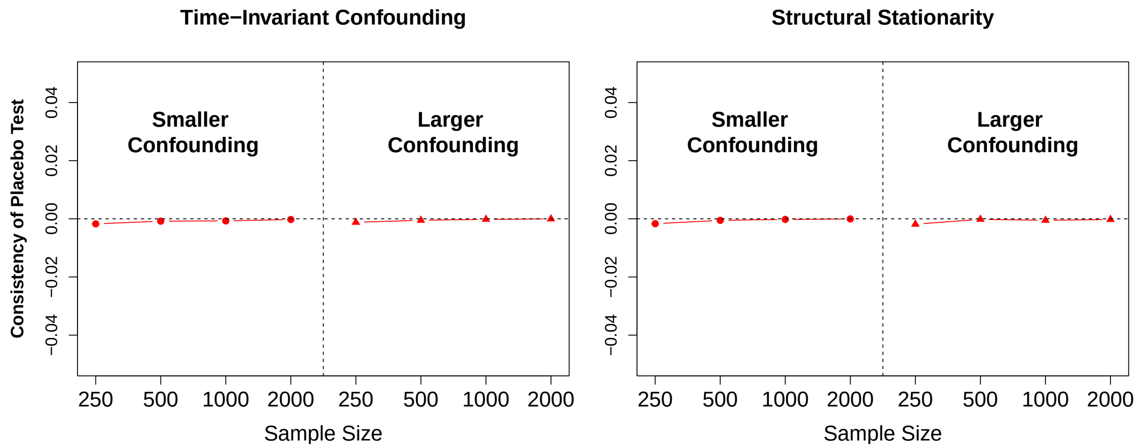


Figure A2: Simulation Results on Consistency of the Placebo Test under the No Omitted Confounders Assumption. *Note:* The Smaller and Larger Confounding corresponds to $\gamma = \{0.05, 0.1\}$, respectively. Results are based on 5000 Monte Carlo draws using four sample sizes.

We also investigate the statistical power of the proposed placebo test when the no omitted confounders assumption is violated. We fit a placebo regression:

$$Y_{it} = \tilde{\alpha}_0 + \tilde{\delta} D_{it} + \tilde{\tau}_0 D_{i,t-1} + \overline{\mathbf{X}}_{it}^{\top} \tilde{\beta}_0 + \tilde{\epsilon}_{it}. \quad \text{(A4)}$$

The key difference is that this regression now ignores contextual confounder U_{it}. Here, $\widehat{\tilde{\delta}}$ serves as a test statistic for the placebo test. We compare this to an oracle test where we fit the following main linear regression,

$$Y_{i,t+1} = \alpha_m + \tau_m D_{it} + \overline{\mathbf{X}}_{i,t+1}^{\top} \beta_m + \xi_{i,t+1}, \quad \text{(A5)}$$

and test $H_0 : \tau_m = \tau$. This test is an "oracle" test because it is available only in the simulation where we know the true ACDE τ. Figure A3 presents the results.

Three findings are worth noting. First, when unmeasured confounding is smaller, it is naturally harder to detect bias (and statistical power of the proposed test is lower). Importantly, however, estimated causal effects are also closer to the true causal effects, and thus, statistical power of the "oracle" test is also lower. Second, when unmeasured confounding is larger, statistical power of the proposed placebo test is closer to that of the oracle test, and the proposed test eventually achieves the same level of power with large enough sample size. On average, statistical power of the proposed test is about 80% of statistical power of the "oracle" test. Finally, as we might expect, it is easier for the proposed method to detect time-invariant unmeasured confounding, but the proposed placebo test can still properly detect time-varying

