Bracketing in the Comparative Interrupted Time-Series Design to Address Concerns about History Interacting with Group: Evaluating Missouri's Handgun Purchaser Law

Raiden B. Hasegawa

University of Pennsylvania

Daniel W. Webster

Johns Hopkins University

Dylan S. Small

University of Pennsylvania*

November 7, 2018

Type of Manuscript: Original Research Article

Running Head: Bracketing in Comparative Interrupted Time-Series.

Conflict of interest statement: There is no conflict of interest.

Source of funding: None.

Data and code availability: The data is provided in Table 1. The code that produced the results in §3 can be found in the electronic supplementary materials.

*Address for Correspondence: Dylan Small, Department of Statistics, The Wharton School of the University of Pennsylvania, Philadelphia, PA 19104 (E-mail: *dsmall@wharton.upenn.edu*).

Abstract

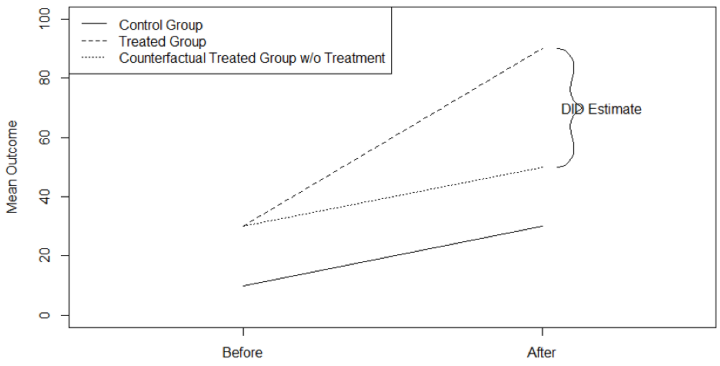
In the comparative interrupted time-series design (also called the method of difference-in-differences), the change in outcome in a group exposed to treatment in the periods before and after the exposure is compared to the change in outcome in a control group not exposed to treatment in either period. The standard difference-in-difference estimator for a comparative interrupted time-series design will be biased for estimating the causal effect of the treatment if there is an interaction between history in the after period and the groups, e.g., there is a historical event besides the start of the treatment in the after period that benefits the treated group more than the control group. We present a bracketing method for bounding the effect of an interaction between history and the groups that arises from a time-invariant unmeasured confounder having a different effect in the after period than the before period. The method is applied to a study of the effect of the repeal of Missouri's permit-to-purchase (PTP) handgun law on its firearm homicide rate. We estimate that the effect of the PTP repeal on Missouri's firearm homicide rate is bracketed between 0.9 and 1.3 homicides per 100,000 people, corresponding to a percentage increase of 17% to 27% (95% CI: [0.6,1.7] or [11%,35%]). A placebo study provides further evidence that the repeal has a causal effect of increasing state-wide firearm homicides.

1 Comparative Interrupted Time-Series Design and Potential Biases

The interrupted time series (ITS) is an observational study design for estimating the causal effect of a treatment on a group when data is available before the group was treated. In the simplest ITS, the before and after treatment outcomes are compared. This before-after design does not account for confounding factors that co-occur with treatment such as historical events or maturation[1]. To strengthen the before-after design, it is common to add time-series data from a control group that never received the treatment over the same period – the comparative ITS

(CITS) design[1,2,3,4], also called the nonequivalent control group design or method of difference-in-differences. The latter name derives from that the simplest CITS analysis is to take the difference between the difference of the after and before outcomes for the treated group and the difference of the after and before outcomes for the control group. This difference-in-differences estimate is an unbiased estimator of the causal effect of treatment if the treatment and control groups would have exhibited parallel trends in the counterfactual absence of treatment[2]; see Figure 1. The parallel trends assumption can be partially assessed if there is more than one

Figure 1: Stylized plot of data from a comparative interrupted time series design. The dotted line shows the assumption that the difference-in-difference (DID) estimate makes about the treatment groups counterfactual mean in the absence of treatment.



time point in the before period by assessing whether the groups exhibit parallel trends in the before period[2]. However, even if the trends are parallel in the before period, there could be historical events in the after period that affect the two groups differently, i.e., history interacts with group (other reasons that parallel trends could be violated include differences in maturation, instrumentation or statistical regression between the groups)[5,6]. For example, the outcome measures poor health, country A (treated group) enacts a policy reform, country B (control group) does not enact the reform, and a worldwide economic recession occurs after the reform that has a greater impact on people starting out in poorer health. If country B started out with poorer health, then parallel trends would be violated because country B's poor health would

have increased more than country A in the after period in the counterfactual absence of the reform because of the worldwide economic recession. This violation of parallel trends would not happen if A and B started with the same level of poor health in the before period. However, it is often difficult to find a control group that has outcomes close to the treated group in the before period.

