# A Negative Correlation Strategy for Bracketing in Difference-in-Differences with Application to the Effect of Voter Identification Laws on Voter Turnout

Ting Ye, Luke Keele, Raiden Hasegawa, and Dylan S. Small

University of Pennsylvania

May 12, 2020

**Abstract**

The method of difference-in-differences (DID) is widely used to study the causal effect of policy interventions in observational studies. DID exploits a before and after comparison of the treated and control units to remove the bias due to time-invariant unmeasured confounders under the parallel trends assumption. Estimates from DID, however, will be biased if the outcomes for the treated and control units evolve differently in the absence of treatment, namely if the parallel trends assumption is violated. We propose a general identification strategy that leverages two groups of control units whose outcomes relative to the treated units exhibit a negative correlation, and achieves partial identification of the average treatment effect for the treated. The identified set is of a union bounds form that previously developed partial identification inference methods do not apply to. We develop a novel bootstrap method to construct uniformly valid confidence intervals for the identified set and parameter of interest when the identified set is of a union bounds form, and we establish the theoretical properties. We develop a simple falsification test and sensitivity analysis. We apply the proposed strategy for bracketing to an application on the effect of voter identification laws in Georgia and Indiana on turnout and find evidence that the laws increased turnout rates.

**Keywords:** bootstrap, difference-in-differences, parallel trends, partial identification, sensitivity analysis, uniform inference

# 1 Introduction

Under the U.S. electoral system, most aspects of election administration are controlled by state governments. As a consequence, state governments oversee both the voter registration process as well as the conduct of voting on election day. State governments regulate key aspects of the voting process such as polling hours on election day, the number of polling stations, whether voters are able to vote by mail, and whether they can register to vote on election day. One key question in political science is whether these regulations have an effect on whether voters decide to vote or not. For example, if states provide a small number of polling locations, it may contribute to long delays in voting and many citizens may be deterred from voting at all. Over the last fifteen years, a number of states have adopted strict voter identification (ID) laws. One area of research in political science has sought to understand whether voter ID laws may affect voter turnout.

## 1.1 Voter Identification Laws

The Help America Vote Act of 2002 primarily focused on improving voting technology in the aftermath of the 2000 presidential election. However, one part of the legislation set minimum voter ID requirements. A number of states subsequently adopted standards that were stricter than the Federal requirements. Some states, such as South Dakota, Michigan, Louisiana, and Hawaii passed laws that allow voting officials to request photo ID from voters, but all voters were not required to show a photo ID in order to vote. Georgia and Indiana passed the first strict photo ID laws that required all voters to provide an acceptable form of photo ID before they can vote. In Georgia, for example, a voter who is unable to show a photo ID is allowed to cast a provisional ballot, but for that ballot to be counted, the voter must travel to the county registrar office within three days of the election to present the required photo ID or provide a list of required documents to obtain a voter ID card from the county registrar office. After a series of legal challenges, voter ID laws in Georgia and Indiana went into effect for the 2008 presidential election. Since that election, many other states have also adopted strict voter ID laws. A number of studies have sought to estimate the effect of voter ID laws and have reported a range of estimates from positive to negative (Barreto et al. 2009; Milyo 2007; Alvarez et al. 2008; Mycoff et al. 2009; Erikson and Minnite 2009; Hajnal et al. 2017; Burden 2018; Hajnal et al. 2018; Barreto et al. 2018; Hood III and Bullock III 2008; Hopkins et al. 2017; Grimmer et al. 2018). These mixed results are most likely due to the fact that causal inference in this setting is a challenging task. In the next section, we review those challenges.

## 1.2 Challenges in Causal Inference for Voter ID Laws

Drawing causal inference about the effect of state laws like voter ID laws is a challenging task for three specific reasons. First, randomized experiments in this setting are nonexistent. While randomized experiments are frequently implemented for education and labor interventions, we know of no such experiments in the context of election administration. Second, natural experiments are also rare as states tend to implement voting laws for all citizens and rarely do natural circumstances produce anything haphazard about the introduction of these laws. Finally, states adopt these laws in a highly purposeful fashion, and therefore exposure to these laws almost certainly depends on unobservable factors.

One common strategy for estimating the effect of state laws on citizen behavior is the method of difference-in-differences (DID). The simplest DID estimate is based on a comparison of the outcome differences for the treated units before and after adopting the treatment and the outcome differences for the control units. More generally, the DID estimate can be obtained using fixed effects regression models and one can adjust for observed variables (Angrist and Pischke 2009, Ch. 5). The key advantage of DID is that it removes time-invariant bias from unobserved confounders. However, the DID method depends on a key assumption that the outcomes in the treated and control units are, in the absence of treatment, evolving in the same way over time. In our context, we must assume that voter turnout in Georgia or Indiana, if these two states had not adopted voter ID laws, would have followed the same dynamics over time as the turnouts in non-voter ID states. This key assumption is often referred to as the parallel trends assumption.

In studies of voter ID laws, bias from state-specific activities would appear to be a non-ignorable threat to the validity of DID methods. In U.S. presidential elections, battleground states (or swing states) are usually targeted by both major-party campaigns, especially in competitive elections. Voters in battleground states are often subject to significant amounts of political advertising and voter mobilization efforts by both the campaigns but also outside groups. For example, political groups may send mail to and knock on doors of voters to increase turnout. Voters in non-battleground states, however, are subject to few presidential campaign activities. States such as Florida, Ohio and Pennsylvania are perennial battleground states, but battleground state status may vary across elections. Alternatively, a key ballot initiative or competitive Senatorial election may also boost voter mobilization efforts in a specific state in a single election.

## 1.3 Our Contributions

Given these key threats to validity, researchers need tools that exploit the strengths of DID, but depend on less stringent assumptions. For example, Abadie (2005) and Callaway and Sant'Anna (2019) assume that the parallel trends assumption holds after conditioning on observed covariates. Athey and Imbens (2006) assume a changes-in-changes model that is general but rules out classic measurement error on the outcome. Daw and Hatfield (2018) and Ryan et al. (2018) focus on matching in DID analyses. Manski and Pepper (2017) and Rambachan and Roth (2019) also consider partial identification in DID settings, but do not exploit the bounding relationship using multiple control groups.

Hasegawa et al. (2019) propose a bracketing method that addresses the bias arising from a historical event interacting with the groups. For example, Indiana is not generally a state with competitive presidential elections. However, in 2008 Barack Obama narrowly won the state, making it the first time a Democrat carried the state since 1964. As such, the first use of voter ID laws in Indiana coincided with a competitive presidential election. Therefore, estimates based on standard DID methods could be biased because the 2008 presidential election may have affected the treated state differently from control states. The bracketing method works by partitioning the control states (states without the voter ID laws) into two groups according to their past levels of the turnout and then constructing two DID estimates using respectively these two control groups. Then, under the assumptions reviewed in Section 2.2, the effect of the 2008 presidential election on the treated state is bounded by its effects on the two control groups. This bracketing method is a partial identification approach in that it produces bounds for the treatment effect–two identifiable parameters that bracket the true causal effect.

In this article, we propose a general strategy for DID bracketing that addresses the concerns about heterogeneity in different units' outcome dynamics. This general strategy includes the original bracketing method (Hasegawa et al. 2019) as a special case, but is more broadly applicable. In a nutshell, we leverage two groups of control units whose outcomes relative to the treated units exhibit a negative correlation, i.e., when the relative outcome for one control group increases, the relative outcome for the other control group decreases. In this case, DID parameters are identifiable based on the two control groups, and can be used to bound (bracket) the average treatment effect for the treated units. We derive the key identification assumption for this general strategy. We show that this new identification assumption accommodates many existing commonly adopted assumptions in the DID literature.

A second contribution of our paper is that we develop inference for a general class of partially

identified parameters that has a union bounds form. Partial identification has a long history in econometrics dating back to the 1920s, see Tamer (2010) for a review. In contrast to the point identification approach, the partial identification approach typically imposes weaker assumptions. This feature is particularly attractive in causal inference because although making assumptions is inevitable, one can now choose to make credible assumptions that can be well-justified from the subject matter, but not other assumptions that have previously been made without strong subject matter justification. As a result, the partial identification approaches are playing an increasingly important role in causal inference (Richardson et al. 2014). For example, partial identification approaches have been developed for ordinal outcomes (Chiba 2017), mediation analysis (Cai et al. 2008; Sjölander 2009; Vanderweele 2011), missing data (Horowitz and Manski 2000; Imai 2008; Yang and Small 2016; Jiang and Ding 2018), and instrumental variable methods (Balke and Pearl 1997; Siddique 2013; Flores and Flores-Lagunes 2013; Swanson et al. 2018). Inference for partially identified parameters has been well studied when the boundary of the identified set has an explicit form and can be consistently estimated (Imbens and Manski 2004; Stoye 2009), or more generally when the identified set can be represented by moment inequalities (Romano and Shaikh 2008; Andrews and Soares 2010; Bugni 2010; Chernozhukov et al. 2013; Bugni et al. 2017; Kaido et al. 2019).

However, the identified set for our proposed bracketing method does not have either an explicit form that can be consistently estimated nor can it be represented by moment inequalities. Instead, the identified set takes a "union bounds" form, namely the bounds can be expressed as the union of several intervals. We develop a novel and easy-to-implement bootstrap method to construct uniformly valid confidence intervals for the identified set and parameters of interest, and we establish key theoretical properties. This new inference method for union bounds can be used in other applications, for example, when bracketing using multiple controls (Campbell, 1969, p. 365; Rosenbaum, 1987*a*).

Finally, we develop a falsification test and sensitivity analysis to probe the identification assumption and evaluate how sensitive the study conclusions are to violations of the assumption.

Our paper proceeds as follows. In Section 2, we introduce notations and our causal framework. We also review the DID method and the bracketing method proposed by Hasegawa et al. (2019). In Section 3, we introduce the general bracketing strategy in DID. We derive the identification assumption and describe how to construct the control groups. In Section 4, we develop a bootstrap inference method and study its theoretical properties. In Section 5, we develop a falsification test and sensitivity analysis. In Section 6, we examine the finite sample empirical

performance of the proposed bootstrap inference method for union bounds in a simulation study. In Section 7, we apply the proposed bracketing methods to study the effect of voter ID laws. In Section 8, we conclude with a discussion. Technical proofs are in the Supplementary Materials.

# 2 Preliminaries

## 2.1 Notation and Causal Framework

We consider applications where data is observed for the treated and control units before and after the treated units adopt the treatment, while the control units are never treated. Suppose the treated units adopt the treatment between two time periods, which we denote as $t = 1$ and $t = 2$, and remain treated afterwards. We will refer to time period $t = 1$ as the pre-treatment period and time periods $t = 2, \ldots, T$ as the post-treatment periods. We write $D_i = 1$ if individual $i$ belongs to the treated units, $D_i = 0$ if individual $i$ belongs to the control units. In our application, Indiana and Georgia enacted the voter ID laws between the 2004 and 2008 presidential elections, and thus, we use $t = 1$ to denote year 2004, $t = 2$ to denote year 2008. We use $D_i = 1$ to represent individual $i$ living in Indiana or Georgia, and $D_i = 0$ to represent individual $i$ living in other states that did not enact the voter ID laws. This represents a common configuration where the units are states, and the treatment is a change in state policy for one or more states. We do not consider the staggered adoption case, where the treatment is adopted by multiple states over time. We leave that case for future work.

As in Neyman (1923) and Rubin (1974), we define treatment effects in terms of potential outcomes. Let $Y_{it}^{(1)}$ represent the potential outcome for individual $i$ at time $t$ if being treated, let $Y_{it}^{(0)}$ represent the potential outcome for individual $i$ at time $t$ if being untreated. We assume throughout the article that at each time $t$, the potential outcomes and the treated unit indicator $(Y_{it}^{(1)}, Y_{it}^{(0)}, D_i), i = 1, \ldots, N_t$, are identically and independently distributed (i.i.d.) realizations of $(Y_t^{(1)}, Y_t^{(0)}, D)$. Relatedly, there have been studies that evaluate the impact of within-cluster correlation arising from a random unit-time specific component (Bertrand et al. 2004; Donald and Lang 2007), see Imbens and Wooldridge (2008, Section 6.5.3) for a review. In this article, we instead view the unit-time specific components as fixed effects and propose to bracket these fixed effects rather than modeling them as random. As such, the unit-time specific components can be accounted for using our bracketing method without creating within-cluster correlation. More specific discussion on this point is given in Section 4 below. Depending on the treatment status, the observed outcomes can be expressed as $Y_{it} = Y_{it}^{(0)}$ for $t \leq 1$, $Y_{it} = D_i Y_{it}^{(1)} + (1 - D_i) Y_{it}^{(0)}$ for $t = 2, \ldots, T$. The observed data $\{Y_{i1}, D_i\}_{i=1,\ldots,N_1}, \ldots, \{Y_{iT}, D_i\}_{i=1,\ldots,N_T}$ can be obtained from

a longitudinal study of the same units over time or a repeated cross-sectional study. Hereafter, we drop the subscript $i$ to simplify the notation.

We are interested in the average treatment effect for the treated units in the post-treatment periods,

$$ATT_t = E[Y_t^{(1)} - Y_t^{(0)}|D = 1], \qquad t = 2, \ldots, T$$

where the expectation is taken with respect to the distribution of $(Y_t^{(1)}, Y_t^{(0)}, D)$. In our application, we focus on two post-treatment time periods–2008 and 2012. We use 2012 as an outcome year to explore the possibility of a delayed treatment effect. We also conduct separate analyses for Indiana and Georgia, and thus identify the average treatment effects for Indiana and Georgia separately. Alternatively, if we group Indiana and Georgia together, we would identify the average treatment effect for the treated population at each post-treatment period, which is a weighted average of the average treatment effect for Indiana and the average treatment effect for Georgia where the weights are the populations for these two states at that time. Note that $E[Y_t^{(1)}|D = 1] = E[Y_t|D = 1]$ for $t = 2, \ldots, T$ can be identified from the observed data, but we never observe $Y_t^{(0)}$ for the treated units in post-treatment periods.

One approach to causal identification is to use the method of Difference-in-Differences (DID). The crucial identifying assumption in DID is that the treated units and the control units would have exhibited parallel trends in the potential outcomes in the absence of treatment. Mathematically, this parallel trends assumption can be expressed as

$$E[Y_t^{(0)} - Y_1^{(0)}|D = 1] = E[Y_t^{(0)} - Y_1^{(0)}|D = 0], \qquad t = 2, \ldots, T. \tag{1}$$

With this assumption, we can leverage the control units to identify the change in the potential outcomes for the treated units had the units counterfactually not been treated. The average treatment effect for the treated units in the post-treatment periods can be identified through

$$\begin{aligned} ATT_t &= E[Y_t^{(1)} - Y_t^{(0)}|D = 1] \\ &= E[Y_t^{(1)} - Y_1^{(0)}|D = 1] - E[Y_t^{(0)} - Y_1^{(0)}|D = 1] \\ &= E[Y_t^{(1)} - Y_1^{(0)}|D = 1] - E[Y_t^{(0)} - Y_1^{(0)}|D = 0] \\ &= E[Y_t - Y_1|D = 1] - E[Y_t - Y_1|D = 0], \end{aligned}$$

where the third equality holds because of the parallel trends assumption in (1). In the simplest scenario, the DID estimator replaces the conditional expectations above with the corresponding

sample averages.

## 2.2 Review: Bracketing in Difference-in-Differences

In extant work, researchers have sought to relax the parallel trends assumption in various ways. The DID bracketing method in Hasegawa et al. (2019) considers the setup with one post-treatment period (i.e., $T = 2$), and works by partitioning the control units into two *groups* (one lower control group and one upper control group), and uses the two standard DID estimators based on theses two control groups to bound the true treatment effect. In this section, we review the DID bracketing method in more detail, as the general identification strategy in Section 3 builds on this section and it is important to compare these two DID bracketing methods.

Let $G$ be a group indicator, where $G = trt$ (equivalent to $D = 1$) denotes the treated group, $G = lc$ denotes the lower control group and $G = uc$ denotes the upper control group. Hasegawa et al. (2019) assume the following model for the potential outcome $Y_t^{(0)}$ that generalizes the standard DID model and changes-in-changes model (Athey and Imbens 2006),

$$Y_t^{(0)} = h(U, t) + \epsilon_t, \tag{2}$$

where $U$ is a time-invariant unmeasured confounder that may have a time-varying effect on the outcome, $\epsilon_t$ is an error term that captures additional sources of variation at time $t$ and $E[\epsilon_t|U, G] = 0$ for every $t$. Critically, $U$ is a time-invariant variable that captures the systematic difference among different groups.

