Bracketing Bounds for Differences-in-Differences with an Application to Voter ID Laws

Luke Keele Raiden Hasegawa Dylan Small

July 11, 2019

Abstract

The method of differences-in-differences (DID) is widely used in political science to estimate causal effects. The primary advantage of DID is that it can account for time-invariant bias from unobserved confounders. However, the standard DID estimator will be biased if there is an interaction between history in the after period and the groups, That is, bias will be present if an event besides the treatment occurs at the same time and affects treated group in a differential fashion. We present a method of bounds based on DID that accounts for an unmeasured confounder that has a different effect in the post-treatment time period. This form of bounds is simple to implement and only requires partitioning the controls into two separate groups. We also develop a new falsification test to probe the key assumption that is necessary for the bounds estimator to provide consistent estimates of the treatment effect. We apply these DID bounds to data on the effect of voter identification laws on turnout. Specifically, we focus on the enactment of voter identification laws in Georgia and Indiana. We find evidence that new voter ID laws in those states actually increased turnout.

1 Introduction

1.1 Causal Effects of Political Institutions

One defining feature of political science is the recognition that the design of political institutions has important effects on both elite and mass behavior. No where is this fact more salient than in the study of voter turnout. In the United States, state and local governments retain considerable latitude over voting procedures, and a central question is whether these governments use voting procedures to alter the composition of the electorate in hopes of shaping electoral outcomes. In this vein, there has been considerable debate over whether voter identification (ID) requirements effect turnout. The Help America Vote Act of 2002 primarily focused on voting technology in the aftermath of the 2000 presidential election, but also set Federal minimum voter identification requirements. A number of states subsequently adopted stricter voter identification requirements. The most controversial of these laws require all voters to present some form of photo identification in order to vote. Empirical assessment of voter ID laws has been decidedly mixed with different studies alternately concluding that these laws have reduced turnout, had no effect at all, or possibly increased turnout (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). Adding to the controversy surrounding the adoption of these laws, scholars have noted partisan voting patterns in the passage of these laws and leveled charges of racial discrimination (Davidson 2009; Sobel 2009).

1.2 Extant Identification Strategies

While one can identify several reasons why empirical results on voter ID laws are mixed, one critical reason is, undoubtedly, that drawing causal inferences in settings of this type is quite difficult. Several key limitations make it particularly difficult to make causal inferences about voter ID laws. First, randomization of voting requirements is nonexistent. Second, natural experiments are quite rare in this setting. States tend to implement voting requirements for all citizens and there is rarely anything haphazard about the introduction of these laws, which often rules out research designs based on natural experiments. Third, states adopt these laws in a highly purposeful fashion to influence voter turnout levels. As such, the exact reason one state adopts a particular set of voting requirements almost certainly depends on unobservable factors.

The extant literature on turnout, and on voter ID more specifically, has relied predominantly

on two empirical strategies that are both ill-equipped to overcome these limitations. In the first approach, analysts assume that adoption of voter ID laws is based on 'selection on observables" (Barnow et al. 1980). That is, one must assume that one can observe and measure all the reasons that a state like Indiana adopts a strict voter ID law while other states did not. Here, one must ask how could it be that two voters that are identical in all meaningful background characteristics nonetheless receive different voter ID requirements? This approach would seem implausible given that most data sets contain a relatively limited set of covariates on which to make voters comparable.

The primary alternative to selection on observables has been the differences-in-differences (DID) identification strategy. DID exploits longitudinal variation to remove time-invariant bias from unobserved confounders. More specifically, DID requires observing treated and control data before and after the intervention occurs. One then calculates the treated over time change in the outcome and the control over time change in the outcome. The estimated treatment effect is the difference between the treated outcome trend and the control outcome trend. That is, one takes the difference in the outcomes after removing the time trends for the treated and control groups. Conveniently, DID can be implemented via a standard linear regression model, which allows for the investigator to control for factors that might influence the over time evolution of the treated and control groups. The simple form of the estimator also makes DID very flexible in terms of data requirements: the investigator does even need to observe the same voters over time. For example, survey data may be collected before and after a treatment goes into effect. Even if the survey data is not collected from the same individuals before and after treatment, one can still apply DID.

For DID to produce a valid treatment effect estimate, investigators must assume that the outcomes in the treated and control units are, absent treatment, evolving the same way over time. That is, one must assume that turnout in a voter ID state like Indiana would have evolved the same way over time as turnout in non-voter ID states. In other words, only the treatment alters the trajectory of turnout in the treated state. One key threat to the DID identification strategy is the possibility of selective maturation. Selective maturation refers to a situation where an event occurs concurrently with intervention (or shortly thereafter) and effects control and treatment groups differentially (Campbell and Stanley 1963). Possible bias due to selective maturation would appear to be a important threat in many studies of political institutions especially those seeking to understand how changes in voter registration affect turnout.

Let's assume we have two states, and one adopts voter ID and the other does not. In a

presidential election, some states are battleground states due to the competitive nature of the presidential election in that state. Voters in battleground states will almost certainly be exposed to mobilization efforts by the campaigns, while voters in non-battleground states may be subject to few, if any, mobilization efforts. These mobilization efforts tend to vary from one election to the next. In the most extreme case, one state may be a battleground state in one election and not in the next. Variation in mobilization in the treated state that occurs in the election after the adoption of voter ID will raise the strong possibility of bias due to selective maturation. However, a range of changes in the electoral environment could also produce bias from selective maturation. For example, a competitive Senate race or ballot initiative are all events that could also be concurrent with the change in voter ID laws and affect the treated state differentially.

In short, the two standard research designs used to understand the effects of interventions like voter ID laws are particularly ill-suited to empirical circumstances. Here, we introduce a method of bounds that is better suited to applications of this type. Our approach is based on using DID estimates to construct plausible bounds. More specifically, we use a method developed in Hasegawa et al. (2019) which imposes weaker assumptions than a traditional DID approach. In this paper, we further develop that method by deriving a set of falsification tests that allow for probing the key identification assumptions.