When there is no control group completely comparable to the treated group, Campbell[7] proposed bracketing to distinguish treatment effects from plausible biases[8]. For comparing treatment and control at one time point, suppose there is concern about an unmeasured confounder U. Bracketing uses two control groups such that, in the first group U tends to be higher than in the treated group and in the second group, U tends to be lower. The effect of U on the treated group is bracketed by its effect on the two control groups. When there is bracketing, if the treated group has a significantly higher outcome than both control groups, then this association between treatment and outcome cannot plausibly be explained away as being bias from U.

In this paper, we show how bracketing can be applied to the CITS to distinguish treatment effects from plausible biases due to history interacting with group. The basic idea is to consider one control group that has a lower expected outcome than the treated group in the before period and another control group that has a higher expected outcome than the treated group in the before period; we show under certain assumptions that the expectations of the two difference-in-difference estimators using the lower control group and higher control group respectively bracket the causal effect of the treatment. Bracketing for the CITS has been mentioned informally[2] but the idea of choosing the bracketing control groups based on expected before period outcomes was not mentioned. We present assumptions and results for our bracketing method (§2) and then apply the method to study the effect of the repeal of Missouri's permit-to-purchase handgun law on its firearm homicide rate (§3).

2 Bracketing

2.1 Notation and Model

Let Y denote outcome and D dose of exposure, $D = 1$ for treatment and $D = 0$ for control. Let $Y_{ip}^{(d)}$ denote the counterfactual outcome that would have been observed for unit i in period p, $p = 0$ for before period and $p = 1$ for after period, had the unit received exposure dose d, i.e., $Y_{ip}^{(1)}$ is the counterfactual outcome under treatment and $Y_{ip}^{(0)}$ is the counterfactual outcome under control. Let $\mathbf{U}_i$ be a vector of time invariant unmeasured confounders for unit i. Let G denote group where the groups are t = treated group, lc = lower control group (control group with expected outcomes lower than treated group in before period) and uc = upper control group (control group with expected outcomes higher than treated group in before period). Finally, let S be an indicator of whether or not a unit belonging to a particular group is in the study population in a given period. Specifically, $S_{ip} = 1$ or 0 when unit i is in the population or not in period p: $S_{i0} = S_{i1} = 1$ for a unit in the population both before and after treatment, $S_{i0} = 1, S_{i1} = 0$ for a unit in the population only before treatment (unit might have moved away or died in after period) and $S_{i0} = 0, S_{i1} = 1$ for a unit in the population only after treatment (unit might have moved into study area or been born in after period).

We consider the following model which generalizes the standard difference-in-difference model and changes-in-changes model.[9] Let $\mathbf{U}_i$ be time-invariant unmeasured confounders and ϵ_{ip} be an error term that captures additional sources of variation for unit i in period p. Then our model can be expressed as

$$Y_{ip}^{(d)} = h(\mathbf{U}_i, p) + \beta d + \epsilon_{ip} \tag{1}$$

where the function $h(\mathbf{U}_i, p)$ is the unobserved expected outcome under control of subject i in period p. We drop the subscript i to refer to a randomly drawn unit from the population of all units in either period, where $Y_p^{(d)}, d = 0, 1$, and ϵ_p are undefined if $S_p = 0$. We make the

following assumptions:

Increasingness of h *in* $\mathbf{U}$: $h(\mathbf{U}, p)$ bounded and increasing in $\mathbf{U}$ for $p = 0, 1$. (2)
($(h(\mathbf{U}, p) \geq h(\mathbf{U}', p)$ whenever all coordinates of $\mathbf{U} \geq$ all coordinates of $\mathbf{U}'$)

Time Invariance of $\mathbf{U}$ *Within Groups*: $\mathbf{U}$ conditionally independent of (3)
$\{S_0, S_1\}$ given group G.

Independence of ϵ *with Time and Group*: Distributions of $\epsilon_p | S_p = 1, G = g$ for (4)
$p = 0, 1$, $g = lc, uc, tc$ all have mean zero and are the same.

Assumptions (2) and (3) match assumptions in the changes-in-changes model. Assumption (2) requires that higher levels of unmeasured confounders correspond to higher levels of outcomes. Such increasingness is natural when the unmeasured confounder is an individual characteristic such as health or ability[9] and Y is a measure of some positive outcome, for example, income. Negative confounders – where higher levels of the confounder correspond to lower levels of the outcome – are not precluded by Assumption (2) as the corresponding coordinates of $\mathbf{U}$ may simply be replaced by their negation. Assumption (3) says that the distribution of confounders in the population of units for a given group remains the same over time. Assumption (4) says that time-varying factors have the same distribution in each group and over time; it would be sufficient for subsequent developments to just assume the distributions of $\epsilon_p | S_p = 1, G = g$ for $p = 0, 1$, $g = lc, uc, tc$ all have mean zero rather than the stronger (4). We can further relax this assumption by assuming zero mean only for components of ϵ_p that are true confounders, that is, factors whose distributions depend on the interaction of time and group. Assumption (4) is weaker than the changes-in-changes model assumption that ϵ_{ip} is always zero which rules out classical measurement error in the outcome when h is non-linear.[9] Our model contains the

standard difference-in-difference model, which can be represented in our model by $h(\mathbf{U}, p) = k(\mathbf{U}) + \tau p$ for some bounded and increasing function k, where $k(\mathbf{U})$ can be viewed as a group fixed effect.