Identification of bounds on the true treatment effect is achieved by imposing assumptions on the distribution of $U$ in different groups and its effect on the outcome and the outcome dynamics.

**Assumption 1.** *The distribution of $U$ within groups is stochastically ordered: $U|G = lc \preccurlyeq U|G = trt \preccurlyeq U|G = uc$.*

Note that two random variables $A, B$ are stochastically ordered $A \preccurlyeq B$ if $E[f(A)] \leq E[f(B)]$ for all bounded non-decreasing functions $f$ (Hadar and Russell 1969). In words, this assumption states that an unmeasured confounder such as measure of political engagement is lowest in the lower control group, intermediate in the treated group, and highest in the upper control group.

**Assumption 2.** *The unspecified function $h(U, t)$ is bounded and increasing in $U$ for every $t$.*

This assumption is natural when a higher level of the unmeasured confounder corresponds to a higher value of the outcome. The other direction is also included in this model because we

can simply replace $U$ with its negation. Model (2) and Assumptions 1-2 combined imply that $E[Y_t^{(0)}|G = lc] \leq E[Y_t^{(0)}|G = trt] \leq E[Y_t^{(0)}|G = uc]$ for every $t$.

**Assumption 3.** *Either one of the following is satisfied: (a)* $h(U,2)-h(U,1) \geq h(U',2)-h(U',1)$ *for all* $U \geq U'$, $U, U' \in$ *supp*$(U)$*; (b)* $h(U,2) - h(U,1) \leq h(U',2) - h(U',1)$ *for all* $U \geq U'$, $U, U' \in$ *supp*$(U)$.

In words, the bounded function $h(U,2) - h(U,1)$ is either non-decreasing in $U$ over the whole support of $U$, or non-increasing in $U$ over the whole support of $U$. Combining Assumptions 1, 3 and the boundedness of $h(U,t)$, we have that $E[h(U,2) - h(U,1)|G = lc] \leq E[h(U,2) - h(U,1)|G = trt] \leq E[h(U,2) - h(U,1)|G = uc]$ or the reverse direction.

Next, we define the two DID parameters as

$$\beta(uc) = E[Y_2 - Y_1|G = trt] - E[Y_2 - Y_1|G = uc] \tag{3}$$

$$\beta(lc) = E[Y_2 - Y_1|G = trt] - E[Y_2 - Y_1|G = lc] \tag{4}$$

where $\beta(uc)$ is the DID parameter using the upper control group, and $\beta(lc)$ is the DID parameter using the lower control group. Under model (2), we can relate $\beta(uc), \beta(lc)$ with the treatment effect of interest $ATT_2$ as follows:

$$\begin{aligned}
\beta(uc) &= ATT_2 + E[Y_2^{(0)} - Y_1^{(0)}|G = trt] - E[Y_2^{(0)} - Y_1^{(0)}|G = uc] \\
&= ATT_2 + E[h(U,2) - h(U,1)|G = trt] - E[h(U,2) - h(U,1)|G = uc] \\
\beta(lc) &= ATT_2 + E[Y_2^{(0)} - Y_1^{(0)}|G = trt] - E[Y_2^{(0)} - Y_1^{(0)}|G = lc] \\
&= ATT_2 + E[h(U,2) - h(U,1)|G = trt] - E[h(U,2) - h(U,1)|G = lc].
\end{aligned}$$

Additionally under Assumptions 1–3, one of the two DID parameters is too large and the other is too small such that we can bound $ATT_2$:

$$\min\{\beta(uc), \beta(lc)\} \leq ATT_2 \leq \max\{\beta(uc), \beta(lc)\}. \tag{5}$$

The DID parameters $\beta(uc), \beta(lc)$ are identifiable from the observed data. For example, one can simply replace the conditional expectations in (3)-(4) with sample analogues to obtain the corresponding DID estimators. As such, the DID bracketing method accounts for the bias arising from the time-varying effect of the unmeasured confounder $U$ by leveraging the connections between the effect of $U$ on the outcome and the effect of $U$ on the outcome dynamics using two

control groups. Facilitated by the control group construction approach discussed in Hasegawa et al. (2019) in which units are designated to the lower (upper) control group if the average outcome is lower (higher) than the average outcome for the treated group in a prior-study period, the average treatment effect for the treated $ATT_2$ is partially identified via (5).

# 3 A General Strategy for Bracketing in DID

The bracketing method in Hasegawa et al. (2019) relies on model (2) and the unmeasured confounder $U$ satisfying Assumptions 1-3 to connect the outcome levels and outcome dynamics in the absence of treatment, such that the changes in outcome for different groups are ordered according to their outcome levels. However, in some applications, if the unmeasured confounder $U$ does not satisfy Assumptions 1-3, and previous outcome levels do not reflect the relative magnitude of changes in outcome in the study period, the control groups constructed following Hasegawa et al. (2019) may fail to produce valid DID brackets. In the context of voter ID laws, some states are battleground states in some elections but not others. Voters in battleground states tend to be exposed to more mobilization efforts by the campaigns, while voters in non-battleground states may be subject to fewer mobilization efforts. In this scenario, levels of turnout may not capture changes in voter turnout due to variation in mobilization efforts across elections. Nonetheless, there may be useful patterns in turnout dynamics that can be exploited to achieve partial identification of the average treatment effect for the treated.

## 3.1 Partial Identification of the treatment effect

In this section, we present our key partial identification assumption based on two control groups which we denote as $a, b$. Let $\Delta_t^{(0)}(g) = E(Y_t^{(0)} - Y_{t-1}^{(0)}|G = g)$ be the expected change in potential outcome for group $g$ from time $t-1$ to time $t$ in the absence of treatment. Let $\Gamma_t^{(0)}(g) = E[Y_t^{(0)}|G = g] - E[Y_t^{(0)}|G = trt]$ be the expected outcome for control group $g$ at time $t$ relative to the treated group in the absence of treatment. The key partial identification assumption is

**Assumption 4.** *(monotone trends) For $t = 2, \ldots, T$,*

$$\min\left\{\Delta_t^{(0)}(a), \Delta_t^{(0)}(b)\right\} \leq \Delta_t^{(0)}(trt) \leq \max\left\{\Delta_t^{(0)}(a), \Delta_t^{(0)}(b)\right\}. \tag{6}$$

After some simple algebra, one can show that $\Delta_t^{(0)}(g) - \Delta_t^{(0)}(trt) = \Gamma_t^{(0)}(g) - \Gamma_{t-1}^{(0)}(g)$. Therefore, Assumption 4 has an equivalent formulation, which we state as Lemma 1.

**Lemma 1.** *Assumption 4 is equivalent to the following: for* $t = 2, \ldots, T$,

$$\left\{\Gamma_t^{(0)}(a) - \Gamma_{t-1}^{(0)}(a)\right\}\left\{\Gamma_t^{(0)}(b) - \Gamma_{t-1}^{(0)}(b)\right\} \leq 0. \tag{7}$$

What does Assumption 4 imply about the behavior of the two control groups relative to the treated group? According to Assumption 4, in every pair of adjacent time periods, the change in outcome for the control group $a$ and the change in outcome for the control group $b$ provide bounds on the change in outcome for the treated group if it were untreated. The equivalent formulation in Lemma 1 provides another interesting and more concrete perspective. That is, the outcomes for the two control groups $a, b$ relative to the treated group are negatively correlated. In other words, if the relative outcome for control group $a$ increases (decreases), the relative outcome for control group $b$ decreases (increases). Figure 1 (a) provides an illustrative example of Assumption 4 and Figure 1 (b) provides an illustrative example of the equivalent formulation as in Lemma 1. In Figure 1 (a), the treated group has the lowest outcome level across all four time periods. However, for every pair of adjacent time periods, the slope for the treated group (i.e., $\Delta_t^{(0)}(trt)$) is bounded by the slopes for the two control groups (i.e., $\Delta_t^{(0)}(a)$ and $\Delta_t^{(0)}(b)$). Specifically, $\Delta_2^{(0)}(b) > \Delta_2^{(0)}(trt) > \Delta_2^{(0)}(a)$, because from $t = 1$ to $t = 2$, the expected outcome for the control group $b$ has a larger increase compared with the treated group and the expected outcome for the control group $a$ decreases, and similarly $\Delta_3^{(0)}(b) > \Delta_3^{(0)}(trt) > \Delta_3^{(0)}(a)$, $\Delta_4^{(0)}(a) > \Delta_4^{(0)}(trt) > \Delta_4^{(0)}(b)$, which implies Assumption 4 is satisfied. Figure 1 (a) is then translated into Figure 1 (b) by plotting the outcomes for the two control groups relative to the treated group (i.e., $\Gamma_t^{(0)}(a)$ and $\Gamma_t^{(0)}(b)$). In Figure 1 (b), for every pair of adjacent time periods, the relative outcome for one control group increases and that for the other control group decreases. Specifically, $\Gamma_2^{(0)}(a) - \Gamma_1^{(0)}(a) < 0, \Gamma_2^{(0)}(b) - \Gamma_1^{(0)}(b) > 0$, so that (7) holds for $t = 2$; $\Gamma_3^{(0)}(a) - \Gamma_2^{(0)}(a) < 0, \Gamma_3^{(0)}(b) - \Gamma_2^{(0)}(b) > 0$, so that (7) holds for $t = 3$; $\Gamma_4^{(0)}(a) - \Gamma_3^{(0)}(a) > 0, \Gamma_4^{(0)}(b) - \Gamma_3^{(0)}(b) < 0$, so that (7) holds for $t = 4$; this implies the equivalent condition in Lemma 1 is satisfied.

Next, we compare Assumption 4 with the assumptions of other similar identification strategies. Compared with the original bracketing method reviewed in Section 2, it is easy to see that Assumption 4 (with $a, b$ being $lc, uc$ and $T = 2$) is implied by the conditions imposed in Hasegawa et al. (2019), so the original bracketing method is a special case of the current general strategy. Moreover, the general strategy no longer requires model (2) or Assumption 2, because we now explicitly impose assumptions on how the outcomes change.

Assumption 4 also accommodates many existing assumptions in the literature. For example,

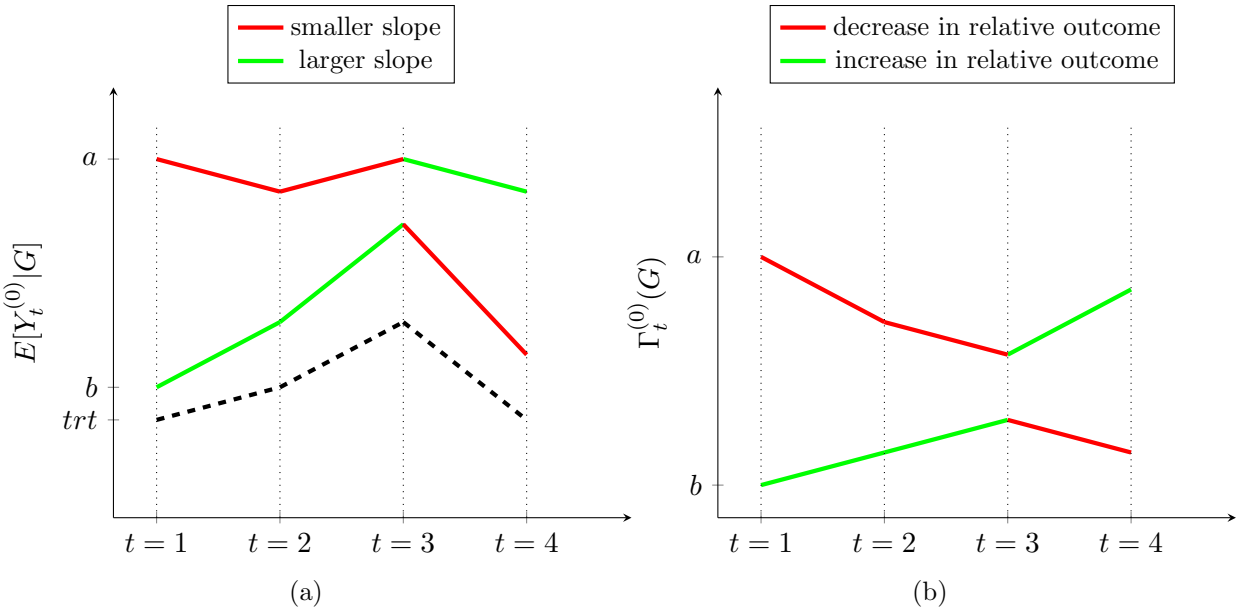


Figure 1: Illustrations of Assumption 4 and Lemma 1. Panel (a) plots the expected outcome for every group, where the slope between every pair of adjacent time periods for the treated group (i.e., $\Delta_t^{(0)}(trt)$) is bounded in between by the slopes for the two control groups (i.e., $\Delta_t^{(0)}(a)$ and $\Delta_t^{(0)}(b)$). Correspondingly, Panel (b) plots the expected outcomes for two control groups relative to the treated group (i.e., $\Gamma_t^{(0)}(a) = E[Y_t^{(0)}|G = a] - E[Y_t^{(0)}|G = trt]$, $\Gamma_t^{(0)}(b) = E[Y_t^{(0)}|G = b] - E[Y_t^{(0)}|G = trt]$), where for every pair of adjacent time periods, the relative outcome for one control group increases and that for the other control group decreases.

the classical DID method requires that the outcome dynamics for every group are the same, i.e., $\Delta_t^{(0)}(a) = \Delta_t^{(0)}(trt) = \Delta_t^{(0)}(b)$ for every $t$, under which Assumption 4 holds. Therefore, the bracketing method is valid under the classical DID assumptions. Assumption 4 also relates to the parallel growth assumption (Mora and Reggio 2012), which requires that $\Delta_t^{(0)}(a) - \Delta_{t-1}^{(0)}(a) = \Delta_t^{(0)}(trt) - \Delta_{t-1}^{(0)}(trt) = \Delta_t^{(0)}(b) - \Delta_{t-1}^{(0)}(b)$ for every $t$. If we construct the control groups such that $\Delta_0^{(0)}(a) \leq \Delta_0^{(0)}(trt) \leq \Delta_0^{(0)}(b)$, then the parallel growth assumption implies that $\Delta_t^{(0)}(a) \leq \Delta_t^{(0)}(trt) \leq \Delta_t^{(0)}(b)$ for every $t$, and thus also implies that Assumption 4 is true. Assumption 4 is more general than both existing identification assumptions because it does not restrict the outcome dynamics to be the same or the "second derivative" of $Y_t^{(0)}$ to be the same for every group. Hence, the DID bracketing method is valid whenever the classical DID method or the parallel growth method is applicable, but is also valid under much broader range of scenarios.

Recall that we define the average treatment effect for the treated $ATT_t = E[Y_t^{(1)} - Y_t^{(0)}|G = trt]$ for $t = 2, \ldots, T$. For $t = 2$, we can relate the DID parameter using each of the two control

groups to $ATT_2$,

$$\begin{aligned}\tau_2(a) &= E[Y_2 - Y_1|G = trt] - E[Y_2 - Y_1|G = a] = ATT_2 + \Delta_2^{(0)}(trt) - \Delta_2^{(0)}(a), \\ \tau_2(b) &= E[Y_2 - Y_1|G = trt] - E[Y_2 - Y_1|G = b] = ATT_2 + \Delta_2^{(0)}(trt) - \Delta_2^{(0)}(b). \end{aligned} \tag{8}$$

where $\tau_2(a), \tau_2(b)$ are standard DID parameters. For the case where $t > 2$, we define

$$\begin{aligned}\tau_t(a) &= E[Y_t - Y_{t-1}|G = trt] - E[Y_t - Y_{t-1}|G = a] = ATT_t - ATT_{t-1} + \Delta_t^{(0)}(trt) - \Delta_t^{(0)}(a), \\ \tau_t(b) &= E[Y_t - Y_{t-1}|G = trt] - E[Y_t - Y_{t-1}|G = b] = ATT_t - ATT_{t-1} + \Delta_t^{(0)}(trt) - \Delta_t^{(0)}(b), \end{aligned} \tag{9}$$

where $\tau_t(a), \tau_t(b)$ are not standard DID parameters because $t-1$ is also a post-treatment period when $t > 2$. Under Assumption 4, it is true that for every $t$, $\min\{\Delta_t^{(0)}(trt) - \Delta_t^{(0)}(a), \Delta_t^{(0)}(trt) - \Delta_t^{(0)}(b)\} \leq 0$ and $\max\{\Delta_t^{(0)}(trt) - \Delta_t^{(0)}(a), \Delta_t^{(0)}(trt) - \Delta_t^{(0)}(b)\} \geq 0$. Therefore, when $t = 2$, the two DID parameters, $\tau_2(a)$ and $\tau_2(b)$, bound the $ATT_2$, i.e., $\min\{\tau_2(a), \tau_2(b)\} \leq ATT_2 \leq \max\{\tau_2(a), \tau_2(b)\}$; when $t > 2$, we have $\min\{\tau_t(a), \tau_t(b)\} \leq ATT_t - ATT_{t-1} \leq \max\{\tau_t(a), \tau_t(b)\}$.