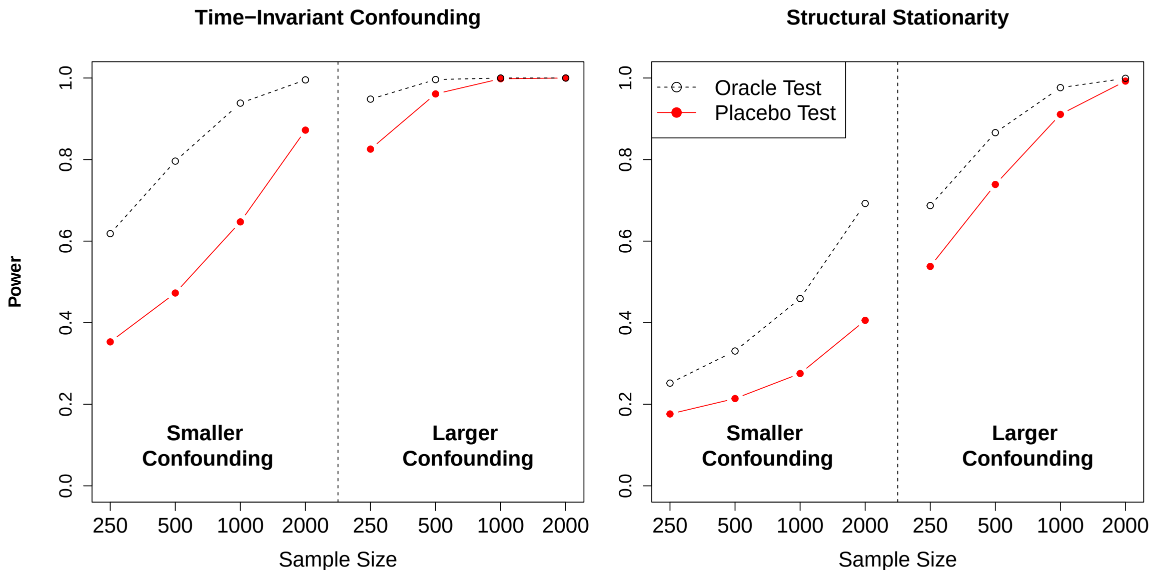


Figure A3: Simulation Results on Statistical Power of Placebo Test. *Note:* The Smaller and Larger Confounding corresponds to $\gamma = \{0.05, 0.1\}$, respectively. Results are based on 5000 Monte Carlo draws using four sample sizes.

unmeasured confounding as long as structural stationarity holds. Given that the oracle test is available only in simulations where the true causal effect is known, these results suggest that the placebo test can serve as a practical tool to detect biases in applied settings.

D.2 Bias-Corrected Estimator

In Section 4.3, we show that the proposed bias-corrected estimator can identify the ACDE for the treated under Assumption 3. Here, we investigate how much the bias-corrected estimator can reduce bias and RMSE even in settings where this required time-invariance assumption is slightly violated.

In particular, we compare an uncorrected estimator, which ignores unobserved contextual confounder U, and the proposed bias-corrected estimator under two scenarios; (1) time-invariant confounding and (2) structural stationarity. The time-invariance assumption required for the bias correction (Assumption 3) holds in the first but not in the second scenario.

Figure A4 presents the simulation results. In the time-invariant confounding case (the first column), whereas the bias in the conventional uncorrected estimator is about 0.12, the bias in the proposed bias-corrected estimator is essentially 0. The bias is corrected as Theorem 2 implies. The RMSE also significantly improves upon the uncorrected conventional estimator. The 95% confidence interval is close to its nominal coverage rate in contrast to that of the uncorrected estimator.

More importantly, even in structural stationarity case (the second column in Figure A4) where the required assumption for the bias correction is slightly violated, the bias-corrected estimator shows reasonable performance. While the bias in the conventional uncorrected estimator is about 0.04, the bias in the proposed bias-corrected estimator is less than 0.01. Although the bias does not vanish, it reduces by about 80%. This benefit is also clear in the results of RMSE. Because the bias-corrected estimator tends to have a larger standard error, the RMSE of the bias-corrected estimator is bigger than the one of the uncorrected estimator when the sample size is small. However, as the sample size grows, the bias-corrected estimator outperforms the uncorrected estimator. Finally, as the required time-invariance assumption is violated, the coverage of the 95% confidence interval for the bias-corrected estimator is slightly smaller than its nominal coverage rate, but it attains more than 90% in contrast to the performance of the uncorrected estimator. These results suggest that the proposed bias-corrected estimator can reduce bias and RMSE in applied settings where the necessary assumption might hold only approximately.