We make two further assumptions about the distribution of $\mathbf{U}$ in groups and how its effect over time changes among the groups. First, we assume the distribution of $\mathbf{U}$ within groups can be stochastically ordered so that $\mathbf{U}$ is lowest in the lower control group, intermediate in the treated group and highest in the upper control group:

$$\mathbf{U}|G = lc \preceq \mathbf{U}|G = t \preceq \mathbf{U}|G = uc \tag{5}$$

where two random vectors $\mathbf{A}, \mathbf{B}$ are stochastically ordered, $\mathbf{A} \preceq \mathbf{B}$, if $E[f(\mathbf{A})] \leq E[f(\mathbf{B})]$ for all bounded increasing functions f[10]. For example, if $\mathbf{U}$ is normally distributed with common variance and group means μ_{lc}, μ_t, and μ_{uc}, then $\mu_{lc} \leq \mu_t \leq \mu_{uc}$ would imply (5) . Second, we assume that higher values of $\mathbf{U}$ either have a bigger effect over time over the whole range of $\mathbf{U}$ or a smaller effect over the whole range:

$$\begin{aligned} \text{Either (i) } & h(\mathbf{U}, 1) - h(\mathbf{U}, 0) \geq h(\mathbf{U}', 1) - h(\mathbf{U}', 0) \text{ for all } \mathbf{U} \geq \mathbf{U}', \mathbf{U}, \mathbf{U}' \in \mathcal{U} \text{ or} \\ \text{(ii) } & h(\mathbf{U}, 1) - h(\mathbf{U}, 0) \leq h(\mathbf{U}', 1) - h(\mathbf{U}', 0) \text{ for all } \mathbf{U} \geq \mathbf{U}', \mathbf{U}, \mathbf{U}' \in \mathcal{U} \end{aligned} \tag{6}$$

An example of this pattern of $\mathbf{U}$ confounding could occur in a study of the effect of a regional policy on average income where the policy change occurred contemporaneously with an easing of trade restrictions. A potential unmeasured confounder for such a study would be $\mathbf{U}$ = share of skilled workers in a region, as a higher share of skilled workers is associated with higher average income. There is considerable evidence that trade liberalization leads to an increase in the skill premium – the relative wage of skilled to unskilled workers – at both the regional and country level[11,12]. Thus, we might expect (i) in (6) to hold if there was an easing of trade restrictions in the after period.

We assume units are randomly sampled from each group in each time period. The data could be obtained from repeated cross sections or a longitudinal study. The supplementary materials discuss inferences under different sampling assumptions.

2.2 Bracketing Result

The standard difference-in-difference estimator using control condition c can be written as $\hat{\beta}_{dd.c} = (\overline{Y}_{t,1} - \overline{Y}_{t,0}) - (\overline{Y}_{c,1} - \overline{Y}_{c,0})$ where $\overline{Y}_{g,p}$ indicates the sample average of units in group g and time period p. This estimate is equivalent to the coefficient on the treatment indicator in a fixed effects regression with full time and group indicator variables. When using data already aggregated at some level, for example by state-year, a fixed effects regression using weights proportional to population will return this estimate. In the following, we show that the expectation of the two standard difference-in-difference estimators computed with the upper and lower controls can be used to bound the treatment effect.

The expected value of the standard difference-in-difference estimator comparing the treated group to the lower control group, $\hat{\beta}_{dd.lc}$, is

$$\begin{aligned} E[\hat{\beta}_{dd.lc}] &= \{E[Y_1|G=t, S_1=1] - E[Y_0|G=t, S_0=1]\} \\ &\quad -\{E[Y_1|G=lc, S_1=1] - E[Y_0|G=lc, S_0=1]\} \\ &= \{\beta + E[h(\mathbf{U},1)|G=t, S_1=1] - E[h(\mathbf{U},0)|G=t, S_0=1]\} \\ &\quad -\{E[h(\mathbf{U},1)|G=lc, S_1=1] - E[h(\mathbf{U},0)|G=lc, S_0=1]\}, \end{aligned}$$

where Y_1, Y_0 denote observed outcomes in after period ($p = 1$) and before period ($p = 0$) respectively. Under the time invariance of $\mathbf{U}$ within groups assumption (3), we have