We state the key partial identification result as follows.

**Theorem 1.** *Under Assumption 4, the average treatment effect for the treated $ATT_t, t = 2, \ldots, T$ can be partially identified through*

$$\sum_{s=2}^{t} \min\{\tau_s(a), \tau_s(b)\} \leq ATT_t \leq \sum_{s=2}^{t} \max\{\tau_s(a), \tau_s(b)\} \tag{10}$$

*where $\tau_2(a)$ and $\tau_2(b)$ are defined in (8), $\tau_s(a)$ and $\tau_s(b)$ for $s > 2$ are defined in (9).*

The tightness of the bounds depends on the magnitude of violation of the parallel trends assumption, since the width of the bounds equals

$$\sum_{s=2}^{t} [\max\{\tau_s(a), \tau_s(b)\} - \min\{\tau_s(a), \tau_s(b)\}] = \sum_{s=2}^{t} \left| \Delta_s^{(0)}(b) - \Delta_s^{(0)}(a) \right|,$$

where $|\cdot|$ denotes the absolute value of a constant. If the parallel trends assumption holds over the study period, i.e., $\Delta_s^{(0)}(a) = \Delta_s^{(0)}(b)$ for every $s$, the width of the bounds equals zero for every post-treatment period.

In fact, for the control groups $a$ and $b$, Theorem 1 holds under a weaker assumption than Assumption 4, that is the monotone trends assumption holds in a cumulative fashion. Specifically,

for the bounds in (10) to be valid, it suffices to assume

$$\sum_{s=2}^{t} \min\left\{\Delta_s^{(0)}(a), \Delta_s^{(0)}(b)\right\} \leq \sum_{s=2}^{t} \Delta_s^{(0)}(trt) \leq \sum_{s=2}^{t} \max\left\{\Delta_s^{(0)}(a), \Delta_s^{(0)}(b)\right\}. \tag{11}$$

This cumulative monotone trends assumption is useful for scenarios when Assumption 4 is subject to mild violations for brief time periods, but the cumulative monotone trends assumption (11) still holds. In the context of the voter ID example, this would occur if control groups $a$ and $b$ have long term turnout trends that bracket the potential turnout trends for Indiana or Georgia in the absence of voter ID laws, but there is one time period where that is not true. For example, Indiana or Georgia may experience an unobserved shock such that Assumption 4 fails between 2004 and 2008, but the cumulative monotone trends assumption (11) still holds for 2012 (i.e., $t = 3$). If true, the DID bracketing method would still be valid with respect to the treatment effect for 2012.

## 3.2 Constructing the Two Control Groups

In Theorem 1, we assume that two control groups satisfying Assumption 4 are given. In practice, we typically must identify these two control groups from the candidate control units. In this section, we present three methods that one could use to construct the control groups.

The first method is based on the equivalent formulation in Lemma 1 and is data-driven. Suppose data from at least two time points in a "prior-study" period (i.e., $t < 1$) are available, we identify two groups of control units whose relative outcomes compared with the treated group are negatively correlated during this prior-study period. First, for each candidate control unit $i$, let $\mathbf{\Gamma}^{(0)}(i)$ be the column vector of its average outcome relative to the treated group at every prior-study time period (i.e., the average outcome for every control unit after subtracting the average outcome for the treated group). Second, calculate the correlation matrix of $\mathbf{\Gamma}^{(0)}(i)$'s. Third, use a hierarchical clustering algorithm to find two clusters of control units that exhibit strong between-cluster negative correlation. These two clusters of control units are then designated as control groups $a, b$. The last two steps can be implemented in R using the *corrplot* package (Wei et al. 2017). By construction, control groups $a, b$ satisfy Assumption 4 in the prior-study period. We then must assume that this pattern persists during the study period. Conversely, if there are no control units which exhibit negative correlation, it indicates that Assumption 4 may not be supported by the data and the proposed strategy for bracketing may not be applicable.

The second method does not require the availability of data from "pre-study" time periods. Instead, we use an additional variable $S_t$ that can largely explain the heterogeneity in different

units' outcome dynamics. Assume that $\Gamma_t^{(0)}(g) = f(\Psi_t(g))$, where $f(\cdot)$ is an unspecified monotone function, $\Gamma_t^{(0)}(g) = E[Y_t^{(0)}|G = g] - E[Y_t^{(0)}|G = trt]$, $\Psi_t(g) = E[S_t|G = g] - E[S_t|G = trt]$, $g \in \{a, b\}$, then the negative correlation in $\Gamma_t^{(0)}(g)$ is implied by the negative correlation in $\Psi_t(g)$, specifically, (7) is implied by $\{\Psi_t(a) - \Psi_{t-1}(a)\}\{\Psi_t(b) - \Psi_{t-1}(b)\} \leq 0$. This relationship also holds in many scenarios when $f$ is time varying (see an example in Section 6.2). In this sense, if the variable $S_t$ is quantitatively available, we can apply the first control group construction method but with the past outcomes replaced by $S_t$ in the study period. The bracketing method that leverages $S_t$ for control group construction can be more robust than adjusting for $S_t$ via linear regression because no parametric assumption is imposed (a simulated comparison is in Section 6.2). In addition, if only some qualitative measure of the variable $S_t$ is available, this information can still be used in the bracketing method to validate the constructed control groups.

The third method is based on model (2). Consider applications in which the heterogeneity in different units' outcome dynamics in the absence of treatment is due to the time-varying effect of the time-invariant unmeasured confounder $U$ and the change in outcome in the absence of treatment is monotone in $U$. Then we can construct two control groups so that the unmeasured confounder $U$ for one control group is stochastically larger than $U$ for the treated group and $U$ for the other control group is stochastically smaller than $U$ for the treated group. A specific example is the group construction method in Hasegawa et al. (2019).

Finally, we remark that all three methods ensure that the data used for control groups construction does not overlap with the data used for estimation and inference, mitigating concerns about measurement error. In Section 7, we illustrate the control group construction methods using the voter ID example.

# 4 Inference for Union Bounds

Before formally introducing the inference method, it is important to discuss the implications of the i.i.d. assumption we invoked at the beginning of Section 2. Consider the standard model in a DID design

$$Y_{itj} = \alpha_i + \lambda_t + \beta Z_{it} + \eta_{it} + \epsilon_{itj},$$

where $i$ indexes cluster, $t$ indexes time, $j$ indexes individual (Imbens and Wooldridge 2008). Here, $Y_{itj}$ is the outcome, $Z_{it}$ is a indicator variable that equals one if cluster $i$ is being treated at time $t$, $\alpha_i, \lambda_t, \beta$ are unknown parameters, respectively representing time-invariant cluster effects, time effects and the treatment effect of interest, $\eta_{it}$ is an unobserved cluster-time specific effect,

and $\epsilon_{itj}$ is an individual-level random error.

Bertrand et al. (2004) and Donald and Lang (2007) noted that in some applied work, the $\eta_{it}$ (i.e., time specific factors which affect the whole cluster) were effectively set to zero and ignored, which can severely understate the standard error of the DID estimator. Donald and Lang (2007) outlined an approach which models the $\eta_{it}$ as mean zero random factors. In our approach, we instead view $\eta_{it}$ as fixed effects and propose to bracket the $\eta_{it}$'s of treated and control groups rather than modeling them as random. As such, the existence of non-zero $\eta_{it}$ can be accounted for using our bracketing method without creating within-cluster correlation.

Here, we introduce a novel bootstrap method to construct confidence intervals for the partially identified parameter of interest $ATT_t$ and its identified set in (10) for $t = 2, \ldots, T$. Differences between confidence intervals for the identified set and for the parameter of interest within that set have been well-addressed in the prior literature. See, for instance, Imbens and Manski (2004) and Stoye (2009). Algebra reveals that the identified set in (10) for $ATT_t$ can be equivalently formulated as

$$\min_{g_s \in \{a,b\}} \left\{ \sum_{s=2}^{t} \tau_s(g_s) \right\} \leq ATT_t \leq \max_{g_s \in \{a,b\}} \left\{ \sum_{s=2}^{t} \tau_s(g_s) \right\}. \tag{12}$$

where the proof is in the Supplementary Materials. For example, when $t = 2$, the bounding parameters are $\{\tau_2(a), \tau_2(b)\}$, and their minimum and maximum form the bounds for $ATT_2$; when $t = 3$, the bounding parameters are $\{\tau_2(a)+\tau_3(a), \tau_2(a)+\tau_3(b), \tau_2(b)+\tau_3(a), \tau_2(b)+\tau_3(b)\}$, and their minimum and maximum form the bounds for $ATT_3$. In general, there are $2^{t-1}$ bounding parameters for $t \geq 2$. We call such bounds "union bounds."

Inference for union bounds is challenging because of the minimum (min) and maximum (max) operators in the lower bound and upper bound as in (12). As noted by Manski and Pepper (2009, 2000), simply estimating the bounds using the minimum and maximum of the estimated bounding parameters tends to produce a wider interval than the true bounds. In fact, Hirano and Porter (2012) demonstrated that finding unbiased estimators for the bounds involving non-differentiable functionals is generally impossible. Recent literature also indicates that the canonical bootstrap is not generally consistent in this case (Shao 1994; Romano and Shaikh 2008; Andrews and Han 2009; Bugni 2010; Canay 2010). In a related setting, where the identified set is defined by moment inequalities, progress has been made on developing valid inferential methods, see Romano and Shaikh (2008); Andrews and Soares (2010); Chernozhukov et al. (2009, 2013); Bugni et al. (2017); Kaido et al. (2019). However, these procedures are not applicable in this setting because the union bounds cannot generally be represented using

moment inequalities. To the best of our knowledge, there is no formal investigation of this problem, and naive confidence intervals tend to be unnecessarily wide (see Section 6). Therefore, it is important that we develop valid and informative inference methods for the union bounds.

We rearrange the data as $\boldsymbol{O}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iT}, R_{i1}, \ldots, R_{iT}, G_i)$, $i = 1, \ldots, N$, which are assumed to be identically and independently distributed (i.i.d.) sequence of random vectors with distribution $P \in \mathcal{P}$, where $\mathcal{P}$ is some class of distributions, $Y_{it}$ is the outcome for individual $i$ at time $t$, $R_{it}$ indicates whether $Y_{it}$ is observed or not, which equals 1 if we observe $Y_{it}$, equals 0 if not, and $N = \sum_{t=1}^{T} N_t$ is the total number of individuals we observe. We assume $R_{it}$ is independent of $(Y_{it}, G_i)$ for every $t$, $P(R_{it} = 1)$ and $P(G_i = g)$ are strictly bounded away from zero for every $t$ and $g$. This data configuration enables the proposed inferential method to account for arbitrary serial correlation among multiple observed outcomes for an individual. Suppose based on $\boldsymbol{O}_i$'s, we compute a vector of sample means denoted by $\bar{\boldsymbol{X}}$ and $\boldsymbol{\mu} = E(\bar{\boldsymbol{X}})$. Let $\{\theta_j = \theta_j(\boldsymbol{\mu}), j = 1, \ldots, k\}$ be the set of bounding parameters and let $\hat{\theta}_j = \theta_j(\bar{\boldsymbol{X}})$ be their estimators, where $k$ is a finite number. The parameter of interest $\psi_0$ belongs to the identified set $\Psi_0 = [\min_j \theta_j, \max_j \theta_j]$. The goal is to construct uniformly valid confidence intervals for $\Psi_0 = [\min_j \theta_j, \max_j \theta_j]$ and $\psi_0$ in the asymptotic sense.

The bootstrap inference method is implemented as follows: Generate $B$ bootstrap samples, compute $\min_j \hat{\theta}_j^{*b}, \max_j \hat{\theta}_j^{*b}$ for each bootstrap sample, and obtain the $1 - \alpha/2$ sample quantile of $\{\min_j \hat{\theta}_j^{*b}, b = 1, \ldots, B\}$ and the $\alpha/2$ sample quantile of $\{\max_j \hat{\theta}_j^{*b}, b = 1, \ldots, B\}$, respectively denoted as $Q_{1-\alpha/2}\left(\{\min_j \hat{\theta}_j^{*b}\}_{b\in[B]}\right), Q_{\alpha/2}\left(\{\max_j \hat{\theta}_j^{*b}\}_{b\in[B]}\right)$. The random interval

$$\left[2\min \hat{\theta}_j - Q_{1-\alpha/2}\left(\{\min_j \hat{\theta}_j^{*b}\}_{b\in[B]}\right), 2\max \hat{\theta}_j - Q_{\alpha/2}\left(\{\max_j \hat{\theta}_j^{*b}\}_{b\in[B]}\right)\right] \tag{13}$$

is a uniformly valid $1-\alpha$ level confidence interval for the identified set $\Psi_0$, and thus the parameter of interest $\psi_0$.

Next, we derive the theoretical properties of the confidence interval (13) and its extensions. Suppose we obtain a nonparametric bootstrap sample $\boldsymbol{O}_1^*, \ldots, \boldsymbol{O}_N^*$ that is drawn from the empirical distribution based on $\boldsymbol{O}_1, \ldots, \boldsymbol{O}_N$. Let $\bar{\boldsymbol{X}}^*$ and $\hat{\theta}_j^* = \theta_j(\bar{\boldsymbol{X}}^*), j = 1, \ldots, k$ be the bootstrap analogues calculated based on the bootstrap sample. Let $L(x) = P\{\sqrt{N}(\min_j \hat{\theta}_j - \min_j \theta_j) \leq x\}$ be the true distribution, $\hat{L}(x) = P_*\{\sqrt{N}(\min_j \hat{\theta}_j^* - \min_j \hat{\theta}_j) \leq x\}$ be the bootstrap estimator of $L(x)$, where $P_*$ is the conditional probability with respect to the random generation of bootstrap sample given the original data. Similarly, let $R(x) = P\{\sqrt{N}(\max_j \hat{\theta}_j - \max_j \theta_j) \leq x\}$ be the true distribution, $\hat{R}(x) = P_*\{\sqrt{N}(\max_j \hat{\theta}_j^* - \max_j \hat{\theta}_j) \leq x\}$ be the bootstrap estimator of $R(x)$.

The following theorem lays a theoretical foundation for the proposed inference procedure.

**Theorem 2.** *Suppose that* $E\bar{\boldsymbol{X}}^2 < \infty$, *and* $\theta_j$ *is continuously differentiable at* $\boldsymbol{\mu}$ *with* $\nabla\theta_j(\boldsymbol{\mu}) \neq \mathbf{0}$ *for every* $j = 1\ldots, k$.

*(a)* $\lim_{N\to\infty} \sup_{P\in\mathcal{P}} \sup_{x\in\mathcal{R}}\{\hat{L}(x) - L(x)\} \leq 0$ *and* $\lim_{N\to\infty} \inf_{P\in\mathcal{P}} \inf_{x\in\mathcal{R}}\{\hat{R}(x) - R(x)\} \geq 0$.