2 Difference-in-Difference Bracketing

2.1 Notation and Estimator

We begin with some basic notation. Given an observational unit (e.g., a citizen) and a set of conditions called treatment and control (whether he or she must show photo ID in order to vote), we associate with each unit and condition a potential outcome (whether she turned out). These outcomes are potential because only one will ultimately be realized and observed. For each citizen, we only observe whether she turned out or not but not both potential outcomes. For each unit i, the treatment indicator $W_i = w$, $w \in \{0, 1\}$, records which potential outcome is actually observed. Here, we define treatment status as voting under either a strict voter ID requirement in Indiana or Gerorgia or not. The outcome Y_i is a binary indicator for whether citizen i voted or not. We write the potential outcomes as Y_{iw}. Since the treatment is binary, there are two potential outcomes for each unit: Y_{i1} and Y_{i0}. The first is the potential outcome under treatment, and the second is the potential outcome without it. Under the Rubin causal model, the causal effect is defined as $Y_{i1} - Y_{i0}$. That is, the definition of a causal effect depends on the potential outcomes rather than the outcome that is actually observed. The fundamental

problem of causal inference as defined by Rubin is that, for any single unit, at most one of the potential outcomes can be revealed (Holland 1986). While the causal effect is well defined, we cannot learn its value from one realized outcome. For example, we will never observe Y_{i0} if i lives in a state with a photo ID requirement, and we will never observe Y_{i1} if i lives in a state without that requirement. By combining the potential outcomes with the treatment indicator, we can define the observed outcome for each unit i:

$$Y_i = W_i Y_{i1} + (1 - W_i) Y_{i0}. \tag{1}$$

Next, we denote time periods with $t \in \{1, 2\}$, where 1 indicates a time period before the treatment and 2 indicates a time period after the treatment has been administered. The response observed from unit i in each time period is $Y_i^{(t)}$. Next, we separate the control units into two groups that we designate upper and lower controls. For $Z_i^{(t)} = 1$ denotes the treated unit (i.e. $W = 1$) and $Z_i^{(t)} = 0$ denotes control units with higher outcomes in a time period or time periods before t. Hereafter, we refer to $Z_i^{(t)} = 0$ units as upper controls. Next, $D_i^{(t)} = 1$ denotes the treated unit and $D_i^{(t)} = 0$ denotes control units with lower outcomes in a time period or time periods before t. We refer to $D_i^{(t)} = 0$ units as lower controls. Note that $Z_i^{(t)} = 1$ if and only if $D_i^{(t)} = 1$. This notation is meant to capture the fact that we have partitioned the units for whom $W_i = 0$ into two separate groups using outcomes from a time period before $t = 1$. Below, we discuss empirical strategies for how one might partition the control units into the lower and upper groups.

The logic behind this partition is as follows and is based on ideas in Campbell (2009). We concerned that the treated and control group differ on U on unobserved covariate. Here, we are creating two control groups that are assumed to differ systematically on U. That is, we assume that values of U tend to be larger for the upper controls–states with $Z_i^{(t)} = 0$–and for lower controls–states with $D_i^{(t)} = 0$. If true, the effect of U on the treated group is bracketed by its effect on the two control groups. If this bracketing relationship holds, an estimated treatment effect that is significantly higher relative to *both* control groups cannot plausibly be explained by a bias due to U. Next, we review a DID bracketing estimator that exploits this possibility..

Hereafter, we drop the notation for t from the treatment indicators, since we assume the intervention is only present when $t = 2$. We then define two different DID estimates:

$$\tau_{uc} = \Big(\mathbb{E}[Y^{(2)}|Z_i = 1] - \mathbb{E}[Y^{(1)}|Z_i = 1]\Big) - \Big(\mathbb{E}[Y^{(2)}|Z_i = 0] - \mathbb{E}[Y^{(1)}|Z_i = 0]\Big)$$

$$\tau_{lc} = \Big(\mathbb{E}[Y^{(2)}|D_i = 1] - \mathbb{E}[Y^{(1)}|D_i = 1]\Big) - \Big(\mathbb{E}[Y^{(2)}|D_i = 0] - \mathbb{E}[Y^{(1)}|D_i = 0]\Big)$$

Here, τ_{uc} represents the treatment effect for the treated unit relative to the upper controls, and τ_{lc} represents the treatment effect for the treated unit relative to the lower controls. Subject to a set of assumptions that we discuss next, we can bound the true causal effect using the following equation:

$$\min\{\hat{\tau}_{lc}, \hat{\tau}_{uc}\} \leq \tau \leq \max\{\hat{\tau}_{lc}, \hat{\tau}_{uc}\} \tag{2}$$

The bounds are based on the upper and lower the estimates from the two different DID estimates using the upper and lower controls. Inference for these bounds is straightforward to calculate. The $1 - \alpha$ confidence interval for the bound in equation (2) can be found by taking the minimum and maximums of the $1 - \alpha$ confidence intervals for $\hat{\tau}_{lc}$ and $\hat{\tau}_{uc}$. See Hasegawa et al. (2019) for details. These DID bounds removes history-by-group interaction bias arising from an unobserved confounder that has an increasing or decreasing effect on voter turnout over time. One key advantage of these bounds is the simple form used for estimation and inference: simple DID estimators can be applied to estimate both the bounds and associated confidence intervals. Next, we review the assumptions that must hold for these bounds to bracket the true causal effect.

2.2 Key Assumptions

For the DID bracketing estimator to bound the true causal effect, a set of identification assumptions must hold. Here, we review those assumptions in the context of voter ID laws. We do not present formal versions of these assumptions as those can be found elsewhere (Hasegawa et al. 2019). Critically, we focus evaluating the plausibility of these assumptions within the context of the voter ID application. Ideally, these assumptions should have greater plausibility than the usual DID identification conditions.