$$E[\hat{\beta}_{dd.lc}] = \beta + \{E[h(\mathbf{U},1) - h(\mathbf{U},0)|G=t]\} - \{E[h(\mathbf{U},1) - h(\mathbf{U},0)|G=lc]\}; \qquad (7)$$

similarly, the expected value of the difference-in-difference estimator comparing the treated

group to the upper control group, $\hat{\beta}_{dd.uc}$, is

$$E[\hat{\beta}_{dd.uc}] = \beta + \{E[h(\mathbf{U}, 1) - h(\mathbf{U}, 0)|G = t]\} - \{E[h(\mathbf{U}, 1) - h(\mathbf{U}, 0)|G = uc]\}. \quad (8)$$

The difference-in-difference estimators $\hat{\beta}_{dd.lc}$ and $\hat{\beta}_{dd.uc}$ are unbiased if $h(\mathbf{U}, 1) - h(\mathbf{U}, 0)$ is constant for all $\mathbf{U}$, i.e., if the effect of the unmeasured confounders is the same in different periods and there is just a time effect between periods that is the same for all levels of $\mathbf{U}$. If the effect of the unmeasured confounders changes between periods, then because of assumptions (5) and (6), we conclude from (7) and (8) that

$$\min\{E[\hat{\beta}_{dd.lc}], E[\hat{\beta}_{dd.uc}]\} \leq \beta \leq \max\{E[\hat{\beta}_{dd.lc}], E[\hat{\beta}_{dd.uc}]\}, \quad (9)$$

i.e., the expected values of the difference-in-difference estimators using the upper control group and lower control group bracket the causal effect (proof in supplementary materials). The tightness of the bracketing bounds in (9) and, to some extent, the width of the corresponding confidence interval developed in following section depend on the magnitude of the group-by-time interaction. For example, if urban poverty concentration varied notably between groups and its effect on firearm homicides were modulated by the Great Recession, one would expect looser bracketing bounds.

2.3 Inference

We would like to make inferences for the causal effect β under the assumption (6) that $h(\mathbf{U}, 1) - h(\mathbf{U}, 0)$ is either an increasing or decreasing function of $\mathbf{U}$ (we do not want to specify which a priori). Let $\theta_{lc.t} = E[\hat{\beta}_{dd.lc}]$ and $\theta_{uc.t} = E[\hat{\beta}_{dd.uc}]$, i.e., the expected values of the difference-in-difference estimators using the lower control group and upper control group, respectively. From

the bracketing results (9), we have

$$\min(\theta_{lc.t}, \theta_{uc.t}) \leq \beta \leq \max(\theta_{lc.t}, \theta_{uc.t}).$$

and the following interval, where CI means confidence interval,

[min(lower endpoint of $1-\alpha$ two sided CI for $\theta_{lc.t}$, lower endpoint of $1-\alpha$ two sided CI for $\theta_{uc.t}$), max(upper endpoint of $1-\alpha$ two sided CI for $\theta_{lc.t}$, upper endpoint of $1-\alpha$ two sided CI for $\theta_{uc.t}$)], (10)

has probability $\geq 1-\alpha$ of containing both $\min(\theta_{lc.t}, \theta_{uc.t})$ and $\max(\theta_{lc.t}, \theta_{uc.t})$, and thus β, where it assumed that the two-sided CIs are constructed by taking the intersection of two one-sided $1-(\alpha/2)$ confidence intervals (proof in supplementary materials).

2.4 Constructing the Lower and Upper Control Groups

The results in Sections 2.2-2.3 assume the lower and upper control groups have been constructed before looking at the data. If the lower control group was constructed by looking at the before period data by choosing units with lower outcomes than the treated in the before period, then the sample average of $Y_0|G = lc, S_0 = 1$ may tend to be lower than $E(Y_0|G = lc, S_0 = 1)$. Consequently, the difference-in-difference estimate using the lower control group may be downward biased even if the parallel trends assumption holds because of regression to the mean[1]; similarly, the difference-in-difference estimated using the upper control group may be upward biased. This may invalidate the bracketing result (9). To avoid the regression to the mean problem, we propose first selecting a "pre-study" time period prior to the before period. Then, the lower control group can be constructed from units with lower outcomes than the treated in this pre-study period and the upper control group from units with higher outcomes. It should then be tested whether the constructed lower control group has smaller expected outcomes than the

constructed upper control group in the before period; see §3 for example.