*(b) Let* $c_L^*(p) = \inf\{x \in \mathcal{R} : \hat{L}(x) \geq p\}$, $c_U^*(p) = \sup\{x \in \mathcal{R} : \hat{R}(x) \leq p\}$, *then*

$$\lim_{N\to\infty} \inf_{P\in\mathcal{P}} P\left\{\sqrt{N}(\min_j \hat{\theta}_j - \min_j \theta_j) \leq c_L^*(p)\right\} \geq p$$
$$\lim_{N\to\infty} \inf_{P\in\mathcal{P}} P\left\{\sqrt{N}(\max_j \hat{\theta}_j - \max_j \theta_j) \geq c_U^*(1-p)\right\} \geq p$$

*(c) Taking* $p = 1 - \alpha/2$, *then*

$$CI_{1-\alpha} \equiv [\min_j \hat{\theta}_j - N^{-1/2} c_L^*(1-\alpha/2), \max_j \hat{\theta}_j - N^{-1/2} c_U^*(\alpha/2)] \tag{14}$$

*is a uniformly valid* $1-\alpha$ *level confidence interval for the identified set* $\Psi_0 = [\min_j \theta_j, \max_j \theta_j]$, *i.e.,* $\lim_{N\to\infty} \inf_{P\in\mathcal{P}} P\{[\min_j \theta_j, \max_j \theta_j] \in CI_{1-\alpha}\} \geq 1-\alpha$.

*(d) Let* $\hat{w}^+ = \hat{w}I(\hat{w} > 0)$, *where* $\hat{w} = \{\max_j \hat{\theta}_j - N^{-1/2} c_U^*(1/2)\} - \{\min_j \hat{\theta}_j - N^{-1/2} c_L^*(1/2)\}$, *and* $\hat{p} = 1 - \Phi(\rho\hat{w}^+)\alpha$, *where* $\Phi(\cdot)$ *is the standard normal cumulative distribution function,* $\rho$ *is a sequence of constants satisfying* $\rho \to \infty$, $N^{-1/2}\rho \to 0$ *and* $\rho|\hat{w}^+ - (\max_j \theta_j - \min_j \theta_j)| \xrightarrow{p} 0$, *where* $\xrightarrow{p} 0$ *denotes convergence in probability, then*

$$CI_{1-\alpha}^{\psi} \equiv [\min_j \hat{\theta}_j - N^{-1/2} c_L^*(\hat{p}), \max_j \hat{\theta}_j - N^{-1/2} c_U^*(1-\hat{p})] \tag{15}$$

*is a uniformly valid* $1-\alpha$ *level confidence interval for the partially identified parameter* $\psi_0$, *i.e.,* $\lim_{N\to\infty} \inf_{P\in\mathcal{P}} \inf_{\psi_0\in\Psi_0} P\left\{\psi_0 \in CI_{1-\alpha}^{\psi}\right\} \geq 1-\alpha$.

The proof is in the Supplementary Materials. From Theorem 2(a), the bootstrap estimators $\hat{L}(x), \hat{R}(x)$ are not consistent for the true distributions $L(x), R(x)$, and this is caused by the possibility that there can be more than one bounding parameters equal to $\min_j \theta_j$ and $\max_j \theta_j$. Also, this inconsistency is directional, particularly, $\hat{L}(x)$ tends to be smaller than $L(x)$, and $\hat{R}(x)$ tends to be larger than $R(x)$. Given that the goal is to construct a confidence interval with asymptotic coverage probability at least $1-\alpha$, using critical values $c_L^*(p), c_U^*(p)$ satisfying $\hat{L}(c_L^*(p)) \geq p, \hat{R}(c_U^*(p)) \leq p$, we have asymptotically $L(c_L^*(p)) \geq \hat{L}(c_L^*(p)) \geq p$, $R(c_U^*(p)) \leq \hat{R}(c_U^*(p)) \leq p$. This property lays the ground for the proposed inference method.

The interval (13) is in fact a Monte Carlo approximation of (14) in Theorem 2(c), because $Q_{1-\alpha/2}\left(\{\min_j \hat{\theta}_j^{*b}\}_{b\in[B]}\right) - \min_j \hat{\theta}_j = Q_{1-\alpha/2}\left(\{\min_j \hat{\theta}_j^{*b} - \min_j \hat{\theta}_j\}_{b\in[B]}\right) \approx N^{-1/2} c_L^*(1-\alpha/2)$ and $Q_{\alpha/2}\left(\{\max_j \hat{\theta}_j^{*b}\}_{b\in[B]}\right) - \max_j \hat{\theta}_j = Q_{\alpha/2}\left(\{\max_j \hat{\theta}_j^{*b} - \max_j \hat{\theta}_j\}_{b\in[B]}\right) \approx N^{-1/2} c_U^*(\alpha/2)$.

Notice that the confidence interval $CI_{1-\alpha}$ in Theorem 2(c) is also a uniformly valid confidence interval for the parameter of interest $\psi_0$, because the event that $CI_{1-\alpha}$ covers the identified set $\Psi_0 = [\min_j \theta_j, \max_j \theta_j]$ implies the event that $CI_{1-\alpha}$ covers the parameter of interest $\psi_0$. This also means that $CI_{1-\alpha}$ as a confidence interval for $\psi_0$ can be further improved by taking into consideration the width of the identified set (Imbens and Manski 2004; Stoye 2009), which motivates the confidence interval $CI^{\psi}_{1-\alpha}$ in Theorem 2(d).

The idea in Theorem 2(d) is to set $\hat{p} = 1 - \alpha$ when the bounds are wide enough in the sense that $\rho(\max_j \theta_j - \min_j \theta_j) \to \lambda \in (0, \infty]$, and set $\hat{p} = 1 - \alpha/2$ if $\rho(\max_j \theta_j - \min_j \theta_j) \to 0$, where $\rho$ satisfies $\rho \to \infty, N^{-1/2}\rho \to 0$. The use of $\Phi(\cdot)$ is simply to smoothly connect these two ends because $\Phi(\rho\hat{w}^+) \in [0, 1/2]$. The intuition behind this construction is that if the bounds are wide relative to the measurement error, the parameter of interest $\psi_0$ can only be close to at most one boundary of the identified set, so the asymptotic probability that $\psi_0$ is more extreme than the other boundary is negligible and the noncoverage risk is one-sided. This reasoning appears in Imbens and Manski (2004), Stoye (2009) and Chernozhukov et al. (2009). In practice, we set $\rho = \{N^{-1/2} \max[c^*_U(3/4) - c^*_U(1/4), c^*_L(3/4) - c^*_L(1/4)]\}^{-1} / \log(N)$. Again, we emphasize that the union bounds we focus on in this article are different from the intersection bounds considered in Chernozhukov et al. (2009, 2013), in which the minimum operator appears in the upper bound and the maximum operator appears in the lower bound. Applying the inference method for intersection bounds developed in Chernozhukov et al. (2009, 2013) to the union bounds tends to produce a confidence interval with insufficient coverage probability.

We emphasize that $CI^{\psi}_{1-\alpha}$ is valid uniformly with respect to the location of $\psi_0$ in the identified set $\Psi_0 = [\min \theta_j, \max_j \theta_j]$ and the width of $\Psi_0$. This uniformity is important because it ensures that the coverage probability is adequate even when $\psi_0$ is at the boundary of $\Psi_0$, or when the width of $\Psi_0$ shrinks towards zero and point identification is established. In particular, the point identification scenario is very salient for the union bounds developed in Section 3, because when the parallel trends assumption holds over the study period, the width of the identified set in (12) equals zero. Theorem 2(d) guarantees that the confidence interval is valid when the parallel trends assumption holds.

The proposed inference method generally applies to an identified set for a parameter that can be expressed as union bounds, i.e., the lower bound can be formulated as the minimum of a set of bounding parameters, and the upper bound can be formulated as the maximum of another set of bounding parameters. These two sets of bounding parameters can be different. The proposed inference method also applies to other data structures, e.g., cross sectional data.

Finally, as an illustration, in the setting when the parameter of interest is $ATT_2$,

$$\bar{\boldsymbol{X}} = \begin{bmatrix} \frac{\sum_{G_i=trt} Y_{i2}R_{i2}}{\sum_{G_i=trt} R_{i2}} - \frac{\sum_{G_i=trt} Y_{i1}R_{i1}}{\sum_{G_i=trt} R_{i1}} \\ \frac{\sum_{G_i=a} Y_{i2}R_{i2}}{\sum_{G_i=a} R_{i2}} - \frac{\sum_{G_i=a} Y_{i1}R_{i1}}{\sum_{G_i=a} R_{i1}} \\ \frac{\sum_{G_i=b} Y_{i2}R_{i2}}{\sum_{G_i=b} R_{i2}} - \frac{\sum_{G_i=b} Y_{i1}R_{i1}}{\sum_{G_i=b} R_{i1}} \end{bmatrix} \qquad \boldsymbol{\mu} = \begin{bmatrix} E(Y_2 - Y_1|G = trt) \\ E(Y_2 - Y_1|G = a) \\ E(Y_2 - Y_1|G = b) \end{bmatrix}$$

and $\theta_1 = E(Y_2 - Y_1|G = trt) - E(Y_2 - Y_1|G = a)$, $\theta_2 = E(Y_2 - Y_1|G = trt) - E(Y_2 - Y_1|G = b)$. Theorem 2 would construct uniformly valid confidence intervals for identified set $[\min(\theta_1, \theta_2), \max(\theta_1, \theta_2)]$ and the parameter of interest $ATT_2$. In Section 6, we conduct empirical studies on the proposed inference methods.

# 5 Falsification Test and Sensitivity Analysis

To enhance the reliability of an observational study, it is useful to include additional analyses that explore whether the key assumptions appear plausible and whether conclusions are robust to violations of the key assumptions (Rosenbaum 2010). In this section, we complete our methodology by adding these two elements of analysis via falsification test and sensitivity analysis.

## 5.1 Falsification Test

A falsification test is one method for exploring whether the identification assumptions are plausible. Falsification tests are often possible due to the fact that causal theories do more than predict the presence of a causal effect; causal theories may also predict an absence of causal effects in other instances. Angrist and Krueger (1999) refer to such tests as instances of "refutability." In the context of DID, investigators often conduct a falsification test by testing for parallel trends in pre-treatment time periods (Angrist and Pischke 2009, ch. 5). Next, we outline how the key partial identification assumption (Assumption 4) can be tested indirectly via falsification.

Recall that Assumption 4 requires that for every pair of adjacent time points in the study period, the changes in outcome for the two control groups bound the change in outcome for the treated group if the treated group were untreated. The falsification test is based on the assumption that if the monotone trends relationship holds in a pair of unused adjacent time periods prior to the study period, say $t = t_1^*, t_2^*$, then it is plausible that the monotone trends relationship also holds in the study period ($t = 2, \ldots, T$).

To this end, we formulate the following null hypothesis that the monotone trends hold during

the time periods $t = t_1^*, t_2^*$, i.e.,

$$H_0 : \min\left[\Delta_{t_2^*}^{(0)}(a), \Delta_{t_2^*}^{(0)}(b)\right] \leq \Delta_{t_2^*}^{(0)}(trt) \leq \max\left[\Delta_{t_2^*}^{(0)}(a), \Delta_{t_2^*}^{(0)}(b)\right],$$

and the alternative hypothesis includes two possible scenarios when $H_0$ is not true:
(i) $\min\left[\Delta_{t_2^*}^{(0)}(a), \Delta_{t_2^*}^{(0)}(b)\right] > \Delta_{t_2^*}^{(0)}(trt)$; (ii) $\max\left[\Delta_{t_2^*}^{(0)}(a), \Delta_{t_2^*}^{(0)}(b)\right] < \Delta_{t_2^*}^{(0)}(trt)$. Next, we define a set of simple hypotheses:

$$\begin{aligned}
H_a^i &: \Delta_{t_2^*}^{(0)}(a) - \Delta_{t_2^*}^{(0)}(trt) \leq 0 \\
H_b^i &: \Delta_{t_2^*}^{(0)}(trt) - \Delta_{t_2^*}^{(0)}(b) \leq 0 \\
H_a^d &: \Delta_{t_2^*}^{(0)}(a) - \Delta_{t_2^*}^{(0)}(trt) \geq 0 \\
H_b^d &: \Delta_{t_2^*}^{(0)}(trt) - \Delta_{t_2^*}^{(0)}(b) \geq 0
\end{aligned}$$

The null hypothesis $H_0$ can be written in a form of a composite null hypothesis, that is

$$H_0 : (H_a^i \cap H_b^i) \cup (H_a^d \cap H_b^d)$$

Let the p-values testing each individual hypothesis $H_a^i, H_b^i, H_a^d, H_b^d$ be $p_a^i, p_b^i, p_a^d, p_b^d$. From the definition of one-sided p-values, $p_a^d = 1 - p_a^i, p_b^d = 1 - p_b^i$. Therefore, following Bonferroni's method and Berger (1982), we reject $H_0$ if

$$\max(\min(p_a^i, p_b^i), \min(1 - p_a^i, 1 - p_b^i)) \leq \alpha/2.$$

Critically, this is a falsification test, such that failing to reject $H_0$ does not mean Assumption 4 is valid, but rejecting $H_0$ implies the data may be inconsistent with Assumption 4. In a falsification test, we prefer larger p-values, since this is better evidence of assumption plausibility. Visual inspection of the outcome trends before the study period can also be performed by plotting average pre-treatment outcomes. If every slope for the treated group is bounded in between by the slopes for the two control groups, it provides visual evidence that Assumption 4 is plausible.

## 5.2 Sensitivity Analysis

In this section, we develop a sensitivity analysis to evaluate how sensitive the conclusion is to violations of Assumption 4. A sensitivity analysis is used to quantify the degree to which a key identification assumption must be violated in order for a researcher's original conclusion to be

reversed. If a causal conclusion is sensitive, a slight violation of the assumption may lead to substantively different conclusions. There is a large and growing literature on sensitivity analysis, e.g., Rosenbaum (1987*b*); Imbens (2003); Richardson et al. (2014); Ding and VanderWeele (2016); Fogarty (2019). See Daniels and Hogan (2008) and Rosenbaum (2010) for textbook references.

There are two scenarios when Assumption 4 is violated at time $t$: (i) $\min\left[\Delta_t^{(0)}(a), \Delta_t^{(0)}(b)\right] > \Delta_t^{(0)}(trt)$; or (ii) $\max\left[\Delta_t^{(0)}(a), \Delta_t^{(0)}(b)\right] < \Delta_t^{(0)}(trt)$. We use two sensitivity parameters $\gamma_t$ and $\delta_t$ for these two scenarios at time $t$, where $\gamma_t$ is for scenario (i) and $\delta_t$ is for scenario (ii).

We introduce the following sensitivity assumption.

**Assumption 5** (Sensitivity)**.** *For the given non-negative sensitivity parameters $\{\gamma_t, \delta_t\}_{t\geq 2}$, and the two control groups $a$ and $b$,*

$$\min\left[\Delta_t^{(0)}(a), \Delta_t^{(0)}(b)\right] - \gamma_t \leq \Delta_t^{(0)}(trt) \leq \max\left[\Delta_t^{(0)}(a), \Delta_t^{(0)}(b)\right] + \delta_t$$

*holds for $t = 2, \ldots, T$,*

When $\gamma_t = \delta_t = 0$ for every $t$, Assumption 5 degenerates to Assumption 4, under which the bounds in (10) are valid. Next, we derive the bounds and confidence interval for $ATT_t$ under Assumption 5, which will serve as a basis for the sensitivity analysis.

**Theorem 3.** *Under Assumption 5, for $t = 2, \ldots, T$,*
*(a) The treatment effect for treated $ATT_t$ can be partially identified through*

$$\sum_{s=2}^{t} \min\{\tau_s(a), \tau_s(b)\} - \sum_{s=2}^{t} \delta_s \leq ATT_t \leq \sum_{s=2}^{t} \max\{\tau_s(a), \tau_s(b)\} + \sum_{s=2}^{t} \gamma_s, \tag{16}$$

*where $\tau_s(a)$ and $\tau_s(b)$ are defined in (8)-(9), $\delta_s, \gamma_s$ are the sensitivity parameters defined in Assumption 5.*
*(b) Further assume the conditions in Theorem 2, let $[\hat{l}_t, \hat{r}_t]$ be the uniformly valid $1-\alpha$ confidence interval for $ATT_t$ (or the identified set) developed using Theorem 2 (c)-(d) under Assumption 4, then $[\hat{l}_t - \sum_{s=2}^{t} \delta_s, \hat{r}_t + \sum_{s=2}^{t} \gamma_s]$ is a uniformly valid $1-\alpha$ confidence interval for $ATT_t$ (or the identified set) under Assumption 5.*

The proof is in the Supplementary Materials. The confidence interval $[\hat{l}_t, \hat{r}_t]$ can be constructed using the method developed in Theorem 2. The form of the confidence interval under Assumption 5 illustrates how large $\gamma_s, \delta_s$ have to be in order for the study conclusion to be materially altered. Given that $[\hat{l}_t, \hat{r}_t]$ is the constructed $1-\alpha$ confidence interval for $ATT_t$ under

Assumption 4, if $\hat{l}_t, \hat{r}_t$ are both positive, $\sum_{s=2}^{t} \delta_s = \hat{l}_t$ would suffice to explain away the treatment effect but large $\gamma_s$'s would not, in which case $\sum_{s=2}^{t} \Delta_s^{(0)}(trt) \geq \sum_{s=2}^{t} \max[\Delta_2^{(0)}(a), \Delta_2^{(0)}(b)] + \hat{l}_t$; if $\hat{l}_t, \hat{r}_t$ are both negative, $\sum_{s=2}^{t} \gamma_s = \hat{r}_t$ would suffice to explain away the treatment effect but large $\delta_s$'s would not, in which case $\sum_{s=2}^{t} \Delta_s^{(0)}(trt) \leq \sum_{s=2}^{t} \min[\Delta_2^{(0)}(a), \Delta_2^{(0)}(b)] - \hat{r}_t$.