First, we introduce some notation. We use $\mathbf{U}$ to denote a set of time invariant confounders. The key identification restrictions all impose constraints on $\mathbf{U}$. Evaluating the plausibility of the assumptions is greatly aided if the investigator outlines likely elements of $\mathbf{U}$. We would argue that the most critical set of time invariant confounders are state level factors that contribute to voter turnout but do not change or change very slowly from election to election. One key set of factors in $\mathbf{U}$ would be the ease of the voter registration process at the state level. In some states, this process is relatively straightforward, while in other states, considerably more effort

by voters is required to register to vote. We would also include state political culture as another key time invariant confounder. Thus we assume that $\mathbf{U}$ contains both of these key confounders: one which measures the ease of voting in a state with higher values representing an easier voting process and one which represents the political culture in a state where higher values represent a more engaged political culture.

The first assumption required for the DID bracketing method is that it must be the case that higher values of $\mathbf{U}$ correspond to higher outcomes. This assumption would appear to be quite plausible in the context of the voter ID. That is, higher values of our two key scales, ease of voting and political culture, should be correspond to higher turnout. Under the second assumption, it must be the case that the distribution of $\mathbf{U}$ is time invariant within the treated and control groups. This assumption would be violated if values of $\mathbf{U}$ varied over time within the treated and control groups. In the voter ID context, this implies that we are assuming that the distribution of $\mathbf{U}$ is time invariant within the voter ID states and the non-voter id states. That is, the distribution of things such as state political culture within the upper and lower controls and the voter ID states should be time invariant during the study period. This assumption is also plausible especially given a relatively short study time period. That is, take state political culture. Under this assumption, the distribution of this unobserved quantity needs to be fixed over a four year time period. While political culture may change from one election to the next, in most cases it is reasonable to view this as fixed.

Next, we must assume the distribution of $\mathbf{U}$ with respect to the treated, lower and upper controls can be stochastically ordered. The distribution of a random variable A is said to be higher than the distribution of random variable B if A stochastically dominates B (Hadar and Russell 1969). This implies that the lowest values of $\mathbf{U}$ are present in the lower control group, medium values are found in the treated group and, the highest values of $\mathbf{U}$ are found in the upper control group. For example, this implies that the distribution of confounders such as political culture and the ease of voting should be lowest in the control states with the lowest turnout and highest in the control states with the highest turnout. The distribution of $\mathbf{U}$ is lowest in the lower control group, $D_i^{(t)} = 0$, medium in the treated group, and highest in the upper control group $Z_i^{(t)} = 0$. Finally, we assume that higher values of $\mathbf{U}$ either have a larger or smaller effect over time over the whole range of $\mathbf{U}$ or have a smaller effect over the range of $\mathbf{U}$. Thus the effect of confounders such as political culture and the ease of voting should have larger effects across time for all units. That is, if the effect of political culture changes it should have a larger or smaller effect across all three groups.

Thus far, we have only focused on instances where the structure of **U** only contains time-invariant confounders. Critically, DID bracketing can bound the true causal effect even if time-varying confounders are present under the following condition. If higher values of **U** have a larger effect over time over the whole range of **U** then the distributional shifts in **U** over time must be more positive for higher starting distributions of **U**. For instance, it may be likely that a larger mobilization efforts have a greater effect over time than a smaller mobilization effort. If larger mobilization efforts beget larger subsequent efforts, then this condition holds and the DID bracket will bound the true causal effect even when voter mobilization is allowed to vary over time and across group. In general, for bias from time invariant confounders to be minimal, we must assume that time-varying confounders have the same distribution within similar groups over time. Again, this assumption is relatively plausible in the voter ID context. Of course, the plausibility of these assumptions will vary from context to context. In general, they would appear to be reasonable in the voter ID context.

2.3 Placebo Testing Based Inference

While the DID bracketing approach relies on a different identification assumptions than DID, it is still prone to a number of issues which can contribute to underestimation of standard errors in DID estimators. It is well understood that regression estimators with fixed effects may understate standard errors in the presence of serially correlated data, yearly state-level shocks, or a small number of policy changes (Bertrand et al. 2004; Donald and Lang 2007; Conley and Taber 2011). As such, the confidence intervals for the DID bounds, when based on regression estimators, may be too narrow.

One alternative method of inference is to construct a distribution of placebo estimates, and compare the DID bracketing estimates to this distribution of placebo estimates (Hasegawa et al. 2019). To construct a placebo distribution, we employed the following set of steps. First, we restricted the data to the 46 states that did not adopt voter ID laws in 2008 or 2012. For each non-voter ID state, we identified a set of lower and upper controls. We then used each non-voter ID state as a treated unit and estimated the DID treatment effect estimate using these control groups. We then repeated this process for every non-voter ID state. The results is a distribution of state specific estimates that should be zero by construction, since they are based on a comparison of one control state to other control states. We can then compare the estimates from the DID bracketing analysis to this distribution to understand whether the estimated treatment effects are large relative to the placebo distribution. Next, we develop a set of falsification tests that can be used to probe the key identification assumptions.

2.4 Falsification Testing

As is the case with any method for estimating treatment effects in an observational study, one component of the analysis should be probing to ensure that the data are consistent with key assumptions. One particular strength of the DID bracketing approach is the ability conduct a series of falsification tests to probe the identification assumptions. Falsification tests are 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 Rosenbaum (2002); Lipsitch et al. (2010). Angrist Angrist and Krueger (1999) refer to such tests as instances of "refutability." Here, we develop a set of falsification tests for DID bracketing.