2.5 Role of Examining the Groups' Relative Trends in the Before Period

In the standard difference-in-difference analysis which assumes parallel trends, when the before period contains multiple time points, it is good practice to test for parallel trends in the before period[2,13]. In our bracketing approach, we do not need the parallel trend assumption to hold, but examining the relative trends of the groups in the before period is still useful for assessing model plausibility and assumptions. Our model (1)-(4) along with assumptions (5)-(6) implies that if we had counterfactual data on the treatment group in the after period in the absence of treatment, then, without sampling variance, we would see either: (i) the differences between the upper control and counterfactual treated groups and the difference between the counterfactual treated and lower control groups in the after period would be at least as large as their respective differences in the before period or (ii) the difference between the upper control and counterfactual treated groups and the difference between the counterfactual treated and lower control groups in the after period would be no larger and possibly smaller than their respective differences in the before period. The following two patterns would violate the model/assumptions: (iii) the difference between the upper control and counterfactual treated groups is larger after than before and the difference between the counterfactual treated and lower control groups is smaller after than before or (iv) the difference between the upper control and counterfactual treated groups is smaller after than before and the difference between the counterfactual treated and lower control groups is larger after than before. Although we do not have the counterfactual treatment group's data in the absence of treatment in the after period, we have the treatment group's data in the absence of treatment in the before period. We can split the before period into two (or more) periods and test whether the pattern in the before period is consistent with the model. Visual inspection of the relative trends of the counterfactual treated group and the upper and lower control groups during the before period can provide additional evidence for or

against the model assumptions.

2.6 Time-Varying Confounders

Our bracketing method addresses an interaction between history and groups that arises because the time-invariant unmeasured confounders that differ between the groups in the before period ($\mathbf{U}$) become more (or less) important in the after period (assumption (6)). When there are time-varying confounders, the bracketing method still works under certain assumptions. Time-varying confounders can be represented in model (1) by letting $\mathbf{U}$ contain all variables that differ in distribution between the groups in the before period, ϵ_{i0} be the effect of factors that do not differ in distribution between the groups in the before period and ϵ_{i1} be the effect of the same factors in ϵ_{i0} in the after period as well as factors not contained in $\mathbf{U}$ that differ in distribution between the groups in the after period (details on model in supplementary materials). If this last set of factors is present, then (4) may not hold. However, the bracketing result (9) still holds as long as (i) in (6) holds,

$$E[\epsilon_{i1}|G = uc] \geq E[\epsilon_{i1}|G = t] \geq E[\epsilon_{i1}|G = lc], \tag{11}$$

or when (ii) in (6) holds,

$$E[\epsilon_{i1}|G = uc] \leq E[\epsilon_{i1}|G = t] \leq E[\epsilon_{i1}|G = lc]; \tag{12}$$

supplementary materials contain proof and sufficient conditions for (11) or (12) to hold. A part of the sufficient conditions (condition (c) in supplement) is analogous to (i) in (6) in that effects on the outcome, be they time effects or those due to contemporaneous shocks to confounders, are amplified at larger values of $\mathbf{U}$.

One type of time-varying confounder is a variable that largely stays the same between time periods but may change modestly. For example, in our study of Missouri's repeal of their

permit-to-purchase law in §3, urban concentration of poverty might be a confounder and $\mathbf{U}$ contain urban concentration of poverty in the before period. Urban concentration of poverty may stay mostly the same over time but change modestly, where the changes are reflected in ϵ_1. If the effect of urban concentration of poverty on firearm homicides increased in the after period, then the bracketing result would still hold (with respect to the confounding from urban concentration of poverty) as long as the impact of changes in urban concentration of poverty on firearm homicides were at least as great in the upper control group as Missouri and at least as great in Missouri as the lower control group.

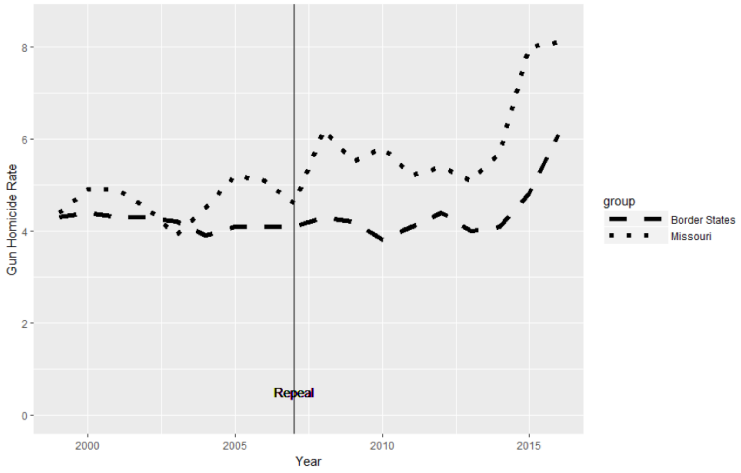
3 Application: Effect of the Repeal of Missouri's Handgun Purchaser Licensing Law on Firearm Homicides