# 6 Simulations

## 6.1 Bootstrap Inference for Union Bounds

In this section, we empirically evaluate the finite sample performance of the proposed bootstrap inference methods for union bounds. We simulate data for the treated group and two control groups at 4 time periods and we consider the following two scenarios:

**Case I: parallel trends:** $E[Y_1^{(0)}|G = trt] = 3, E[Y_1^{(0)}|G = a] = 10, E[Y_1^{(0)}|G = b] = 4$, $\Delta_t^{(0)}(trt) = \Delta_t^{(0)}(a) = \Delta_t^{(0)}(b) \equiv \Delta_t$ for every $t$, where $\Delta_1 = 1, \Delta_2 = -2, \Delta_3 = -1$.

**Case II: partially parallel trends:** $E[Y_1^{(0)}|G = trt] = 3, E[Y_1^{(0)}|G = a] = 10, E[Y_1^{(0)}|G = b] = 4$, $\Delta_1^{(0)}(trt) = 1, \Delta_2^{(0)}(trt) = -4, \Delta_3^{(0)}(trt) = 1, \Delta_1^{(0)}(a) = 1, \Delta_2^{(0)}(a) = -1, \Delta_3^{(0)}(a) = 1, \Delta_1^{(0)}(b) = 2, \Delta_2^{(0)}(b) = -4, \Delta_3^{(0)}(b) = 1$.

In both scenarios, the simulated data resembles a longitudinal study where $N = 1000$ individuals are followed for $T = 4$ time points. The group indicators $G_i$ are given, where $P(G = trt) = 0.3$, $P(G = a) = 0.2$, $P(G = b) = 0.5$. The average treatment effects for the treated group are $ATT_2 = 2, ATT_3 = ATT_4 = 1$. The observed outcomes are generated from $Y_{it} = E(Y_t|G) + \varepsilon_{it}$, where $\varepsilon_{it}$'s are independent from the standard normal distribution. We set the number of bootstrap iterations $B = 300$ and significance level $\alpha = 0.05$.

We compare with the naive method, where one constructs the confidence interval for $ATT_t$ by taking the union of every $1 - \alpha$ confidence interval for $\sum_{s=2}^{t} \tau_s(g_s), g_s \in \{a, b\}$, due to the union bounds form in (12). Following a similar proof as in Hasegawa et al. (2019), one can show that this naive confidence interval has asymptotically at least $1 - \alpha$ coverage probability for the identified set and the parameter of interest.

Table 1 contains the simulation results. The results contain the average length and the empirical probability that the confidence interval covers the parameter of interest $ATT_t$ (i.e., coverage probability) for the following confidence intervals (CIs): the proposed bootstrap confidence interval for the identified set (Theorem 2(c)), the proposed bootstrap confidence interval for the parameter of interest $ATT_t$ (Theorem 2(d)), and the naive confidence interval described above. We summarize the key findings as follows. First, the proposed bootstrap method and the naive method produce confidence intervals with adequate coverage probability, indicating the

Table 1: A comparison of the proposed bootstrap method and the naive method with respect to the average confidence interval (CI) length and coverage probability (CP) of 95% CIs ($N = 1000, B =$ 300, 1000 simulation runs).

| | Bootstrap CI | | | | Naive CI | |
|---|---|---|---|---|---|---|
| | **Identified Set** | | $\boldsymbol{ATT_t}$ | | | |
| | Length | CP | Length | CP | Length | CP |
| **Case I** | | | | | | |
| $t = 2$ | 0.481 | 0.961 | 0.476 | 0.961 | 0.554 | 0.990 |
| $t = 3$ | 0.581 | 0.979 | 0.573 | 0.979 | 0.732 | 0.998 |
| $t = 4$ | 0.670 | 0.983 | 0.659 | 0.983 | 0.892 | 1.000 |
| **Case II** | | | | | | |
| $t = 2$ | 1.453 | 0.982 | 1.404 | 0.969 | 1.444 | 0.980 |
| $t = 3$ | 4.559 | 0.980 | 4.469 | 0.960 | 4.550 | 0.978 |
| $t = 4$ | 4.631 | 0.985 | 4.539 | 0.964 | 4.694 | 0.991 |

validity of all three types of confidence intervals. Second, confidence intervals constructed using the naive method can be unnecessarily wide, especially when the width of the bounds is small (Case I) and the number of the bounding parameters is large ($t = 4$). Overall, the proposed bootstrap method improves significantly over the naive method, as it produces tighter confidence intervals with correct coverage probability. Lastly, when one is interested in a confidence interval for the partially identified parameter itself rather than the identified set, the bootstrap confidence interval tailored for the parameter can be tighter than that for the identified set, and the improvement is more evident when the width of the bounds is large (Case II).

## 6.2 Bracketing and linear fixed effect regression

We have discussed in Section 3.2 that when an observed variable $S_t$ is driving the heterogeneity in different units' outcome dynamics, the bracketing strategy may still be more preferable than adjusting for $S_t$ via linear fixed effect regression, because inference based on the bracketing method can be more robust, since it does not depend on correct specification of the regression model. In this section, we present a simulation study that illustrates this point.

We simulate a population of $N = 1000$ individuals from one treated group and two control groups at 4 time periods. The group indicators $G_i$ are given, with $P(G = trt) = 0.3, P(G = a) = 0.2, P(G = b) = 0.5$. Let $\mu(a) = (3, 5, 1, 1), \mu(b) = (10, 12, 8, 11), \mu(trt) = (4, 6, 2, 3)$ respectively be the mean of $S_{it}$ at 4 time periods for three groups, the observed variable $S_{it}$ is generated from a normal distribution with the specified mean and with standard deviation equal to 0.5. The observed outcome is generated from $Y_{it} = 2 - t + 0.2tS_{it} + ATT_t I(G_i = trt) + \epsilon_{it}$, where $t =$

$1, \ldots, 4, ATT_1 = 0, ATT_2 = 2, ATT_3 = 3, ATT_4 = 1$, $\epsilon_{it}$ follows a standard normal distribution. The Pearson correlation between the average relative outcome for control groups $a, b$ compared with the treated group are smaller than -0.975 in all 1000 simulation runs, suggesting a strong negative correlation in the relative $S_t$ between the two control groups.

From the data generating process, it is true that $Y_{it}^{(0)}$ is independent of $G_i$ conditional on $S_{it}$. In this case, it is common to fit the linear fixed effect model that includes a full set of group dummies, a full set of time dummies, variable $S_t$, and three treatment indicators that respectively equal one for the treated group at $t = 2, t = 3, t = 4$. Clearly this is a misspecified model because one fails to model the time-varying effect of $S_t$. In Table 2, we compare the bracketing method with the linear fixed effect model approach based on 1000 simulation runs. For the bracketing method, Table 2 shows the average 95% bootstrap confidence interval (CI) for the identified set, and the average 95% bootstrap confidence interval for the parameter of interest $ATT_t$, where the number of bootstrap iterations is set as $B = 300$; for the linear fixed effect model, Table 2 shows the mean and standard deviation (SD) of the estimated coefficients, the average of estimated standard errors (SE) and the corresponding average 95% confidence interval.

Table 2: A comparison of the bracketing method and the linear fixed effect regression with $ATT_2 = 2, ATT_3 = 3, ATT_4 = 1$ ($N = 1000, B = 300$, 1000 simulation runs). CI shows the lower and upper confidence interval means. The empirical coverage probabilities for bracketing confidence intervals are all larger than 99.9%, while that for linear fixed effect confidence intervals are 0.

| | Bracketing | | Linear Fixed Effect | | | |
|---|---|---|---|---|---|---|
| | **Identified Set** | $\boldsymbol{ATT_t}$ | | | | |
| | CI | CI | mean | SD | SE | CI |
| $t = 2$ | [0.598, 2.464] | [0.624, 2.429] | 1.200 | 0.106 | 0.113 | [0.979, 1.421] |
| $t = 3$ | [0.389, 3.666] | [0.421, 3.623] | 1.398 | 0.104 | 0.113 | [1.176, 1.619] |
| $t = 4$ | [-4.412, 2.668] | [-4.380, 2.625] | -1.052 | 0.109 | 0.117 | [-1.280, -0.823] |

The bracketing method shows good performance, with confidence intervals covering the true $ATT_t$ in almost all the simulation runs (all the empirical coverage probabilities $\geq 99.9\%$). In contrast, the linear fixed effect treatment effect estimators are biased, with none of the confidence interval covering the true $ATT_t$ across the simulation runs. Moreover, the confidence interval for $ATT_4$ is entirely in the opposite direction, which can result in misleading interpretations. Therefore, the simulation results support that when information on time-varying covariates are available, the bracketing method may still be more preferable than the linear fixed effect regression, because inference based on bracketing is more robust, and does not depend on model

specification.

# 7 Application to Effect of Voter Identification Laws on Voter Turnout

We now apply the proposed methods to the analysis of voter ID laws in Indiana and Georgia.

## 7.1 Constructing Control Groups

Our first task is constructing control groups respectively for Indiana and Georgia. As described in Section 3.2, when data before the study period is available, the most straightforward way to construct control groups is to use a hierarchical clustering algorithm to find two clusters of control states whose relative turnouts compared with the treated state exhibit strong between-cluster negative correlation. For this task, we built a dataset using the voter turnout data from the United States Elections Project, which contains state turnout rates for the voting-eligible population during 1980-2000 (McDonald 2020). By only using data up to year 2000, we separate the data used for control group construction from the data used for analysis and inference.

Figure 2 contains the correlation matrix of relative turnout when Indiana is the treated group. We observe that most candidate control states' relative turnout rates are positively correlated after subtracting the turnout rates for Indiana. The exceptions are Utah and West Virginia, whose relative turnout rates display strong negative correlation. Based on this analysis, we would designate Utah as control group $a$ and West Virginia as control group $b$. Figure 3 contains the correlation matrix of relative turnouts when Georgia is the treated group. Here, we find that Wyoming exhibits negative correlations with almost all the other candidate control states. Therefore, we designate Wyoming as control group $a$, and states whose relative turnout rates show strong negative correlations with Wyoming as control group $b$. Specifically, control group $b$ includes Michigan, Oregon and Iowa.

Next, we implement the second method introduced in Section 3.2 to validate the constructed control groups. A good candidate variable $S_t$ would be a measure of political advertising for each state at time $t$, because difference in the change in turnout over time between two groups of states is likely to reflect the difference in the change in how much of a battleground the states are over time, which may be well captured by differences in the change in political advertising over time. If we find the constructed control groups' relative measure of political advertising compared with the treated group are also negatively correlated during 2004-2012, it provides additional evidence that the control groups satisfy Assumption 4. Using a ranking of how much

of a battleground a state is in Ostermeier (2020) based on the number of tightly contested presidential elections the state has had, we notice that Georgia is a weak battleground state while the state in its control group $a$ (i.e., Wyoming) has almost the weakest battleground level while states in its control group $b$ (i.e., Michigan, Oregon and Iowa) are all strong battleground states. During 2004-2012, Wyoming remained a safe state that was most likely exposed to very little political advertising effort. Georgia was a safe state in 2004 and 2012, and had a relatively more competitive election in 2008; Michigan, Oregon, Iowa were all considered as battleground states for every election during 2004-2012 and were likely exposed to much political advertising effort. Hence, it seems plausible that the turnout rate for control group $b$ had a larger increase (or a smaller decrease) than Georgia, while the turnout rate for control group $a$ had a smaller increase (or a larger decrease) than Georgia, both from 2004 to 2008, and from 2004 to 2012 (recall that we only need cumulative monotonicity from (11)). Similarly when Indiana is the treated group, Indiana is a weak battleground state, the state in its control group $a$ (i.e., Utah) is the weakest battleground state of all states and the state in its control group b (i.e., West Virginia) is a slightly stronger battleground state than Indiana. Thus, consideration of states' battleground rankings provides support for the control groups for Georgia and Indiana satisfying the key partial identification assumption for bracketing (Assumption 4).

## 7.2 Estimation and Inference

Based on the constructed control groups, we apply the developed methods to study the effect of voter ID laws on turnout rates in Georgia and Indiana. All the following analyses rely on data from the Current Population Survey (CPS) voting supplement conducted in November of each election year (U.S. Department of Commerce 2004, 2008, 2012).

We conducted two versions of the bracketing analysis. The unadjusted version simply estimates the DID parameters using average turnout rates; the adjusted version estimates the DID parameters controlling for education, income, residency status, sex, race, hispanic and employment status. The inference results are in Table 3.

From Table 3, the voter ID laws in fact lead to significantly higher turnout rates for Georgia in 2008 (95% CI: unadjusted $[2.89, 13.14]$, adjusted $[4.08, 12.04]$) and 2012 (95% CI: unadjusted $[2.50, 15.22]$, adjusted $[5.14, 13.28]$). The effect of the voter ID laws on turnout rates for Indiana in 2008 is not significant when not adjusting for covariates (95% CI: unadjusted $[-1.62, 19.88]$), but is significant when adjusting for covariates (95% CI: adjusted $[0.89, 14.46]$); the effect for Indiana in 2012 is not significant (95% CI: unadjusted $[-5.46, 26.72]$, adjusted $[-2.58, 17.48]$). In this example, the confidence intervals for the identified set and the parameter of interest are

nearly identical, which is because the widths of the bounds are not large compared with the measurement error and thus $\hat{p}$ defined in Theorem 2(d) is very close to $1 - \alpha/2$.

Our finding that the voter ID laws increased turnout in Georgia in both 2008 and 2012 and increased turnout in Indiana in 2008 (when adjusting for covariates), is consistent with other work that has found voter ID laws cause higher turnout Hopkins et al. (2017). Although this conclusion seems counterintuitive, there are two possible explanations of why the turnout rates are not lowered by the voter ID laws according to Valentino and Neuner (2017) and Highton (2017). First, the implementation of voter ID laws may trigger countermobilization strategies, for example, groups may be motivated to help those who do not already have proper ID to obtain it. Second, the way that the popular media frames the passing of the voter ID laws make some voters angry and those voters become mobilized because they perceive themselves as being disadvantaged by the laws.

Table 3: DID Bracketing 95% Confidence Intervals (CIs) for the effect of voter ID laws on turnout rates in Georgia and Indiana (in %, $B = 300$).

| | Unadjusted | | Adjusted | |
|---|---|---|---|---|
| | **Identified Set** CI | $\boldsymbol{ATT_t}$ CI | **Identified Set** CI | $\boldsymbol{ATT_t}$ CI |
| **Georgia** | | | | |
| 2008 | [2.89, 13.14] | [2.89, 13.14] | [4.08, 12.04] | [4.08, 12.04] |
| 2012 | [2.49, 15.30] | [2.50, 15.22] | [5.14, 13.28] | [5.14, 13.28] |
| **Indiana** | | | | |
| 2008 | [-1.70, 20.01] | [-1.62, 19.88] | [0.81, 14.87] | [0.89, 14.46] |
| 2012 | [-5.81, 27.23] | [-5.46, 26.72] | [-2.90, 17.68] | [-2.58, 17.48] |

### 7.3 Falsification Test and Sensitivity Analysis

Finally, we report the results of falsification test and sensitivity analysis. We implement the falsification test for Georgia and Indiana using the CPS data from election years 1988 and 1992, two presidential election years before the study period (U.S. Department of Commerce 1988, 1992). The CPS data used for the falsification test is non-overlapping with the data used for control group construction (which is based on a different data source) and inference. For Georgia, the p-value for the falsification test (unadjusted) equals 0.32, and the p-value for the falsification test (adjusted) equals 0.33; for Indiana, the p-value for the falsification test (unadjusted) equals 0.14, and the p-value for the falsification test (adjusted) equals 0.27. Therefore, we find no evidence that the data in the prior study period are inconsistent with Assumption 4. Moreover,

a visual inspection of Figure 4-5 shows that the changes in turnout for Georgia and Indiana are reasonably bounded in between by the changes in turnout for their two control groups during 1988-2004, providing additional evidence that Assumption 4 is plausible.