This falsification test requires a counterfactual treated group – a treated group that was not in fact treated. That is, we need to observe a counterfactual version of Indiana or Georgia where the state did not in fact adopt voter id laws. Under our assumptions, we should expect both of the following patterns for counterfactual Indiana to hold in the time period after voter ID went into effect: (i) the difference between the upper controls and Indiana and the difference between the lower controls and Indiana in 2008 would be at least as large as the differences in 2004 or (ii) the difference between the upper controls and Indiana and the difference between Indiana and the lower controls in 2008 would be no larger and possibly smaller than the differences in 2004. As such, the following patterns would be inconsistent with our identification assumptions: (iii) the difference between the upper control and counterfactual Indiana is larger in 2008 than in 2004 and the difference between the counterfactual Indiana and the lower control group is smaller in 2008 than in 2004 and (iv) the difference between the upper control and counterfactual Indiana is smaller in 2008 than in 2004 and the difference between the counterfactual Indiana and the lower control group is larger in 2008 than in 2004.

Of course, we never observe counterfactual Indiana or Georgia. However, we can use the time periods before the intervention goes into effect as counterfactual versions of each state. For example, Indiana first used its voter ID law in 2008. We use data from 2004 and 2008 with the DID bracketing approach to bound the true causal effect. For the falsification test, we would use data from 2000 and 2004–two time periods in which voter ID was not in effect. Indiana in 2000 and 2004 serves as a counterfactual, untreated Indiana. The following patterns would be inconsistent with the DID bracketing assumptions: (iii) the difference between the upper control and counterfactual Indiana is in larger in 2004 than in 2000 and the difference between counterfactual Indiana and the lower control group is smaller in 2004 than in 2000 and (iv) the difference between the upper control and counterfactual Indiana is smaller in 2004 than in 2000

and the difference between counterfactual Indiana and the lower control group is larger in 2004 than in 2000.

Next, we outline how we can conduct a unified set of tests across the upper and lower controls and the counterfactual treated state. First, we introduce notation to allow us to formally describe the falsification tests for DID bracketing. The falsification test analysis is conducted with data from a set of time periods before $t = 1$. We denote these time periods with $-t \in \{-2, -1\}$, where -1 indicates a time period before $t = 1$ and -2 indicates a time period prior to $t = -1$. We now define two new DID estimates:

$$\tau_{uc,-t} = \Big(\mathbb{E}[Y^{(-1)}|Z_i = 0] - \mathbb{E}[Y^{(-2)}|Z_i = 0]\Big) - \Big(\mathbb{E}[Y^{(-1)}|Z_i = 1] - \mathbb{E}[Y^{(-2)}|Z_i = 1]\Big)$$

$$\tau_{lc,-t} = \Big(\mathbb{E}[Y^{(-1)}|D_i = 1] - \mathbb{E}[Y^{(-2)}|D_i = 1]\Big) - \Big(\mathbb{E}[Y^{(-1)}|D_i = 0] - \mathbb{E}[Y^{(-2)}|D_i = 0]\Big)$$

These estimates are simply DID estimates based on data from a set of time periods before the treatment goes into effect. We use these DID estimates to test whether the data are consistent with pattern (iii) and pattern (iv). If these patterns are present in the data, this would be inconsistent with the identification assumptions. Pattern (iii) is the intersection of two configurations of the data. The first configuration is

$$\Big(\mathbb{E}[Y^{(-1)}|Z_i = 1] - \mathbb{E}[Y^{(-2)}|Z_i = 1]\Big) < \Big(\mathbb{E}[Y^{(-1)}|Z_i = 0] - \mathbb{E}[Y^{(-2)}|Z_i = 0]\Big)$$

If the first configuration holds, the following will be true $\tau_{uc,-t} < 0$. We write the following hypothesis for this configuration: $H_{(iii)[a]} : \tau_{uc,-t} < 0$. The second configuration is

$$\Big(\mathbb{E}[Y^{(-1)}|D_i = 1] - \mathbb{E}[Y^{(-2)}|D_i = 1]\Big) < \Big(\mathbb{E}[Y^{(-1)}|D_i = 0] - \mathbb{E}[Y^{(-2)}|D_i = 0]\Big).$$

Under this pattern we should find: $\tau_{lc,-t} < 0$. We write the following hypothesis for this pattern: $H_{(iii)[b]} : \tau_{lc,-t} < 0$. We seek to reject the following compound hypothesis:

$$H_{(iii)} : H_{(iii)[a]} \cap H_{(iii)[b]}$$

Using Simes test, we can reject this hypothesis using the following p-value: $p_{(iii)} = min\{2p_{(a)}, p_{(b)}\}$, where $p_{(a)}$ is the one-sided p-value from testing $H_{(iii)[a]}$ and $p_{(b)}$ is the one-sided p-value from

testing $H_{(iii)[b]}$.

Next, we test whether the data are consistent with pattern (iv). Pattern (iv) is the intersection of two different configurations of the data. The first configuration is

$$\Big(\mathbb{E}[Y^{(-1)}|Z_i = 1] - \mathbb{E}[Y^{(-2)}|Z_i = 1]\Big) > \Big(\mathbb{E}[Y^{(-1)}|Z_i = 0] - \mathbb{E}[Y^{(-2)}|Z_i = 0]\Big)$$

We can test the following hypothesis for this pattern: $H_{(iv)[a]} : \tau_{uc,-t} > 0$. The second pattern is

$$\Big(\mathbb{E}[Y^{(-1)}|D_i = 1] - \mathbb{E}[Y^{(-2)}|D_i = 1]\Big) > \Big(\mathbb{E}[Y^{(-1)}|D_i = 0] - \mathbb{E}[Y^{(-2)}|D_i = 0]\Big)$$