American federal gun law requires background checks and record keeping for gun sales by federally licensed firearm dealers but exempts these regulations for private sales. However, some states have laws requiring all purchasers of handguns from licensed dealers *and* private sellers to acquire a permit-to-purchase (PTP) license that verifies the purchaser has passed a background check. Missouri passed a PTP law in 1921, requiring handgun purchasers to obtain a license from the local sheriff's office which facilitated the background check, but repealed the law on August 28, 2007. Webster et al.[14] examined the effect of Missouri's repeal on firearm homicide rates (the rate of homicides committed using a firearm). One of their analyses used a CITS design, comparing Missouri to the eight states bordering Missouri using a before-period of 1999-2007 and after-period of 2008-2010 (the only available post-repeal data at the time of their analysis), finding evidence that the repeal of Missouri's PTP law increased firearm homicide rates (see their Table 1). None of the border states introduced new or made changes to existing PTP laws during the study period. Using a fixed effect regression and adjusting for several background crime and economic covariates, they estimated that the Missouri PTP repeal was

associated with an increase in the firearm homicide rate by 1.1 per 100,000 persons (95% CI: 0.8,1.4) , a 22% (95% CI: 16 %, 29%) increase. Non-gun related homicides remained virtually unchanged. In what follows, we re-examine the effect of Missouri's repeal using bracketing and the now available after-period data from 2008-2016 to address possible biases arising from unobserved state-by-time interactions.

Figure 3 shows the age-adjusted firearm homicide rates in Missouri and the border states over the study period using data from the CDC's Wide-ranging Online Data for Epidemiologic Research (WONDER) system[15]. The standard difference-in-difference estimate using all neighboring control states, shown in the top row of Table 2, is that Missouri's PTP repeal increased firearm homicides by 1.2 per 100,000 persons (95% CI: 1.0,1.4), corresponding to a 24% increase (95% CI: 18%,31%). In the before-period, Missouri had generally higher firearm

Figure 2: Age-adjusted firearm homicide rates in Missouri and states bordering Missouri (population-weighted averages), 1999-2016.



homicide rates than the control border states, suggesting a lack of comparability between the groups. One concern is that the start of the after period coincided with the beginning of the Great Recession. The economic downturn was followed by a decline in homicide rates. Possible reasons for the effect of the downturn on homicide rates and violence generally include changing alcohol affordability, disposable income, unemployment and income inequality[16,17,18]. The ef-

fects of the economic downturn on firearm homicides might interact with the starting level of firearm homicides in a state. To address this concern, we constructed upper and lower control groups that bracket Missouri's firearm homicide rate in the before period. To avoid regression to the mean (§2.4), we use data from 1994-1998, the five years prior to our before period, to choose the upper and lower control groups; see Table 1 for data. The lower control group is Iowa, Kansas, Kentucky, Nebraska and Oklahoma and the upper control group is Arkansas, Illinois and Tennessee. The population-weighted firearm homicide rate in the before period of 1999-2007 is 5.2, 4.7 and 2.7 in the upper control states, Missouri and the lower control states, respectively (95% CI for difference between upper control and Missouri: 0.2,0.8; 95% CI for difference between Missouri and lower controls: 1.8,2.2).

Table 1: Age-adjusted firearm homicide rates per 100,000 persons from periods 1994-1998 (pre-study period used to construct lower and upper control groups), 1999-2007 (before repeal period where repeal refers to repeal of Missouri's permit-to-purchase handgun licensing law) and 2008-2016 (after repeal period).

	1994-1998	1999-2007	2008-2016
Missouri	6.1	4.7	6.1
Arkansas	7.3	5.1	5.5
Illinois	7.1	5.1	5.2
Iowa	1.2	0.9	1.2
Kansas	4.2	3.0	3.0
Kentucky	4.1	3.3	3.7
Nebraska	2.2	1.8	2.4
Oklahoma	4.8	3.8	4.8
Tennessee	6.9	5.5	5.4
Population-weighted All Controls	5.6	4.2	4.4
Population-weighted Upper Controls	7.1	5.2	5.3
Population-weighted Lower Controls	3.5	2.7	3.2

Figure 3 shows firearm homicides rates (age-adjusted and population-weighted) in the bracketed control groups compared to Missouri. The bottom two rows of Table 2 show the difference-

in-difference estimates using the lower and upper control groups and 95% CIs. Both the lower and upper control groups provide evidence that Missouri's repeal of its permit-to-purchase handgun law increased firearm homicides. The interval (10) that has a $\geq 95\%$ chance of containing the effect of the repeal on the firearm homicide rate is $[0.6, 1.7]$, corresponding to an 11% to 35% increase in firearm homicides, providing evidence that the repeal increased firearm homicides.

Figure 3: Age-adjusted gun homicide rates per 100,000 persons in Missouri, lower control states bordering Missouri (population-weighted averages) and upper control states bordering Missouri, 1999-2016.