Next, we report the results for sensitivity analysis using the method developed in Section 5.2. We only show the unadjusted version, as the adjusted version is similar. According to Table 3, for the average treatment effect for Georgia in 2008 to be insignificant, we need $\delta_2 = 2.89$, which is the scenario that the potential change in turnout rate for Georgia from 2004 to 2008 in the absence of the voter ID laws is larger than 2.89% plus the maximum of the changes in turnout rate for the two control groups from 2004 to 2008. This scenario is unlikely in practice because the states in control group $b$ were still exposed to heavy campaign efforts in 2008, so the turnout rate change for control group $b$ from 2004 to 2008 is likely larger than the turnout rate change in Georgia in the absence of the voter ID laws. Hence, the significant effect for Georgia in 2008 is not very sensitive to possible violations of Assumption 4. For the average treatment effect for Georgia in 2012 to be insignificant, we need $\delta_2 + \delta_3 = 2.5$, which is the scenario that the potential change in turnout rate in Georgia from 2004 to 2012 in the absence of the voter ID laws is larger than 2.5% plus the maximum of the changes in turnout rate for the two control groups from 2004 to 2008 plus the maximum of the changes in turnout rate for the two control groups from 2008 to 2012. We think this scenario is unlikely in practice via a similar reasoning as above. Hence, the significant effect for Georgia in 2012 is not sensitive to plausible violations of Assumption 4.

# 8 Discussion

The method of difference-in-differences (DID) is widely used to study the effect of interventions in the social and medical sciences. The DID method has the advantage of allowing flexible data structures (e.g., it works for repeated cross sectional studies or longitudinal data) and being able to remove time-invariant systematic differences between the treated and control groups. However, it is also well known that the DID method relies on a strong parallel trends assumption, which requires that in the absence of treatment, the treated and the control groups would experience the same outcome dynamics. To relax the stringent parallel trends assumption, recent work by Hasegawa et al. (2019) outlined a bracketing method, that uses two control groups with different outcome levels in the prior-study period and addresses the bias arising from a historical event interacting with the groups.

In this work, we consider a general strategy for bracketing in DID that addresses the concerns

about the heterogeneity in different units' outcome dynamics in the absence of treatment, and includes the original bracketing method in Hasegawa et al. (2019) as a special case. Critically, we leverage two control groups whose outcomes relative to the treated group exhibit a negative correlation. This negative correlation appears to be reasonable in our voter ID application, as for Georgia, both empirical evidence and domain knowledge suggest that one control group consists of typical battleground states (Michigan, Oregon, Iowa) and the other control group consists of a typical safe state (Wyoming) are likely to have their relative turnout rates being negatively correlated during 2004-2012. A similar pattern is observed when Indiana is the treated group.

Besides our voter ID application, our negative correlation strategy for bracketing in DID could be used in other settings. For example, in studies evaluating the effect of a labor market policy that affects workers in one type of job, a DID study might be done comparing workers in the affected job to workers in other unaffected jobs. Different sectors of the labor market might be negatively correlated relative to the trend of the affected job, enabling bracketing using negative correlations. Another example is in marketing where one might be interested in considering the effect of an advertising campaign for one product and use DID comparing the advertised product to unadvertised products. Products which are substitutes and which are complements for the advertised product might be expected to have negative correlations relative to the demand of the advertised product, enabling bracketing using negative correlations.

Another important contribution in this work is that we develop a novel and easy-to-implement bootstrap inference method to construct uniformly valid confidence intervals for union bounds (i.e., the identified set can be formulated as the union of several intervals). This bootstrap inference method for union bounds has the potential of being applied to broader settings. We apply this new bootstrap inference method to construct confidence intervals for the average treatment effects for the treated and their identified sets. A simple falsification test and sensitivity analysis are also discussed as a means to evaluate the study conclusion.

We demonstrate our bracketing methodology on an application to study the effects of voter identification laws on turnouts for Georgia and Indiana, and we find evidence that the voter identification laws in Georgia increased turnout rates in both 2008 and 2012 and in Indiana in 2008 (when adjusting for covariates). We did not find evidence that the voter identification law changed turnout in Indiana in 2012.

## Supplementary Material

We include the technical proofs and R codes in the supplementary materials.

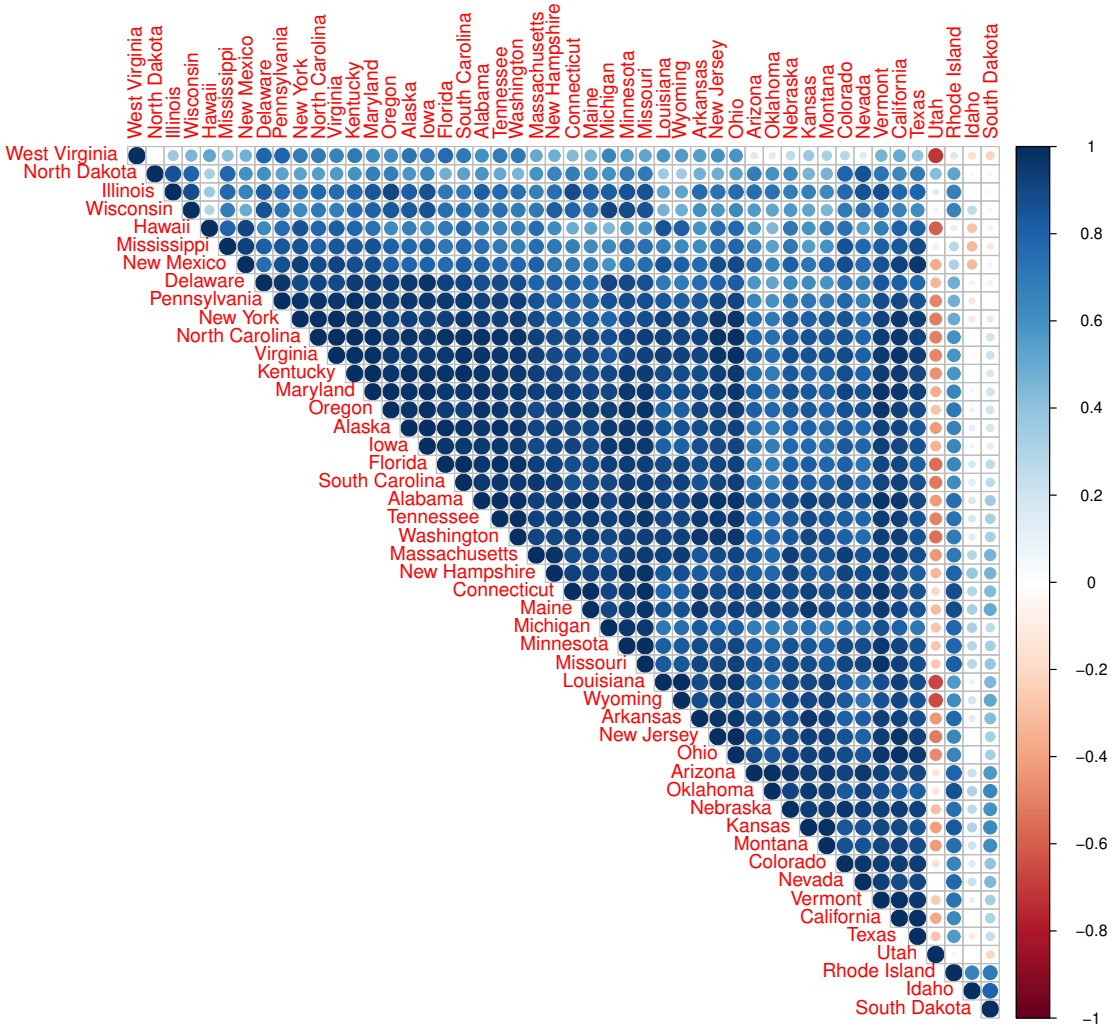


Figure 2: Visualization of the correlation matrix between control states' relative turnout compared with Indiana.

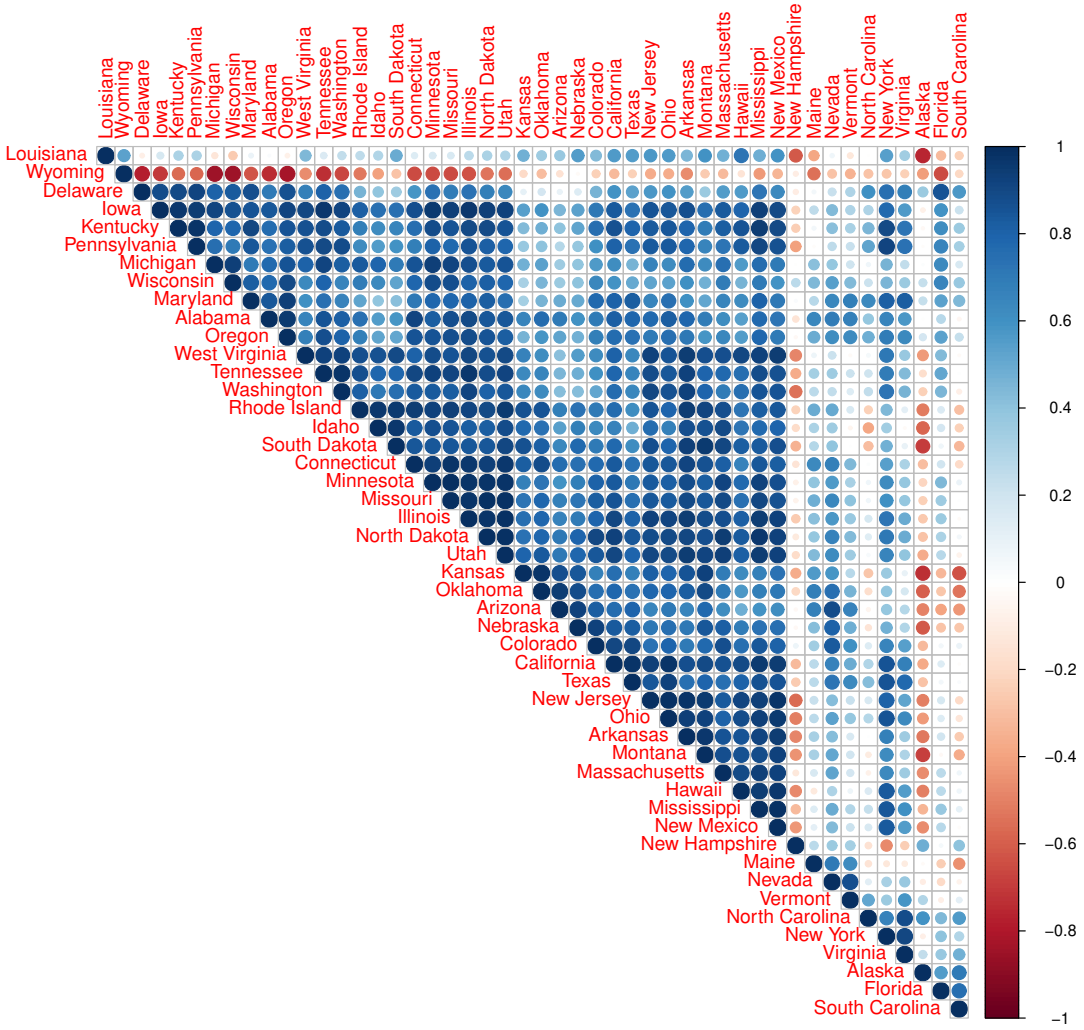


Figure 3: Visualization of the correlation matrix between control states' relative turnout compared with Georgia.

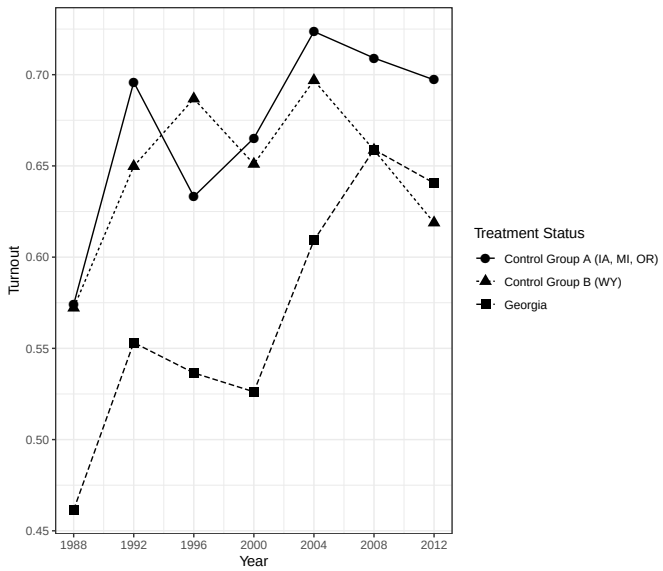


Figure 4: Turnout rates for Georgia and the constructed control groups $a, b$ using the CPS data.

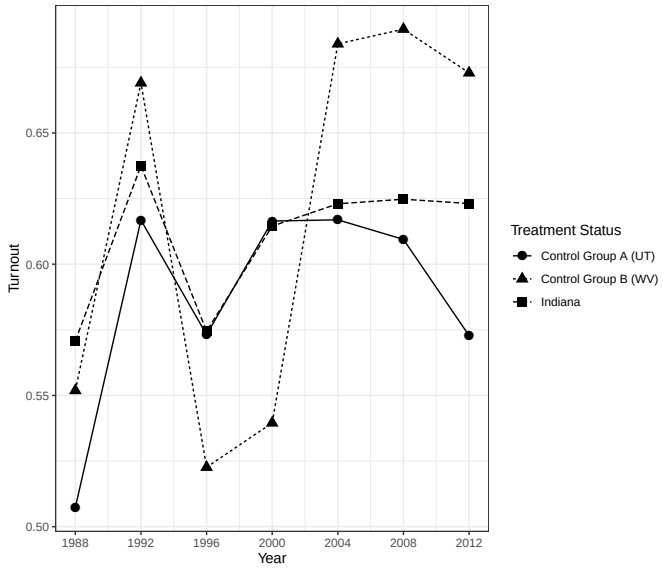


Figure 5: Turnout rates for Indiana and the constructed control groups $a, b$ using the CPS data.

# References


Abadie, A. (2005), 'Semiparametric difference-in-difference estimators', *Review of Economic Studies* **75**(1), 1–19.

Alvarez, R. M., Bailey, D. and Katz, J. N. (2008), The effect of voter identification laws on turnout. Social Science Working Paper 1267R.

Andrews, D. W. K. and Han, S. (2009), 'Invalidity of the bootstrap and the m out of n bootstrap for confidence interval endpoints defined by moment inequalities', *The Econometrics Journal* **12**(s1), S172–S199.

Andrews, D. W. K. and Soares, G. (2010), 'Inference for parameters defined by moment inequalities using generalized moment selection', *Econometrica* **78**(1), 119–157.

Angrist, J. D. and Krueger, A. B. (1999), Empirical strategies in labor economics, *in* O. Ashenfelter and D. Card, eds, 'Handbook of Labor Economics', Vol. 3A, Elsevier Science Publishers, pp. 1277–1366.

Angrist, J. D. and Pischke, J̈.-S. (2009), *Mostly Harmless Econometrics*, Princeton University Press, Princeton, NJ.

Athey, S. and Imbens, G. W. (2006), 'Indentification and inference in nonlinear difference-in-difference models', *Econometrica* **74**(2), 431–497.

Balke, A. and Pearl, J. (1997), 'Bounds on treatment effects from studies with imperfect compliance', *Journal of the American Statistical Association* **92**(439), 1171–1176.

Barreto, M. A., Nuno, S. A. and Sanchez, G. R. (2009), 'The disproportionate impact of voter-id requirements of the electorate-new evidence from indiana.', *PS: Political Science & Politics* **42**(1), 111–116.

Barreto, M. A., Nuño, S., Sanchez, G. R. and Walker, H. L. (2018), 'The racial implications of voter identification laws in america', *American Politics Research* **47**(2), 238–249.