We can test the following hypothesis for this pattern: $H_{(iv)[a]} : \tau_{lc,-t} > 0$. Jointly, we seek to reject the following hypothesis:

$$H_{(iv)} : H_{(iv)[a]} \cap H_{(iv)[b]}$$

While these tests are of some interest individually, we are primarily interested in whether either pattern (iii) and (iv) are plausible. We can test this proposition using the following compound hypothesis:

$$H_0 : H_{(iii)} \cup H_{(iv)}$$

which can be tested using the intersection union test which rejects if $max\{p_{(iii)}, p_{(iv)}\} \leq \alpha$, where $p_{(iii)}$ and $p_{(iv)}$ are p-values from testing $H_{(iii)}$ and $H_{(iv)}$ respectively (Lehmann 1952; Berger 1982). This testing plan allows the investigator to test all the key patterns of the falsification test jointly. That is, if the p-value from the global test is less than α, this indicates that at least one of the four patterns in the data are inconsistent with the identification conditions. As is true for falsification tests generally, we prefer larger p-values rather than those above close α. That is, we interpret larger p-values as better evidence from the falsification test.

Other falsification tests are also possible. For example, one could identify a negative control or placebo outcome (Lipsitch et al. 2010; Rosenbaum 2002). Demographic variable such as education or residency status are obvious examples of placebo outcomes in the voter ID context. Here, one would simply apply the DID bracketing method to the placebo outcome in the study period.

2.5 Plausible Violations: A Sensitivity Analysis

It might be the case that we cannot reject $H_{(iii)}$ and $H_{(iv)}$, but we can find evidence to reject patterns (iii) and (iv) of a certain magnitude. That is, we might find that only small violations of the relative trend assumption are plausible. Consider, for instance, the tests for whether the

data are consistent with pattern (iv). First, we generalize the hypotheses associated with these tests. Let $H'_{(iv)[a]} : \tau_{uc,-t} \geq \delta_{uc}, H'_{(iv)[b]} : \tau_{lc,-t} \geq \delta_{lc}$, and $H'_{(iv)} : H'_{(iv)[a]} \cup H'_{(iv)[b]}$, where δ_{uc} and δ_{lc} represent hypothesized values. For a α level test, we can collect the set of values $(\delta_{uc}, \delta_{lc})$ for which we cannot reject $H'_{(iv)}$. We can think of this as the set of plausible violations that are consistent with pattern (iv). An equivalent set of hypotheses and hypothesized values can be formulated for pattern (iii).

These set of plausible patterns can be visualized as regions in a plot with $\tau_{uc,-t}$ on the y-axis and $\tau_{lc,-t}$ on the x-axis. In a plot of this type, the first quadrant corresponds to pattern (iv), the third quadrant corresponds to pattern (iii), and the second and fourth quadrants correspond to patterns where the relative trend assumptions hold. To the extent that the plausible patterns determined from the falsification study are relevant to potential violations of the relative trends assumption, we can use this plot to assess the sensitivity of our results to plausible violations of these assumptions. If the relative trends fall in the second or fourth quadrant, then the assumptions of our model are satisfied and the 95% confidence interval based on DID bracketing are valid. If the relative trends fall in the first or third quadrant then the bracketing interval is no longer valid.

However, even if the results indicate that the data are not consistent with the assumptions, we can use these results to adjust the estimates for plausible violations. If we knew the true relative trends $(\delta_{lc}, \delta uc)$, we could shift the upper and lower control 97.5% interval accordingly to obtain two valid, unbiased intervals for τ, each constructed as the intersection of two one-sided 98.75% confidence intervals. Consequently, the intersection of these 97.5% intervals for τ, rather than the union of the 95% intervals for τ_{lc} and τ_{uc} will be a valid 95% interval for τ. We do not know the true values of $(\delta_{lc}, \delta uc)$, but we can minimize and maximize the lower and upper limits of this 95% CI over the plausible violation region. This will return a valid $(0.95 - \gamma)x100\%)$ confidence interval (Berger and Boos 1994). Denote the two plausible regions $(iii)'$ and $(iv)'$. We can construct this interval as follows

$$\min_{(x,y)\in(iii)'\cup(iv)'}, \max\{\hat{\tau}_{uc}^{0.0125} - y, \hat{\tau}_{lc}^{0.0125} - x\}, \max_{(x,y)\in(iii)'\cup(iv)'}, \min\{\hat{\tau}_{uc}^{0.9875} - y, \hat{\tau}_{lc}^{0.9875} - x\}$$

This interval is an alternative way to estimate the bounds on the true causal effect for a given violation of the relative trends assumption. The resulting procedure provides a valid confidence interval for τ that relies on a different set of assumptions than the bracketing-based interval. Instead of assuming certain patterns of relative trends in the present, we instead assume that the relative trends of the past (the pre-study period) continue in the present (the study period). If it

agrees with the bracketing-based interval that excludes zero, we have strengthened the evidence for the ostensible causal effect by providing two similar confidence intervals valid under different sets of assumptions.

3 Empirical Analysis

Next, we implement the methods outline to estimate the effects of strict voter ID laws in Indiana and Georgia. All our analyses rely on data from the Current Population Survey's (1996; 2002; 2004; 2006; 2008) voting supplement conducted in November of each election year. The Current Population Surveys (CPS) have long been the standard data for trying to assess the costs of voting imposed by states. In all the analyses, we controlled for for whether a respondent was African American, Hispanic, and female. We also controlled for employment status, income, education, and residency status. We conduct separate analyses for Georgia and Indiana. For each state, we use 2004 as the pretreatment time period. We use 2008 and 2012 as separate post-treatment time periods to allow for some possible delay in the effect of voter ID laws which were first in effect in 2008.

3.1 Upper and Lower Control Group Selection

First, we discuss how we selected states for either the lower or upper control groups for the two treated states. First, Hasegawa et al. (2019) outline that control group selection should use data from time periods that are prior to the pre-treatment time period used to construct the DID bracketing estimates. If we use data from the pretreatment period to select the controls groups, it can introduce bias due to regression to the mean. In our analysis, we use 2004 as the pre-treatment time period. To select the two control groups, we pooled CPS data from the 1996 and 2000 presidential elections.