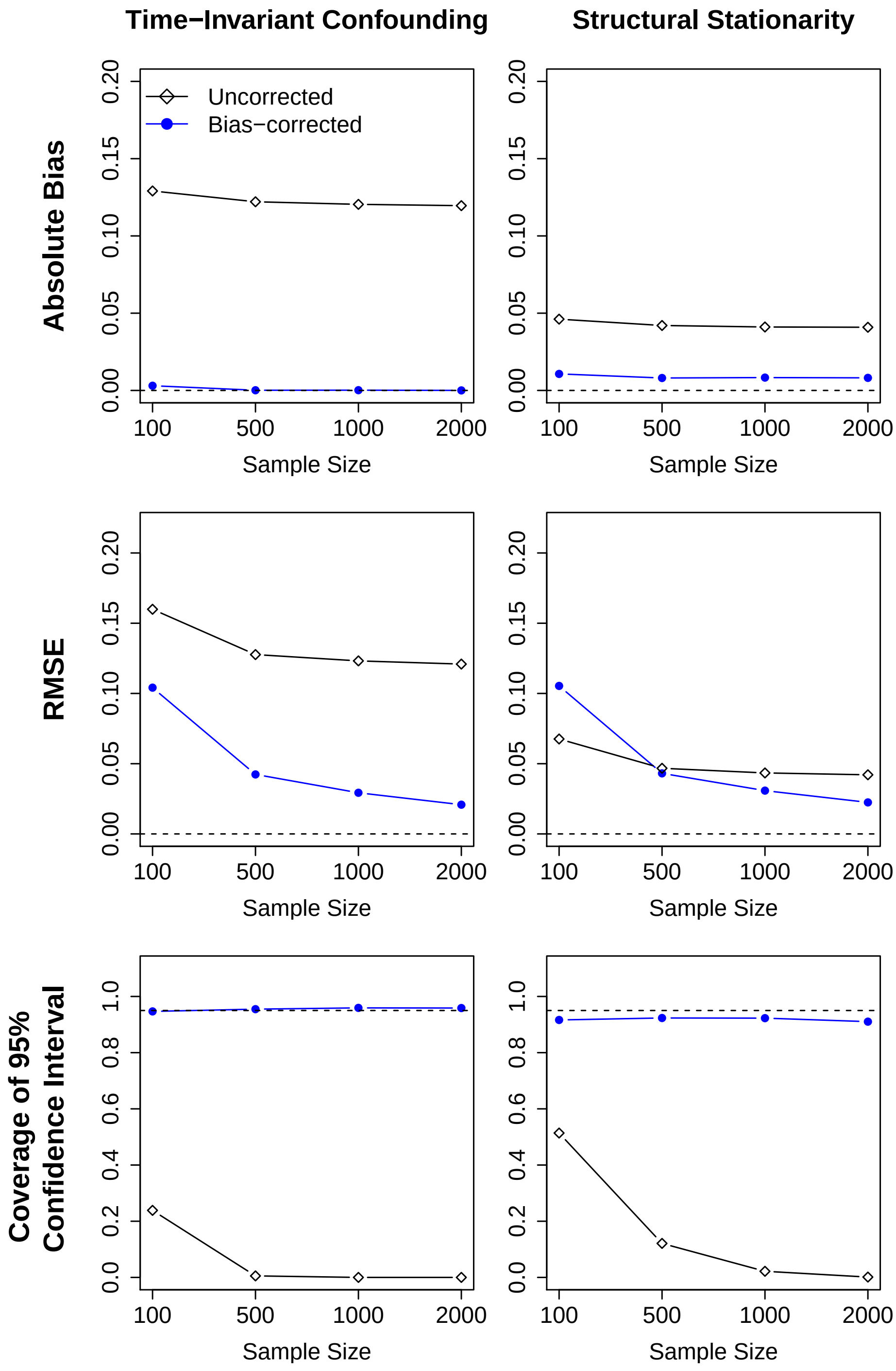


Figure A4: Simulation Results on Bias-Corrected Estimator. *Note:* The first row compares the absolute bias of the uncorrected estimator (empty black square) and the bias-corrected estimator (solid blue circle). The second row examines the root mean squared error (RMSE) and the third row shows the coverage of the 95% confidence interval. The first and second columns correspond to the time-invariant confounding and structural stationarity, respectively. Results are based on 5000 Monte Carlo draws using four sample sizes.

E Empirical Analysis in Section 5

E.1 Control Sets and Placebo Sets

We investigate five different control sets to illustrate how to use the proposed placebo test and bias-corrected estimator. Table A1 describes types of variables we 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.

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.