Berger, R. L. (1982), 'Multiparameter hypothesis testing and acceptance sampling', *Technometrics* **24**(4), 295–300.

Bertrand, M., Duflo, E. and Mullainathan, S. (2004), 'How much should we trust differences-in-differences estimates?', *The Quarterly Journal of Economics* **119**(1), 249–275.

Bugni, F. A. (2010), 'Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set', *Econometrica* **78**(2), 735–753.

Bugni, F. A., Canay, I. A. and Shi, X. (2017), 'Inference for subvectors and other functions of partially identified parameters in moment inequality models', *Quantitative Economics* **8**(1), 1–38.

Burden, B. C. (2018), 'Disagreement over id requirements and minority voter turnout', *The Journal of Politics* **80**(3), 1060–1063.

Cai, Z., Kuroki, M., Pearl, J. and Tian, J. (2008), 'Bounds on direct effects in the presence of confounded intermediate variables', *Biometrics* **64**(3), 695–701.

Callaway, B. and Sant'Anna, P. H. C. (2019), Difference-in-differences with multiple time periods. Working Paper.

Campbell, D. T. (1969), Prospective: Artifact and control, *in* R. Rosenthanl and R. Rosnow, eds, 'Artifact in Behavioral Research', Academic Press, New York, NY, pp. 35l–382.

Canay, I. A. (2010), 'El inference for partially identified models: Large deviations optimality and bootstrap validity', *Journal of Econometrics* **156**(2), 408–425.

Chernozhukov, V., Lee, S. and Rosen, A. M. (2009), Intersection bounds: estimation and inference. cemmap working paper CWP19/09.

Chernozhukov, V., Lee, S. and Rosen, A. M. (2013), 'Intersection bounds: estimation and inference', *Econometrica* **81**(2), 667–737.

Chiba, Y. (2017), 'Sharp nonparametric bounds and randomization inference for treatment effects on an ordinal outcome', *Statistics in Medicine* **36**(25), 3966–3975.

Daniels, M. J. and Hogan, J. W. (2008), *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*, CRC Press.

Daw, J. R. and Hatfield, L. A. (2018), 'Matching and regression to the mean in difference-in-differences analysis', *Health Services Research* **53**(6), 4138–4156.

Ding, P. and VanderWeele, T. J. (2016), 'Sensitivity analysis without assumptions', *Epidemiology* **27**(3), 368.

Donald, S. G. and Lang, K. (2007), 'Inference with differences-in-differences and other panel data', *The Review of Economics and Statistics* **89**(2), 221–233.

Erikson, R. S. and Minnite, L. C. (2009), 'Modeling problems in the voter identification-voter turnout debate', *Election Law Journal* **8**(2), 85–101.

Flores, C. A. and Flores-Lagunes, A. (2013), 'Partial identification of local average treatment effects with an invalid instrument', *Journal of Business & Economic Statistics* **31**(4), 534–545.

Fogarty, C. B. (2019), 'Studentized sensitivity analysis for the sample average treatment effect in paired observational studies.', *Journal of the American Statistical Association,* **in press**.

Grimmer, J., Hersh, E., Meredith, M., Mummolo, J. and Nall, C. (2018), 'Comment on 'voter identification laws and the suppression of minority votes", *Journal of Politics* **80**(3), 1045–1051.

Hadar, J. and Russell, W. R. (1969), 'Rules for ordering uncertain prospects', *The American economic review* **59**(1), 25–34.

Hajnal, Z., Kuk, J. and Lajevardi, N. (2018), 'We all agree: Strict voter id laws disproportionately burden minorities', *The Journal of Politics* **80**(3), 1052–1059.

Hajnal, Z., Lajevardi, N. and Nielson, L. (2017), 'Voter identification laws and the suppression of minority votes', *The Journal of Politics* **79**(2), 363–379.

Hasegawa, R. B., Webster, D. W. and Small, D. S. (2019), 'Bracketing in the comparative interrupted time-series design to address concerns about history interacting with group: Evaluating missouri handgun purchaser law', *Epidemiology* **30**(3), 371–379.

Highton, B. (2017), 'Voter identification laws and turnout in the united states', *Annual Review of Political Science* **20**(1), 149–167.

Hirano, K. and Porter, J. R. (2012), 'Impossibility results for nondifferentiable functionals', *Econometrica* **80**(4), 1769–1790.

Hood III, M. and Bullock III, C. S. (2008), 'Worth a thousand words? an analysis of georgia's voter identification statute', *American Politics Research* **36**(4), 555–579.

Hopkins, D. J., Meredith, M., Morse, M., Smith, S. and Yoder, J. (2017), 'Voting but for the law: Evidence from virginia on photo identification requirements', *Journal of Empirical Legal Studies* **14**(1), 79–128.

Horowitz, J. L. and Manski, C. F. (2000), 'Nonparametric analysis of randomized experiments with missing covariate and outcome data', *Journal of the American Statistical Association* **95**(449), 77–84.

Imai, K. (2008), 'Sharp bounds on the causal effects in randomized experiments with "truncation-by-death"', *Statistics & Probability Letters* **78**(2), 144–149.

Imbens, G. M. and Wooldridge, J. M. (2008), 'Recent developments in the econometrics of program evaluation', *Journal of Economic Literature* **47**(1), 5–86.

Imbens, G. W. (2003), 'Sensitivity to exogeneity assumptions in program evaluation', *The American Economic Review Papers and Proceedings* **93**(2), 126–132.

Imbens, G. W. and Manski, C. F. (2004), 'Confidence intervals for partially identified parameters', *Econometrica* **72**(6), 1845–1857.

Jiang, Z. and Ding, P. (2018), 'Using missing types to improve partial identification with application to a study of hiv prevalence in malawi', *Ann. Appl. Stat.* **12**(3), 1831–1852.

Kaido, H., Molinari, F. and Stoye, J. (2019), 'Confidence intervals for projections of partially identified parameters', *Econometrica* **87**(4), 1397–1432.

Manski, C. F. and Pepper, J. V. (2000), 'Monotone instrumental variables: With an application to the returns to schooling', *Econometrica* **68**(4), 997–1010.

Manski, C. F. and Pepper, J. V. (2009), 'More on monotone instrumental variables', *The Econometrics Journal* **12**, S200–S216.

Manski, C. F. and Pepper, J. V. (2017), 'How do right-to-carry laws affect crime rates? coping with ambiguity using bounded-variation assumptions', *The Review of Economics and Statistics* **100**(2), 232–244.

McDonald, M. P. (2020), 'United states elections project, http://www.electproject.org'.

Milyo, J. (2007), The effects of photographic identification on voter turnout in indiana: A county-level analysis. Institute of Public Policy Working Paper.

Mora, R. and Reggio, I. (2012), Treatment effect identification using alternative parallel assumptions. Working Paper 12-33.

Mycoff, J. D., Wagner, M. and Wilson, D. C. (2009), 'The empirical effect of voter-id laws: Present or absent?', *PS: Political Science & Politics* **42**(1), 121–126.

Neyman, J. (1923), 'On the application of probability theory to agricultural experiments. essay on principles. section 9.', *Statistical Science* **5**(4), 465–472. Trans. Dorota M. Dabrowska and Terence P. Speed (1990).

Ostermeier, E. J. (2020), 'Smart politics, https://editions.lib.umn.edu/smartpolitics/2011/02/14/presidential-battleground-stat/'.

Rambachan, A. and Roth, J. (2019), An honest approach to parallel trends. Working Paper.

Richardson, A., Hudgens, M. G., Gilbert, P. B. and Fine, J. P. (2014), 'Nonparametric bounds and sensitivity analysis of treatment effects', *Statist. Sci.* **29**(4), 596–618.

Romano, J. P. and Shaikh, A. M. (2008), 'Inference for identifiable parameters in partially identified econometric models', *Journal of Statistical Planning and Inference* **138**(9), 2786–2807.

Rosenbaum, P. R. (1987*a*), 'The role of a second control group in an observational study', *Statist. Sci.* **2**(3), 292–306.

Rosenbaum, P. R. (1987*b*), 'Sensitivity analysis for certain permutation inferences in matched observational studies', *Biometrika* **74**(1), 13–26.

Rosenbaum, P. R. (2010), *Design of Observational Studies*, Springer-Verlag, New York.

Rubin, D. B. (1974), 'Estimating causal effects of treatments in randomized and nonrandomized studies', *Journal of Educational Psychology* **6**(5), 688–701.

Ryan, A. M., Kontopantelis, E., Linden, A. and Burgess, J. F. (2018), 'Now trending: Coping with non-parallel trends in difference-in-differences analysis', *Statistical Methods in Medical Research* **28**(12), 3697–3711.

Shao, J. (1994), 'Bootstrap sample size in nonregular cases', *Proceedings of the American Mathematical Society* **122**(4), 1251–1262.

Shao, J. and Tu, D. (2012), *The jackknife and bootstrap*, Springer Science & Business Media.

Siddique, Z. (2013), 'Partially identified treatment effects under imperfect compliance: the case of domestic violence', *Journal of the American Statistical Association* **108**(502), 504–513.

Sjölander, A. (2009), 'Bounds on natural direct effects in the presence of confounded intermediate variables', *Statistics in Medicine* **28**(4), 558–571.

Stoye, J. (2009), 'More on confidence intervals for partially identified parameters', *Econometrica* **77**(4), 1299–1315.

Swanson, S. A., Hernán, M. A., Miller, M., Robins, J. M. and Richardson, T. S. (2018), 'Partial identification of the average treatment effect using instrumental variables: Review of methods for binary instruments, treatments, and outcomes', *Journal of the American Statistical Association* **113**(522), 933–947.

Tamer, E. (2010), 'Partial identification in econometrics', *Annu. Rev. Econ.* **2**(1), 167–195.

U.S. Department of Commerce, B. o. t. C. (1988), *CURRENT POPULATION SURVEY: VOTER SUPPLEMENT FILE [Computer File]*, Inter-university Consortium for Political and Social Research, Ann Arbor, MI.

U.S. Department of Commerce, B. o. t. C. (1992), *CURRENT POPULATION SURVEY: VOTER SUPPLEMENT FILE [Computer File]*, Inter-university Consortium for Political and Social Research, Ann Arbor, MI.

U.S. Department of Commerce, B. o. t. C. (2004), *CURRENT POPULATION SURVEY: VOTER SUPPLEMENT FILE [Computer File]*, Inter-university Consortium for Political and Social Research, Ann Arbor, MI.

U.S. Department of Commerce, B. o. t. C. (2008), *CURRENT POPULATION SURVEY: VOTER SUPPLEMENT FILE [Computer File]*, Inter-university Consortium for Political and Social Research, Ann Arbor, MI.

U.S. Department of Commerce, B. o. t. C. (2012), *CURRENT POPULATION SURVEY: VOTER SUPPLEMENT FILE [Computer File]*, Inter-university Consortium for Political and Social Research, Ann Arbor, MI.

Valentino, N. A. and Neuner, F. G. (2017), 'Why the sky didn't fall: Mobilizing anger in reaction to voter id laws', *Political Psychology* **38**(2), 331–350.

Vanderweele, T. J. (2011), 'Controlled direct and mediated effects: Definition, identification and bounds', **38**(3), 551–563.

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y. and Zemia, J. (2017), 'R package "corrplot": visualization of a correlation matrix (version 0.84)'.

Yang, F. and Small, D. S. (2016), 'Using post-outcome measurement information in censoring-by-death problems', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78**(1), 299–318.

# A Proof of (12)

We will prove that the two bounds in 10 and (12) are equivalent. We will only prove the lower bounds for $ATT_t$ in (10) and (12) are equal, i.e., $\sum_{s=2}^{t}\min\{\tau_s(a),\tau_s(b)\} = \min_{g_s\in\{a,b\}}\{\sum_{s=2}^{t}\tau_s(g_s)\}$, the upper bound can be proved similarly.

First, it is easy to see that $\sum_{s=2}^{t}\min\{\tau_s(a),\tau_s(b)\} \geq \min_{g_s\in\{a,b\}}\{\sum_{s=2}^{t}\tau_s(g_s)\}$, because $\sum_{s=2}^{t}\min\{\tau_s(a),\tau_s(b)\} \in \{\sum_{s=2}^{t}\tau_s(g_s) : g_s \in \{a,b\}\}$ and the right hand side is the minimum.

To prove the other direction, because $\min\{\tau_s(a),\tau_s(b)\} \leq \tau_s(g_s)$ for every $s$, we have that $\sum_{s=2}^{t}\min\{\tau_s(a),\tau_s(b)\} \leq \sum_{s=2}^{t}\tau_s(g_s),\ g_s \in \{a,b\}$. Hence, $\sum_{s=2}^{t}\min\{\tau_s(a),\tau_s(b)\} \leq \min_{g_s\in\{a,b\}}\{\sum_{s=2}^{t}\tau_s(g_s)\}$.

# B Proof of Theorem 2

*Proof.* (a) Consider the lower bound. Without loss of generality, assume the bounding parameters $\theta_j$'s are ordered such that $\min_j \theta_j = \theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$. Let $\mathcal{M} = \{j = 1,\ldots,k : \theta_j = \min_{j'}\theta_{j'}\}$ index the $\theta_j$'s that are equal to the minimum. Then, we have

$$\begin{aligned}
\hat{L}(x) &= P_*\left\{\sqrt{N}(\min_j \hat{\theta}_j^* - \min_j \hat{\theta}_j) \leq x\right\} = P_*\left\{\sqrt{N}\min_j(\hat{\theta}_j^* - \min_{j'}\hat{\theta}_{j'}) \leq x\right\} \\
&= 1 - P_*\left\{\sqrt{N}(\hat{\theta}_j^* - \min_{j'}\hat{\theta}_{j'}) > x, j = 1,\ldots,k\right\} \\
&\leq 1 - P_*\left\{\sqrt{N}(\hat{\theta}_m^* - \hat{\theta}_m) > x, m \in \mathcal{M},\ \sqrt{N}(\hat{\theta}_j^* - \hat{\theta}_1) > x, j \notin \mathcal{M}\right\}
\end{aligned}$$

where the third inequality is because $\hat{\theta}_j \geq \min_{j'}\hat{\theta}_{j'}$ for every $j$.

Next, consider $L(x) = P\left\{\sqrt{N}(\min_j \hat{\theta}_j - \theta_1) \leq x\right\}$. Similarly,

$$\begin{aligned}
L(x) &= P\left\{\sqrt{N}(\min_j \hat{\theta}_j - \theta_1) \leq x\right\} = P\left\{\sqrt{N}\min_j(\hat{\theta}_j - \theta_1) \leq x\right\} \\
&= 1 - P\left\{\sqrt{N}\min_j(\hat{\theta}_j - \theta_1) > x\right\} \\
&= 1 - P\left\{\sqrt{N}(\hat{\theta}_m - \theta_m) > x, m \in \mathcal{M},\ \sqrt{N}(\hat{\theta}_j - \theta_1) > x, j \notin \mathcal{M}\right\}
\end{aligned}$$

Under conditions in Theorem 2, $\sqrt{N}(\bar{\boldsymbol{X}} - \boldsymbol{\mu})$ is asymptotically normally distributed uniformly in $P \in \mathcal{P}$, and thus $\sqrt{N}(\hat{\theta}_j - \theta_j), j = 1,\ldots,k$ jointly are asymptotically normally distributed uniformly in $P \in \mathcal{P}$ by delta method. Hence, $P_*\left\{\sqrt{N}(\hat{\theta}_m^* - \hat{\theta}_m) > x, m \in \mathcal{M},\ \sqrt{N}(\hat{\theta}_j^* - \hat{\theta}_1) > x, j \notin \mathcal{M}\right\}$ is strongly consistent for $P\left\{\sqrt{N}(\hat{\theta}_m - \theta_m) > x, m \in \mathcal{M},\ \sqrt{N}(\hat{\theta}_j - \theta_1) > x, j \notin \mathcal{M}\right\}$ based on standard bootstrap theory, e.g., see Shao and Tu (2012). Therefore, for any $\epsilon > 0$, there exists an $N_{01}$ s.t. for $N > N_{01}$, $\sup_P \sup_x[\hat{L}(x) - L(x)] \leq \epsilon$. Because $\epsilon$ can be arbitrarily small, it leads to the conclusion that $\lim_{N\to\infty}\sup_P \sup_{x\in\mathcal{R}}\{\hat{L}(x) - L(x)\} \leq 0$. The statement for $\hat{R}(x)$ can be similarly proved and is omitted.