For control group selection, we pooled the 1996 and 2000 data sets, we then estimated two multivariate regression models for turnout that included all the control variables and state fixed effects. In the first model, we set Indiana as the reference category for the state fixed effects. In the second model, we set Georgia as the reference category for the state fixed effects. We used the state fixed effects to identify states with either lower or higher turnout for the 1996 and 2000 elections. For Indiana, the lower control group we selected is made up of the following states: WV, HI, and PA. These states had turnout levels that were at least 2% lower and were statistically significantly lower. For Indiana, the upper control is AK, ME, OR, MN, LA, and ND. These states had turnout levels that were at least 8% higher and statistically significant. For Georgia, the lower control group is: WV, HI, and AZ. These states had turnout levels that

were at least 1% lower and were statistically significant. For Georgia, the upper control is: AK, ME, OR, WI, LA, and MN. These states had turnout levels that were at least 12% higher and statistically significant. In general, we found that both Georgia and Indiana were among the state with the lowest turnout levels in the country, hence, we had to use a smaller set of states for the lower controls.

3.2 DID Bracketing Estimation

For each treated state, we simply estimated two separate linear regressions. We specified each of these models as a DID specification that includes the full set of control variables. In one DID specification, we use the lower controls as the control group. In the second DID specification, we use the higher controls as the control group. In these models, we used cluster-robust variance estimates clustered at the state level, since treatment assignment occurs at the state level. Based on these estimates, we form both the DID bracketing bounds and the 95% confidence intervals for those bounds.

4 Results

4.1 Bracketing Estimates

Next, we present the results for the DID bracketing analysis for both Indiana and Georgia. Table 1 contains the results for Indiana. If we use turnout in 2008 as the outcome, we find that turnout in Indiana increased by 4.4% compared to the lower control group and by 5.7% compared to the upper control group. Subject to the identification assumptions, these estimates imply that the true causal effect is between 3.3% and 7.1%. The results are nearly identical if we use 2012 as the outcome year. In 2012, turnout was higher by 5.9% compared to the lower controls and 4.3% compared to the upper controls. This implies that the bounds on the true causal effect are 3.1% and 8.6%.

Next, we review the results for Georgia. Table 2 contains the estimates for Georgia. When 2008 is designated the outcome year, voter ID laws in Georgia increased turnout by 9.8% compared to the lower controls and by 7.6% compared to the upper controls. The bounds on the true causal effect are 6.3% and 8.9%. When 2012 is the outcome year, turnout increased by 13.4% in Georgia relative to the lower controls and by 8.5% compared to the upper controls, and the bounds on the causal effect are 6.7% and 14.8%. In sum, we find a consistent pattern where states that adopted voter ID laws had higher turnout relative to non-voter ID states. As such, our estimates are consistent with other work that has found that voter ID laws cause

Table 1: DID Bracketing Estimates for Indiana using 2008 and 2012 as Separate Outcome Periods

	2008	
	DID Point Estimate	95% CI
Lower Control Group	4.4	[3.3, 5.4]
Upper Control Group	5.7	[4.2, 7.1]
Bounds	[3.3, 7.1]	
	2012	
	DID Point Estimate	95% CI
Lower Control Group	5.9	[3.1, 8.6]
Upper Control Group	4.3	[3.3, 5.4]
Bounds	[3.1, 8.6]	

Note: Estimates adjusted for education, income, residence type, sex, race, hispanic, and employment status. The estimated bounds should bracket the true causal effect.

higher turnout (Hopkins et al. 2017; Grimmer et al. 2018).

4.2 Placebo Tests

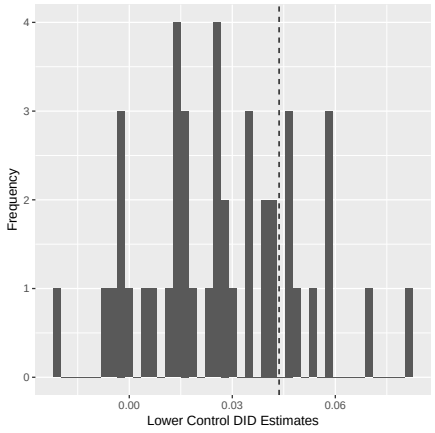
Next, we review the results from applying the placebo test method to the data. We estimated placebo distributions using both 2008 and 2012 as the outcome years. Figure 1 contains the results for Indiana. For Indiana, the DID bracketing study may not be robust to alternative sources of variation. If we use 2008 as the outcome year, using lower controls, we find that 10 states had larger estimates than Indiana. Using upper control states, we find that only one state had a larger estimate (NC). When we use 2012 as the outcome year, 17 states had larger estimates than Indiana when we use the lower controls, and 3 states had larger estimates when we use the upper controls. Thus, when we fully account for statistical variability in the data, the estimates from Indiana are consistent with a placebo distribution of estimates. That is, the estimated increase in turnout in Indiana may be no different from the controls once we fully account for statistical uncertainty.

Figure 2 contains the results for Georgia. For Georgia, the DID bracketing results appear to be robust to alternative sources of variation. For Georgia, we find that its estimate is larger than the upper and lower control state estimates in both 2008 and 2012.