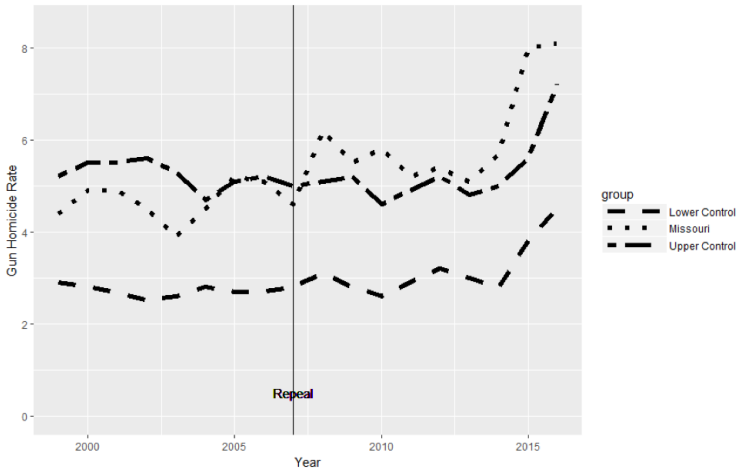


Table 2: Difference-in-difference estimates of effect of repeal of Missouri's permit-to-purchase handgun licensing requirement on firearm homicide rates per 100,000 persons

Control Group	Estimate	95% CI	Corresponding % Change Estimate	95% CI
All Controls	1.2	[0.9, 1.5]	24%	[18% ,31%]
Upper Controls	1.3	[0.9, 1.7]	27%	[19% ,35%]
Lower Controls	0.9	[0.6, 1.2]	17%	[11% ,23%]

3.1 Assessing Model Assumptions: Time-Varying Confounders and Relative Trends

A type of time-varying confounder that is relevant to the Missouri PTP study is a factor that only arises in the after period. The Ferguson unrest in 2014 might have led to less effective policing (spikes in violence typically follow social unrest) in Missouri compared to other states. Such a time-varying confounder would be unlikely to satisfy (11) or (12) because it arises only in the treated group (Missouri) in the after period. However, this confounder alone does not change our finding that the repeal increased firearm homicides. If we limit the study to 2008-2013, Missouri still has significantly higher increases in firearm homicide rates than both the upper and lower control groups (supplementary materials).

To assess the plausibility of our model (1)-(4) and assumptions (5)-(6), we apply the relative trends test described in §2.5. Applying the test to our study of the repeal of Missouri's PTP law, we do not find evidence that our model assumptions are violated. Visual inspection of the relative trends of counterfactual Missouri and the upper and lower controls in the before period further supports the plausibility of our model assumptions; see supplementary materials.

3.2 Standard Error Estimates: A Poisson Model for Death Counts

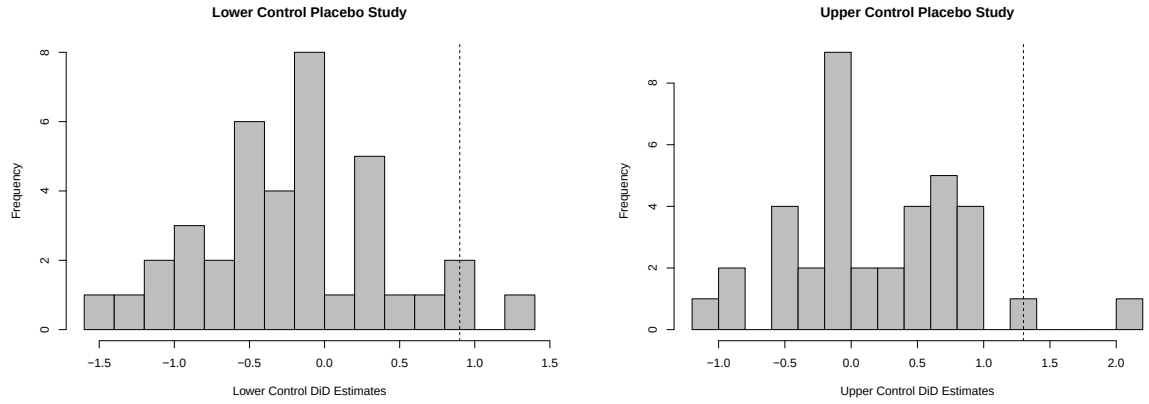
The standard errors used for inference in the previous section come directly from the CDC WONDER system. Vital statistics that derive from complete counts of deaths (by cause) are not subject to sampling error. Nonetheless, a stochastic model of vital statistics may be justified by the presence of biological, environmental, sociological, and other natural sources of variability[19]. For inferential purposes, a census may be viewed as a realization from such a stochastic process under similar conditions to those observed[20]. In particular, the observed firearm homicide death rate in any state-year may be viewed as one of a large series of possible Poisson distributed outcomes under similar conditions[21]. The standard errors reported by the CDC are computed under this Poisson model.