(b) From the definition of $c_L^*(p)$, it satisfies $\hat{L}(c_L^*(p)) \geq p$. From part (a), for any $\epsilon > 0$, there exists an $N_{01}$ s.t. for $N > N_{01}$, $\sup_P \sup_x\{\hat{L}(x) - L(x)\} \leq \epsilon$, and for every $P$,

$$\begin{aligned}
L(c_L^*(p)) &\geq \hat{L}(c_L^*(p)) - \sup_x\{\hat{L}(x) - L(x)\} \\
&\geq p - \epsilon
\end{aligned}$$

This completes the proof of the first part in (b).

From the definition of $c_U^*(1-p)$, it satisfies $\hat{R}(c_U^*(1-p)) \leq 1 - p$. From part (a), for any

$\epsilon > 0$, there exists an $N_{02}$ s.t. for $N > N_{02}$, $\inf_P \inf_x \{\hat{R}(x) - R(x)\} \geq -\epsilon$, and for every $P$,

$$\begin{aligned} R(c_U^*(1-p)) &\leq \hat{R}(c_U^*(1-p)) - \inf_x \{\hat{R}(x) - R(x)\} \\ &\leq 1 - p + \epsilon \end{aligned}$$

Therefore,

$$P\left\{\sqrt{N}(\max_j \hat{\theta}_j - \max_j \theta_j) \geq c_U^*(1-p)\right\} \geq 1 - R(c_U^*(1-p)) \geq p - \epsilon$$

This completes the proof of (b).

(c) Let $p = 1 - \alpha/2$ and rearrange,

$$\begin{aligned} \lim_{N\to\infty} \inf_P P\left\{\min_j \hat{\theta}_j \leq \min_j \theta_j + N^{-1/2} c_L^*(1-\alpha/2)\right\} \geq 1 - \alpha/2 \\ \lim_{N\to\infty} \inf_P P\left\{\max_j \hat{\theta}_j \geq \max_j \theta_j + N^{-1/2} c_U^*(\alpha/2)\right\} \geq 1 - \alpha/2 \end{aligned}$$

By Bonferroni's inequality, we have

$$\begin{aligned} &P\left\{[\min_j \theta_j, \max_j \theta_j] \in CI_{1-\alpha}\right\} \\ &\geq 1 - P\left\{\min_j \theta_j < \min_j \hat{\theta}_j - N^{-1/2} c_L^*(1-\alpha/2)\right\} - P\left\{\max_j \theta_j > \max_j \hat{\theta}_j - N^{-1/2} c_U^*(\alpha/2)\right\} \\ &= P\left\{\min_j \theta_j \geq \min_j \hat{\theta}_j - N^{-1/2} c_L^*(1-\alpha/2)\right\} + P\left\{\max_j \theta_j \leq \max_j \hat{\theta}_j - N^{-1/2} c_U^*(\alpha/2)\right\} - 1 \end{aligned}$$

Therefore,

$$\lim_{n\to\infty} \inf_P P\left\{[\min_j \theta_j, \max_j \theta_j] \in CI_{1-\alpha}\right\} \geq 1 - \alpha/2 + 1 - \alpha/2 - 1 = 1 - \alpha$$

(d) Define $p^\psi = 1 - \Phi\{\rho(\max_j \theta_j - \min_j \theta_j)\}\alpha$. From the condition, we have that $\hat{p} = p^\psi + o_p(1)$.

Decompose the probability that $\psi_0$ is outside $CI_{1-\alpha}^\psi$ as

$$P\left\{\psi_0 \notin CI_{1-\alpha}^\psi\right\} \leq \underbrace{P\left\{\psi_0 < \min_j \hat{\theta}_j - N^{-1/2} c_L^*(\hat{p})\right\}}_{A_L} + \underbrace{P\left\{\psi_0 > \max_j \hat{\theta}_j - N^{-1/2} c_U^*(1-\hat{p})\right\}}_{A_U}$$

Because with probability approaching 1, we have $\hat{p} \geq p^\psi - \epsilon/2$, and thus $c_L^*(\hat{p}) \geq c_L^*(p^\psi - \epsilon/2)$ and $c_U^*(1-\hat{p}) \leq c_U^*(1-p^\psi+\epsilon/2)$ because $c_L^*(p)$ and $c_U^*(p)$ are both non-decreasing functions of $p$. Hence, the first component satisfies

$$A_L = P\left\{\psi_0 + N^{-1/2} c_L^*(\hat{p}) < \min_j \hat{\theta}_j\right\} \leq \underbrace{P\left\{\psi_0 + N^{-1/2} c_L^*(p^\psi - \epsilon/2) < \min_j \hat{\theta}_j\right\}}_{\tilde{A}_L} + o(1)$$

the second component satisfies

$$A_U = P\left\{\psi_0 + N^{-1/2} c_U^*(1-\hat{p}) > \max_j \hat{\theta}_j\right\} \leq \underbrace{P\left\{\psi_0 + N^{-1/2} c_U^*(1-p^\psi+\epsilon/2) > \max_j \hat{\theta}_j\right\}}_{\tilde{A}_U} + o(1)$$

In the following, we will show that for any $\epsilon > 0$, there exists an $N_0$, s.t. for $N > N_0$, $\tilde{A}_L + \tilde{A}_U \leq \alpha + \epsilon$. Define $\lambda$ as the limit: $\rho(\max_j \theta_j - \min_j \theta_j) \to \lambda \in [0, \infty]$.

Suppose first $\lambda = 0$ and $p^\psi = 1 - \alpha/2 + o(1)$. For the same $\epsilon$, there exists an $N_{04}$, s.t. for $N > N_{04}$, $p^\psi \geq 1 - \alpha/2 - \epsilon/2$. In this case,

$$\tilde{A}_L \leq P\left\{\min_j \theta_j + N^{-1/2} c_L^*(p^\psi - \epsilon/2) < \min_j \hat{\theta}_j\right\} \leq \alpha/2 + \epsilon$$

where the first inequality is because $\psi_0 \geq \min_j \theta_j$, the second inequality uses $p^\psi \geq 1 - \alpha/2 - \epsilon/2$ and Theorem 2(b). Similarly,

$$\tilde{A}_U \leq P\left\{\max_j \theta_j + N^{-1/2} c_U^*(1 - p^\psi + \epsilon/2) > \max_j \hat{\theta}_j\right\} \leq \alpha/2 + \epsilon$$

Hence, when $N > N_{04}$, $\lambda = 0$, we have $\tilde{A}_L + \tilde{A}_U \leq \alpha + 2\epsilon$.

Then, consider $\lambda \in (0, \infty]$ and $p^\psi = 1 - \Phi(\lambda)\alpha + o(1)$. For the same $\epsilon$, there exists an $N_{05}$, s.t. for $N > N_{05}$, $p^\psi \geq 1 - \Phi(\lambda)\alpha - \epsilon/2$. Also since $\lambda > 0$, $N^{1/2}(\max_j \theta_j - \min_j \theta_j) \to \infty$. Without loss of generality, assume $N^{1/2}(\psi_0 - \min_j \theta_j) \to \infty$. Under this circumstance,

$$\begin{aligned}
\tilde{A}_L &= P\left\{\sqrt{N}(\min_j \hat{\theta}_j - \min_j \theta_j) > c_L^*(p^\psi - \epsilon/2) + N^{1/2}(\psi_0 - \min_j \theta_j)\right\} \\
&\leq P\left\{\sqrt{N}(\hat{\theta}_1 - \theta_1) > c_L^*(p^\psi - \epsilon/2) + N^{1/2}(\psi_0 - \min_j \theta_j)\right\} \qquad (\theta_1 = \min_j \theta_j) \\
&= o(1) \\
\tilde{A}_U &= P\left\{\psi_0 - \max_j \theta_j + N^{-1/2} c_U^*(1 - p^\psi + \epsilon/2) > \max_j \hat{\theta}_j - \max_j \theta_j\right\} \\
&\leq P\left\{N^{1/2}(\max_j \hat{\theta}_j - \max_j \theta_j) < c_U^*(1 - p^\psi + \epsilon/2)\right\} \\
&\leq P\left\{N^{1/2}(\max_j \hat{\theta}_j - \max_j \theta_j) < c_U^*\left(\Phi(\lambda)\alpha + \epsilon\right)\right\} \leq \Phi(\lambda)\alpha + \epsilon
\end{aligned}$$

For the same $\epsilon$, there exists an $N_{06}$, when $N > N_{06}$, $\tilde{A}_L \leq \epsilon$. Hence, when $N > \max(N_{05}, N_{06})$, $\tilde{A}_L + \tilde{A}_U \leq \Phi(\lambda)\alpha + 2\epsilon$.

Combined, we have for any $\epsilon > 0$, for $N > \max(N_{04}, N_{05}, N_{06})$, $P\left\{\psi_0 \notin CI_{1-\alpha}^\psi\right\} \leq A_L + A_U \leq \alpha + 2\epsilon$, which completes the proof for Theorem 2(d). □

# C Proof of Theorem 3

(a) Define $ATT_1 = 0$. From (8)-(9), we have for every $s \geq 2$,

$$\begin{aligned}
\min(\tau_s(a), \tau_s(b)) &= ATT_s - ATT_{s-1} + \Delta_s^{(0)}(trt) - \max(\Delta_s^{(0)}(a), \Delta_s^{(0)}(b)) \\
\max(\tau_s(a), \tau_s(b)) &= ATT_s - ATT_{s-1} + \Delta_s^{(0)}(trt) - \min(\Delta_s^{(0)}(a), \Delta_s^{(0)}(b))
\end{aligned}$$

Then, from Assumption 5,

$$\begin{aligned}
\min(\tau_s(a), \tau_s(b)) &\leq ATT_s - ATT_{s-1} + \delta_s \\
\max(\tau_s(a), \tau_s(b)) &\geq ATT_s - ATT_{s-1} - \gamma_s
\end{aligned}$$

Hence,

$$\begin{aligned}
\min(\tau_2(a), \tau_2(b)) - \delta_2 \leq \quad & ATT_2 \quad && \leq \max(\tau_2(a), \tau_2(b)) + \gamma_2 \\
\min(\tau_2(a), \tau_2(b)) - \delta_s \leq \quad & ATT_s - ATT_{s-1} && \leq \max(\tau_2(a), \tau_2(b)) + \gamma_s
\end{aligned}$$

Theorem 3(a) is proved by summing these inequalities.
(b) Let $[\hat{l}_t, \hat{r}_t]$ be the confidence interval for the identified set under Assumption 4. Then,

$$
\begin{aligned}
&P\left(\left[\sum_{s=2}^{t}\min\{\tau_s(a),\tau_s(b)\}, \sum_{s=2}^{t}\max\{\tau_s(a),\tau_s(b)\}\right] \in [\hat{l}_t, \hat{r}_t]\right) \\
=&P\left(\sum_{s=2}^{t}\min\{\tau_s(a),\tau_s(b)\} \geq \hat{l}_t, \sum_{s=2}^{t}\max\{\tau_s(a),\tau_s(b)\} \leq \hat{r}_t\right) \\
=&P\left(\sum_{s=2}^{t}\min\{\tau_s(a),\tau_s(b)\} - \sum_{s=2}^{t}\delta_s \geq \hat{l}_t - \sum_{s=2}^{t}\delta_s, \sum_{s=2}^{t}\max\{\tau_s(a),\tau_s(b)\} + \sum_{s=2}^{t}\gamma_s \leq \hat{r}_t + \sum_{s=2}^{t}\gamma_s\right) \\
=&P\left(\left[\sum_{s=2}^{t}\min\{\tau_s(a),\tau_s(b)\} - \sum_{s=2}^{t}\delta_s, \sum_{s=2}^{t}\max\{\tau_s(a),\tau_s(b)\} + \sum_{s=2}^{t}\gamma_s\right] \in [\hat{l}_t - \sum_{s=2}^{t}\delta_s, \hat{r}_t + \sum_{s=2}^{t}\gamma_s]\right)
\end{aligned}
$$

The uniform validity of $[\hat{l}_t - \sum_{s=2}^{t}\delta_s, \hat{r}_t + \sum_{s=2}^{t}\gamma_s]$ is directly from the uniform validity of $[\hat{l}_t, \hat{r}_t]$.

Next, we prove the results hold for the parameter of interest $ATT_t$, where $ATT_t$ lies in the identified set (16). This proof is based on and is similar to the proof of Theorem 2(d), and thus some details are omitted. Let $[\hat{l}_t, \hat{r}_t]$ be the confidence interval for $ATT_t$ under Assumption 4. Under Assumption 5, decompose the probability that $ATT_t$ is outside $[\hat{l}_t - \sum_{s=2}^{t}\delta_s, \hat{r}_t - \sum_{s=2}^{t}\gamma_s]$ as

$$
P\left(ATT_t \notin \left[\hat{l}_t - \sum_{s=2}^{t}\delta_s, \hat{r}_t - \sum_{s=2}^{t}\gamma_s\right]\right) \leq P\left(ATT_t < \hat{l}_t - \sum_{s=2}^{t}\delta_s\right) + P\left(ATT_t > \hat{r}_t + \sum_{s=2}^{t}\gamma_s\right)
$$

Consider the two scenarios when constructing $[\hat{l}_t, \hat{r}_t]$ as in the proof of Theorem 2(d). Define $\lambda$ still as the limit based on the identified set in (10),

$$
\rho\left[\sum_{s=2}^{t}\max\{\tau_s(a),\tau_s(b)\} - \sum_{s=2}^{t}\min\{\tau_s(a),\tau_s(b)\}\right] \to \lambda \in [0,\infty].
$$

First, consider $\lambda = 0$, that is when $p^{\psi} = 1 - \alpha/2 + o(1)$ in constructing $[\hat{l}_t, \hat{r}_t]$. In this scenario,

$$
\begin{aligned}
&P\left(ATT_t < \hat{l}_t - \sum_{s=2}^{t}\delta_s\right) \leq P\left(\sum_{s=2}^{t}\min\left\{\tau_s(a),\tau_s(b)\right\} - \sum_{s=2}^{t}\delta_s < \hat{l}_t - \sum_{s=2}^{t}\delta_s\right) \\
&= P\left(\sum_{s=2}^{t}\min\left\{\tau_s(a),\tau_s(b)\right\} < \hat{l}_t\right) \leq \alpha/2 + o(1)
\end{aligned}
$$

Similarly,

$$
P\left(ATT_t > \hat{r}_t + \sum_{s=2}^{t}\gamma_s\right) \leq \alpha/2 + o(1)
$$

Consider the second scenario when $\lambda \in (0,\infty]$ and $p^{\psi} = 1 - \Phi(\lambda)\alpha + o(1)$ in constructing $[\hat{l}_t, \hat{r}_t]$. Without loss of generality, assume $N^{1/2}[ATT_t - \{\sum_{s=2}^{t}\min\left\{\tau_s(a),\tau_s(b)\right\}\}] \to \infty$, and

thus, $N^{1/2}[ATT_t + \sum_{s=2}^{t} \delta_s - \{\sum_{s=2}^{t} \min\{\tau_s(a), \tau_s(b)\}\}] \to \infty$. Under this circumstance,

$$
\begin{aligned}
&P\left(ATT_t < \hat{l}_t - \sum_{s=2}^{t} \delta_s\right) \leq P\left(ATT_t + \sum_{s=2}^{t} \delta_s < \hat{l}_t\right) = o(1) \\
&P\left(ATT_t > \hat{r}_t + \sum_{s=2}^{t} \gamma_s\right) \leq P\left(\sum_{s=2}^{t} \max\{\tau_s(a), \tau_s(b)\} + \sum_{s=2}^{t} \gamma_s > \hat{r}_t + \sum_{s=2}^{t} \gamma_s\right) \\
=&P\left(\sum_{s=2}^{t} \max\{\tau_s(a), \tau_s(b)\} > \hat{r}_t\right) \leq \Phi(\lambda)\alpha + o(1)
\end{aligned}
$$

Combining two scenarios and using similar arguments as in the proof of theorem 2(d), uniform validity is established, that is

$$
\lim_{N\to\infty} \inf_{P\in\mathcal{P}} \inf_{ATT_t} P\left(ATT_t \notin \left[\hat{l}_t - \sum_{s=2}^{t} \delta_s, \hat{r}_t + \sum_{s=2}^{t} \gamma_s\right]\right) \leq \alpha
$$