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	Share of school leavers
	without lower secondary education graduation	without lower secondary education graduation
Political variables	Turnout rate in 2013	Turnout rate in 2013
	Vote share of extreme right and	Vote share of extreme right and
	populist right-wing parties in 2013	populist right-wing parties in 2013

Table A1: Five Control Sets and Placebo Sets: Spatial Diffusion of Hate Crimes.

E.2 Conditional ACDEs by Education

We 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 them (Figure A5), we estimate the conditional ACDE by proportions of school dropouts, separately for the East and the West.

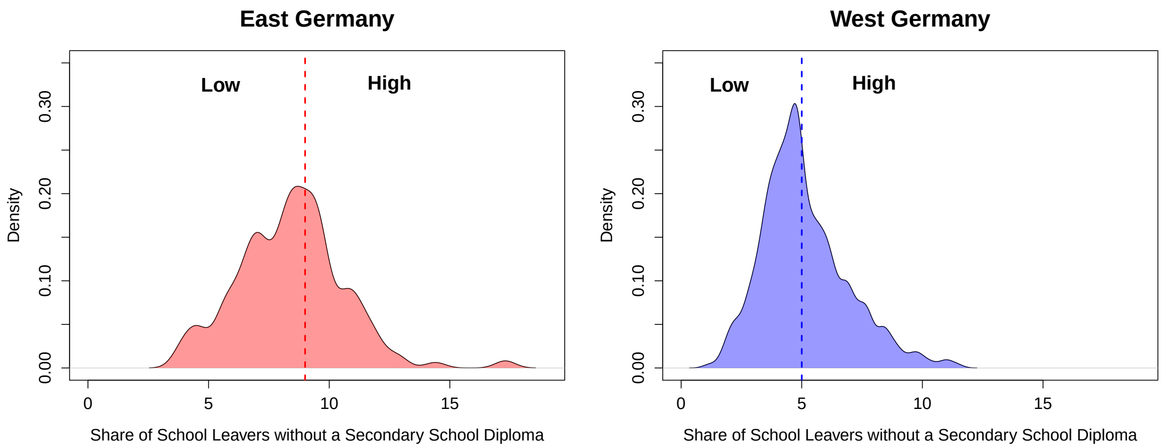
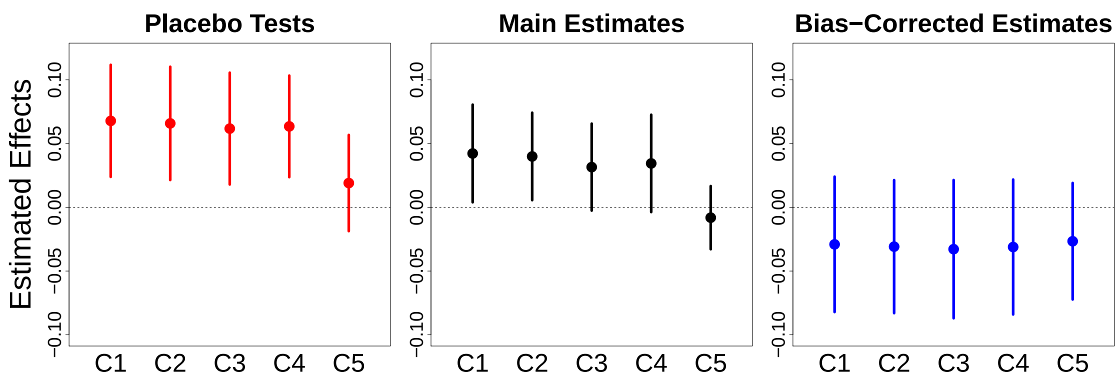


Figure A5: Distribution of Proportions of School Dropouts. Note: For East Germany, we use 9% as a cutoff for high and low proportions of school dropouts, which is approximately the median value in East Germany. For West Germany, we use 5% as a cutoff for high and low proportions of school dropouts, which is approximately the median value in West Germany.

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



(a) (b) (c)

Figure A6: 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, we 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.