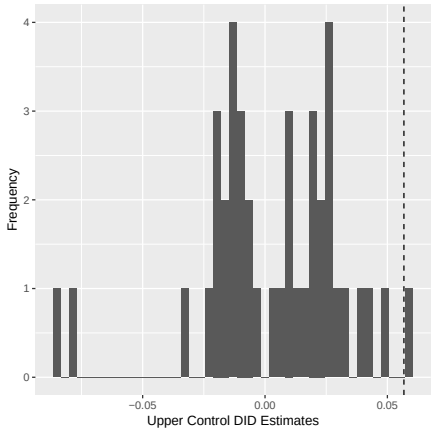
Table 2: DID Bracketing Estimates for Georgia using 2008 and 2012 as Separate Outcome Periods

	2008	
	DID Point Estimate	95% CI
Lower Control Group	9.8	[6.5, 13.1]
Upper Control Group	7.6	[6.3, 8.9]
Bounds	[6.3, 13.1]	
	2012	
	DID Point Estimate	95% CI
Lower Control Group	13.4	[12.0, 14.8]
Upper Control Group	8.5	[6.7, 10.4]
Bounds	[6.7, 14.8]	

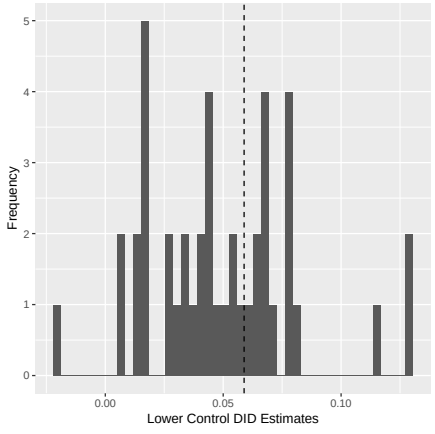
Note: Estimates adjusted for education, income, residence type, sex, race, hispanic, and employment status. The estimated bounds should bracket the true causal effect.



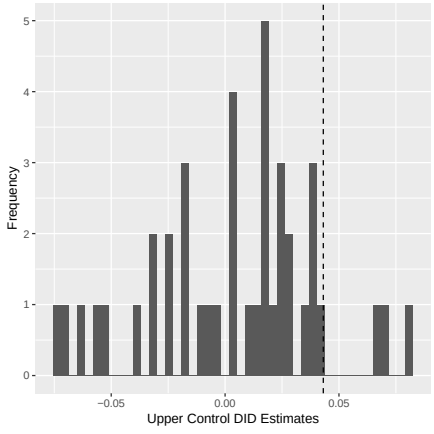
(a) Lower Controls, 2008 As Outcome Year



(b) Upper Controls, 2008 As Outcome Year

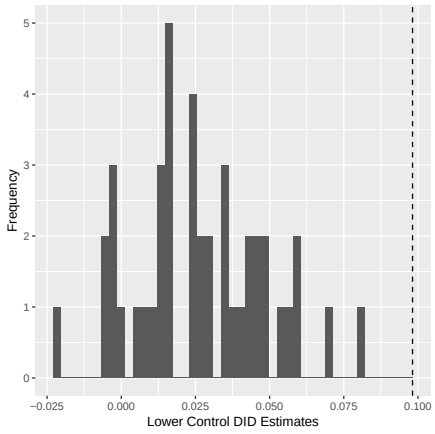


(c) Lower Controls, 2012 As Outcome Year

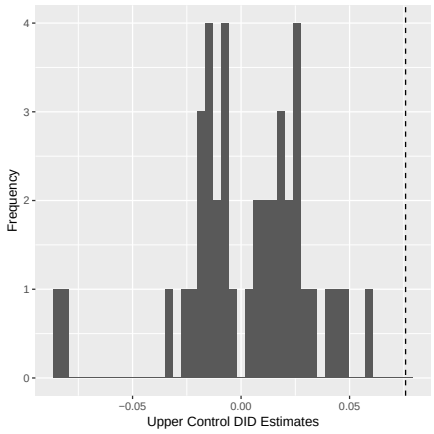


(d) Upper Controls, 2012 As Outcome Year

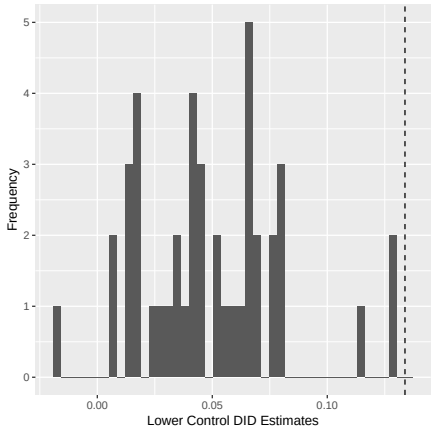
Figure 1: Histogram of DID estimates for Indiana



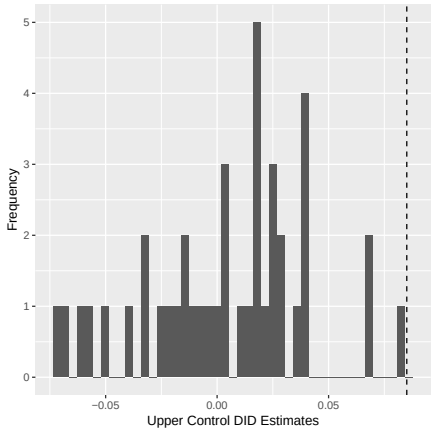
(a) Lower Controls, 2008 As Outcome Year



(b) Upper Controls, 2008 As Outcome Year



(c) Lower Controls, 2012 As Outcome Year



(d) Upper Controls, 2012 As Outcome Year

Figure 2: Histogram of DID estimates for Georgia

4.3 Falsification Tests

Next, we sought to check whether the DID bracketing assumptions are plausible using the proposed falsification tests. We implemented the falsification tests for each treated state by estimating DID estimates using data from 2000 and 2004, the two presidential election years before voter ID laws were in place. Using the data from these years, we again estimated separate treatment effects using the upper and lower control groups. For both Indiana and Georgia, the falsification test results are consistent with the DID bracketing assumptions if we are unable to reject $H_0 : H_{(iii)} \cup H_{(iv)}$. Next, we report the results from the falsification test for each treated state.