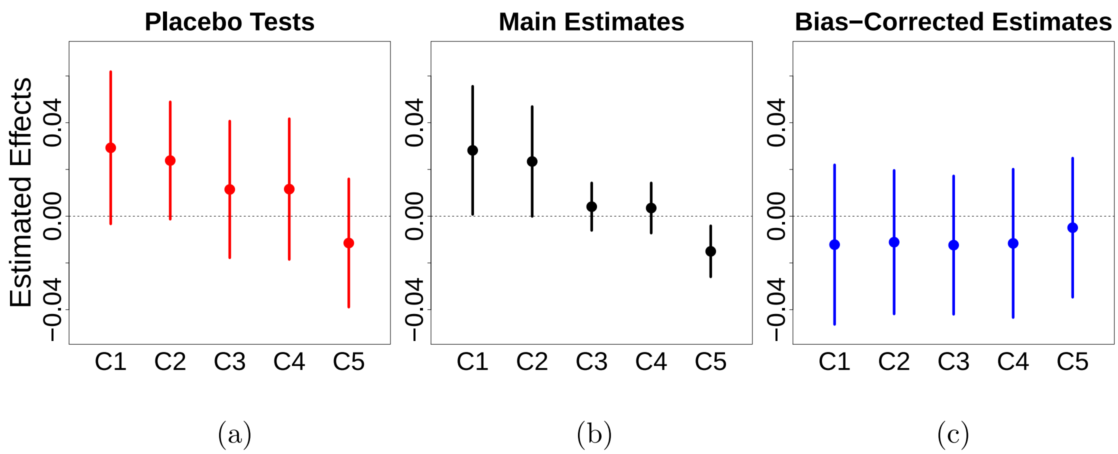


Figure A7: Results of the conditional ACDE (High Proportion of School Dropouts, West). Note: Figure (a) shows that the third, fourth and fifth sets produce small placebo estimates. Focusing on these sets, point estimates of the ACDE in Figure (b) are close to zero and sometimes negative. Figure (c) shows that bias-corrected estimates are similar regardless of the selection of control variables and all of their 95% confidence intervals cover zero.

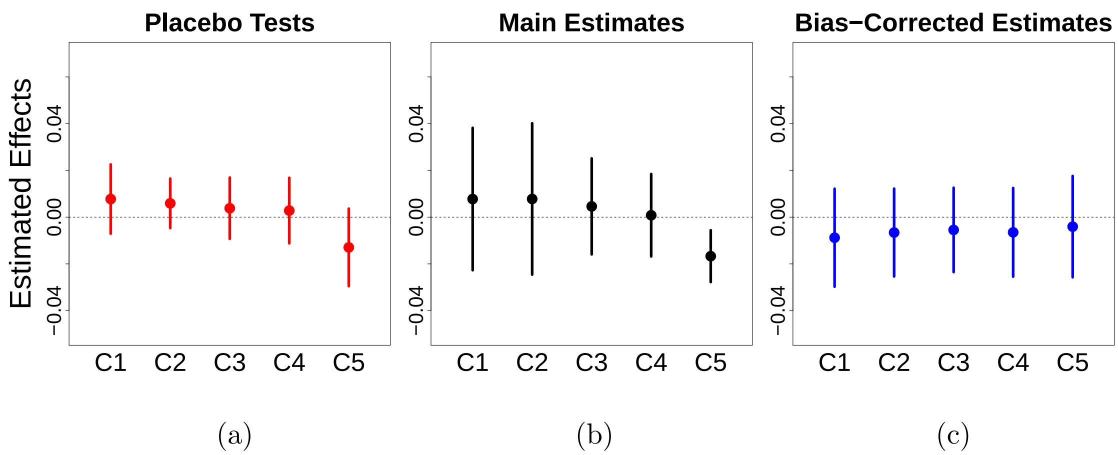


Figure A8: Results of the conditional ACDE (Low Proportion of School Dropouts, West). Note: Figure (a) shows that all the sets produce small placebo estimates. This is partly because there are few hate crimes in this area and hence, there is no variation in outcomes and treatments. In addition, point 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.

F Example Code

```
# ##################
# Placebo Tests
# #################
model.s <- glm(cbind(Physical.bin_lag1, 1- Physical.bin_lag1) ~
                 PhysicalW_lag1 + PhysicalW_lag2 +
                 Physical.bin_lag2 + Physical.bin_lag3 +
                 nei.num + nei.var +
                 log_Total.Ref + log_forgn_total_all + log_pop_mf_total_total.2015 +
                 log_Crime.num_cases_100thInhab + Crime.perc_cases_solved +
                 log_bus_reg_total.2015 + log_insolv_total.2015 +
                 log_vek_percap.2015 + log_empl_ss_resid_total.2016 + log_unemplr_wa18.2015 +
                 rechte_kreis_anteil + school.leave +
                 wahlbeteiligung_kreis_anteil+ maenner_anteil_kreis +
                 bs(Month.num) + factor(ags_state),
               data=hc_data, family=binomial)
vcov_s <- sHAC(fit = model.s, K_base = K_base, space_cut = 100, time_cut = 6)
pl.M <- EstimateQOI_space_one(fit = model.s, vcov = vcov_s, data = hc_data)

# ####################
# Main Estimates
# ####################
model.M <- glm(cbind(Physical.bin, 1- Physical.bin) ~
                  PhysicalW_lag1 +
                  PhysicalW_lag2 +
                  Physical.bin_lag1 + Physical.bin_lag2 +
                  nei.num + nei.var +
                  log_Total.Ref + log_forgn_total_all + log_pop_mf_total_total.2015 +
                  log_Crime.num_cases_100thInhab + Crime.perc_cases_solved +
                  log_bus_reg_total.2015 + log_insolv_total.2015 +
                  log_vek_percap.2015 + log_empl_ss_resid_total.2016 + log_unemplr_wa18.2015
                  rechte_kreis_anteil + school.leave +
                  wahlbeteiligung_kreis_anteil+ maenner_anteil_kreis +
                  bs(Month.num) + factor(ags_state),
                data=hc_data, family=binomial)
vcov_M <- sHAC(fit = model.M, K_base = K_base, space_cut = 100, time_cut = 6)
model.M <- EstimateQOI_space_one(fit = model.M, vcov = vcov_M, data = hc_data)

# ###################
# Bias Correction
# ##################
model.B <- glm(cbind(Physical.bin, 1- Physical.bin) ~
                  PhysicalW_lag1 + PhysicalW_lag2 + PhysicalW_lag3 +
                  Physical.bin_lag1 + Physical.bin_lag2 +
                  nei.num + nei.var +
                  Total.Ref + forgn_total_all + pop_mf_total_total.2015 +
                  Crime.num_cases_100thInhab + Crime.perc_cases_solved +
                  log_bus_reg_total.2015 + log_insolv_total.2015 +
```

```
                  vek_percap.2015 + log_empl_ss_resid_total.2016 +  unemplr_wa18.2015 +
                  rechte_kreis_anteil + school.leave +
                  wahlbeteiligung_kreis_anteil+ maenner_anteil_kreis +
                  factor(ags_state) + bs(Month.num),
                data=hc_data, family=binomial)

pl.B <- glm(cbind(Physical.bin_lag1, 1- Physical.bin_lag1) ~  PhysicalW_lag1 +
               PhysicalW_lag1 + PhysicalW_lag2 + PhysicalW_lag3 +
               Physical.bin_lag2 + Physical.bin_lag3 +
               nei.num + nei.var +
               Total.Ref + forgn_total_all + pop_mf_total_total.2015 +
               Crime.num_cases_100thInhab + Crime.perc_cases_solved +
               log_bus_reg_total.2015 + log_insolv_total.2015 +
               vek_percap.2015 + log_empl_ss_resid_total.2016 +  unemplr_wa18.2015 +
               rechte_kreis_anteil + school.leave +
               wahlbeteiligung_kreis_anteil+ maenner_anteil_kreis +
               factor(ags_state) + bs(Month.num),
             data=hc_data, family=binomial)

vcov_main_B <- sHAC(fit = model.B, K_base = K_base, space_cut = 100, time_cut = 6)
vcov_pl_B <- sHAC(fit = pl.B, K_base = K_base, space_cut = 100, time_cut = 6)
BC_est <- EstimateQOI_space_BC(fit_main = model.B, vcov_main = vcov_main_B,
                               fit_pl = pl.B, vcov_pl = vcov_pl_B,