First, we review the results for Indiana. For Indiana, we find that $p_{(iii)} = .36$ and $p_{(iv)} = .64$. As such, $max\{p_{(iii)}, p_{(iv)}\} \leq \alpha$ and we are unable to reject H_0 which implies that the data in the pre-treatment time period are consistent with the DID bracketing assumptions. Next, for Georgia, we find $p_{(iii)} = .04$ and $p_{(iv)} = .96$. Again, we find that in the placebo time period, the data are consistent with the identification conditions.

5 Discussion

DID is a commonly used identification strategy in political science and the social sciences more broadly. DID is often used to test how changes in political institutions affect mass behavior. The effect of voter identification laws on turnout is one prototypical application where DID has been frequently used in the empirical literature. However, DID is vulnerable to bias from selective maturation where an event other than the treatment occurs concurrently with the treatment but affects treated and control groups differently. Bias from selective maturation is a key threat in many studies where changes in the electoral environment may have differential effects on the treated and control groups.

Here, we reviewed an alternative identification strategy that invokes an alternative set of assumptions that are less vulnerable to bias from selection maturation. This identification strategy depends on the presence of two separate sets of controls: one set of controls that had systematically higher outcomes and a second control group with lower outcomes in the pre-treatment periods. Using these two set of control group, one can estimate bounds on the true causal effect using assumptions that tend be more plausible than the DID assumptions. One advantage of the method is that it can be easily implemented using standard DID estimation methods. Moreover, inference for the bounds is straightforward. One disadvantage is that, depending on the application, one may not be able to identify two separate control groups with

systematically higher and lower outcomes.

In the paper, we demonstrated how, when several periods of data are available from the pre-treatment time period, the primary assumptions can be probed with a falsification test. We developed a unified test across both sets of control groups that investigators can use to decide whether the data are consistent with the assumptions. We also developed an alternative nuisance parameter approach for estimating bounds on the true causal effect. The DID bracketing approach adds another tool to the kit of the applied investigator. It will not be appropriate in every circumstance, but offers an important alternative to DID methods in many contexts.

References

Alvarez, R. M., Bailey, D., and Katz, J. N. (2008), "The Effect of Voter Identification Laws on Turnout," Social Science Working Paper 1267R.

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

Barnow, B., Cain, G., and Goldberger, A. (1980), "Issues in the Analysis of Selectivity Bias," in *Evaluation Studies*, eds. Stromsdorfer, E. and Farkas, G., San Francisco, CA: Sage, vol. 5, pp. 43–59.

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*.

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, 111–116.

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

Berger, R. L. and Boos, D. D. (1994), "P values maximized over a confidence set for the nuisance parameter," *Journal of the American Statistical Association*, 89, 1012–1016.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004), "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics*, 119, 249–275.

Burden, B. C. (2018), "Disagreement over ID Requirements and Minority Voter Turnout," *The Journal of Politics*, 80, 1060–1063.

Campbell, D. T. (2009), "Prospective: Artifact and control," *Artifacts in Behavioral Research: Robert Rosenthal and Ralph L. Rosnow's Classic Books*, 264.

Campbell, D. T. and Stanley, J. C. (1963), *Experimental and Quasi-Experimental Designs for Research*, Chicago, IL: Rand McNally.

Conley, T. G. and Taber, C. R. (2011), "Inference with difference in differences with a small number of policy changes," *The Review of Economics and Statistics*, 93, 113–125.

Davidson, C. (2009), "The Historical Content of Voter Photo-ID Laws," *PS: Political Science & Politics*, 42, 93–96.

Donald, S. G. and Lang, K. (2007), "Inference With Differences-In-Differences and Other Panel Data," *The Review of Economics and Statistics*, 89, 221–233.

Erikson, R. S. and Minnite, L. C. (2009), "Modeling Problems in the Voter Identification-Voter Turnout Debate," *Election Law Journal*, 8, 85–101.

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, 1045–1051.

Hadar, J. and Russell, W. R. (1969), "Rules for ordering uncertain prospects," *The American economic review*, 59, 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, 1052–1059.

Hajnal, Z., Lajevardi, N., and Nielson, L. (2017), "Voter identification laws and the suppression of minority votes," *The Journal of Politics*, 79, 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*, in press.

Holland, P. W. (1986), "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81, 945–960.

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, 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, 79–128.

Lehmann, E. (1952), "Testing multiparameter hypotheses," *The Annals of Mathematical Statistics*, 541–552.

Lipsitch, M., Tchetgen, E. T., and Cohen, T. (2010), "Negative controls: a tool for detecting confounding and bias in observational studies," *Epidemiology (Cambridge, Mass.)*, 21, 383–388.

Milyo, J. (2007), "The Effects of Photographic Identification on Voter Turnout in Indiana: A County-Level Analysis," Institute of Public Policy Working Paper.

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, 121–126.

Rosenbaum, P. R. (2002), *Observational Studies*, New York, NY: Springer, 2nd ed.

Sobel, R. (2009), "Voter-ID Laws Discourage Participation, Particularly among Minorities, and Trigger a Constitutional Remedy in Lost Representation," *PS: Political Science & Politics*, 42, 107–110.

U.S. Department of Commerce, B. o. t. C. (1996), *CURRENT POPULATION SURVEY: VOTER SUPPLEMENT FILE [Computer File]*, Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

— (2002), *CURRENT POPULATION SURVEY: VOTER SUPPLEMENT FILE [Computer File]*, Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

— (2004), *CURRENT POPULATION SURVEY: VOTER SUPPLEMENT FILE [Computer File]*, Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

— (2006), *CURRENT POPULATION SURVEY: VOTER SUPPLEMENT FILE [Computer File]*, Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

— (2008), *CURRENT POPULATION SURVEY: VOTER SUPPLEMENT FILE [Computer File]*, Ann Arbor, MI: Inter-university Consortium for Political and Social Research.