                               data = hc_data)

## ##################
## R Functions
## ##################

# Estimating Spacial HAC Variance
sHAC <- function(fit, K_base,
                 space_cut = 100, time_cut = 6,
                 space_W_use, time_W_use){

  # create K_mat
  if(missing(space_W_use)){
    space_W <- K_base$space_W
    # space_cut
    space_W_use0 <- space_W/space_cut
    space_W_use0[space_W_use0 > 1] <- 1
    space_W_use <- 1 - space_W_use0
  }else{
    space_W_use <- space_W_use
  }
  if(missing(time_W_use)){
    time_W <- K_base$time_W
    # time_cut
    time_W_use0 <- time_W/time_cut
```

```
    time_W_use0[time_W_use0 > 1] <- 1
    time_W_use <- 1 - time_W_use0
  }else{
    time_W_use <- time_W_use
  }

  K_mat <- space_W_use*time_W_use
  vcov_s <- sHAC0(fit = fit, K_mat = K_mat)

  return(vcov_s)
}

sHAC0 <- function(fit, K_mat){
  X <- model.matrix(fit)
  residual <- fit$y - predict(fit, type = "response")
  score <- as.matrix(residual*X)
  V0 <- (t(score) %*% K_mat %*% score)
  B_inv <- vcov(fit)
  vcov_s <- B_inv %*% V0 %*% B_inv
  return(vcov_s)
}

# Estimating Quantities of Interest
EstimateQOI_space_BC <- function(fit_main, vcov_main,
                                 fit_pl, vcov_pl,
                                 data, treat.var = 0.27, cont.var = 0, seed = 1234){
  set.seed(seed)
  data.d.T <-  data.d.C <- data
  data.d.T$PhysicalW_lag1 <- treat.var
  data.d.C$PhysicalW_lag1 <- cont.var

  formula_main <- formula(fit_main)
  formula_pl <- formula(fit_pl)

  X_pl.T <- model.matrix(formula_pl, data=data.d.T)
  X_pl.C <- model.matrix(formula_pl, data=data.d.C)
  X_m.T <- model.matrix(formula_main, data=data.d.T)
  X_m.C <- model.matrix(formula_main, data=data.d.C)

  # Sample with the new vcov-matrix
  sim.coef_pl <- mvrnorm(n=1000, mu = coef(fit_pl), Sigma= vcov_pl)
  sim.coef_m <- mvrnorm(n=1000, mu = coef(fit_main), Sigma= vcov_main)

  # pl
  X_pl.T.lin <- X_pl.T %*% t(sim.coef_pl)
  X_pl.C.lin <- X_pl.C %*% t(sim.coef_pl)
  pred_pl.T <- inv.logit(X_pl.T.lin)
  pred_pl.C <- inv.logit(X_pl.C.lin)
```

```
  # m
  X_m.T.lin <- X_m.T %*% t(sim.coef_m)
  X_m.C.lin <- X_m.C %*% t(sim.coef_m)
  pred_m.T <- inv.logit(X_m.T.lin)
  pred_m.C <- inv.logit(X_m.C.lin)

  # BC
  qoi_mat <- (pred_m.T - pred_m.C) - (pred_pl.T - pred_pl.C)
  qoi.sim <- apply(qoi_mat, 2, mean)
  qoi.mean <-  mean(qoi.sim)
  qoi.sd <- sd(qoi.sim)
  output <- c("qoi.mean" = qoi.mean, "qoi.sd" = qoi.sd)
  return(output)
}

EstimateQOI_space_one <- function(fit, vcov, data,
                                  treat.var = 0.27, cont.var = 0, seed = 1234){
  set.seed(seed)
  data.d.T <-  data.d.C <- data
  data.d.T$PhysicalW_lag1 <- treat.var
  data.d.C$PhysicalW_lag1 <- cont.var

  formula_use <- formula(fit)

  X.T <- model.matrix(formula_use, data=data.d.T)
  X.C <- model.matrix(formula_use, data=data.d.C)

  # Sample with the new vcov-matrix
  sim.coef <- mvrnorm(n=1000, mu = coef(fit), Sigma = vcov)

  # predict
  X.T.lin <- X.T %*% t(sim.coef)
  X.C.lin <- X.C %*% t(sim.coef)
  pred.T <- inv.logit(X.T.lin)
  pred.C <- inv.logit(X.C.lin)

  # BC
  qoi_mat <- (pred.T - pred.C)
  qoi.sim <- apply(qoi_mat, 2, mean)
  qoi.mean <-  mean(qoi.sim)
  qoi.sd <- sd(qoi.sim)
  output <- c("qoi.mean" = qoi.mean, "qoi.sd" = qoi.sd)
  return(output)
}
```

References

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic Blockmodels: First Steps. *Social Networks*, **5**(2), 109–137.

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*.

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

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge University Press, Cambridge.

Sävje, F., Aronow, P. M., and Hudgens, M. G. (2017). Average Treatment Effects in the Presence of Unknown Interference. *arXiv preprint arXiv:1711.06399*.

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.