3.3 A Placebo Study: Assessing Alternative Sources of Uncertainty

There may be other sources of uncertainty unaccounted for by the natural variability of a Poisson model for yearly state-level firearm homicides. Several recent papers suggest that such sources of uncertainty, if ignored, may yield substantially different inferential conclusions. Serially correlated data[22], yearly state-level shocks[23], and small numbers of policy changes[24] can cause the standard errors returned by a fixed effects regression to be downwardly biased. We address the inferential challenges that arise from the presence of time-varying state-level shocks by conducting a placebo study[25,22].

Akin to permutation inference, a placebo study in the context of the Missouri PTP repeal analysis applies the bracketing method to every state to create a placebo treatment effect distribution. Specifically, for each state where there was no PTP repeal we construct lower and upper control groups of neighboring states, when available, in exactly the same way we did so for Missouri. We then compute the difference-in-difference estimates using both control groups for a placebo "repeal" on August 28, 2007. This results in two permutation distributions for the placebo effect estimates, one estimated using lower controls and the other using upper controls. If the PTP repeal effect in Missouri is not spurious, we'd expect to see few placebo effects greater than the ones reported in our study using either control condition. The histograms of the placebo effects in Figure 4 suggest that the Missouri bracketing study is relatively robust to these alternative sources of variability. Of the 38 states that had lower control neighbors, only two (Oklahoma and Delaware) had placebo effect estimates using lower controls that were larger than Missouri (dashed line, left panel). Of the 37 states that had upper control neighbors, only one (Delaware) had a placebo effect estimate using upper controls that was larger than Missouri (dashed line, right panel). Alaska, Hawaii, DC and three states with missing data in either the pre-study, before or after period were excluded from the analysis.

Figure 4: (Left Panel): Histogram of placebo difference-in-difference (DiD) estimates using lower control states ($n = 38$ states with lower control neighbors – includes Missouri). Two states had a larger estimate than Missouri (dashed line).(Right Panel): Histogram of placebo DiD estimates using upper control states ($n = 37$ states with upper control neighbors – includes Missouri). One state had a larger estimate than Missouri (dashed line).



4 Conclusion and Discussion

We developed a bracketing method for CITS to account for concerns that history may interact with groups. In a study of the repeal of Missouri's permit-to-purchase handgun law, the method addressed a concern that on average, control states started out with lower firearm homicide rates than Missouri before the repeal. Comparing both to states that started with higher firearm homicide rates than Missouri and states that started with lower rates, the repeal was associated with a significant increase in firearm homicides, thus strengthening the evidence that the repeal had a causal effect of increasing firearm homicides.

A limitation of our estimated impact of the repeal of Missouri's PTP law is that a Stand Your Ground (SYG) law was simultaneously adopted in Missouri. However, in the original study by Webster et al.[14], the inclusion of an SYG indicator in the regression did not dramatically change the estimated effect. Additionally, a recent comparative interrupted time series study examining firearm homicide rates in large urban counties found that PTP laws were associated with significant reductions in firearm homicides after controlling for the effects of SYG laws[26]. Further evidence that the contemporaneous SYG law does not change the qualitative conclusion

of our study can be found in the placebo study. There were 16 additional states that adopted SYG laws within a few years of Missouri's PTP repeal[26]. Only one state (Oklahoma) of the 16 had a difference-in-difference placebo effect estimate using lower controls that was larger than Missouri and none of the states had placebo effect estimates using upper controls that were larger than Missouri.

Although only one of many potential patterns of bias, the history-by-group interaction bias addressed in this paper has been mentioned in the literature since at least the middle of the 20th century – it is referred to *selection-maturation interaction* in a taxonomy of possible threats to the validity of experimental and quasi-experimental designs presented in Campbell and Stanley[27]. Fundamentally, bracketing relies on constructing control groups across which this potential source of confounding is systematically varied[28]. Other methods for constructing adequate control groups in the presence of history-by-group interactions, such as the synthetic control method[25], have also found success in comparative case studies of the effect of PTP laws on firearm homicide rates[29]. While we do not argue that bracketing is uniformly superior to the synthetic control method, the practitioner may find that each has strengths that lend themselves to different settings. When the researcher believes that unmeasured history-by-group confounding, $h(\mathbf{U}, p)$, can be expressed as a linear factor model with time-varying slopes and group-specific loadings, the synthetic control method provides an asymptotically unbiased point estimate of the causal effect of treatment while bracketing can only provide bounds on the treatment effect. However, when the practitioner suspects that only the weaker assumptions of the model outlined in §2.1 hold, the bracketing bounds will remain unbiased, in that they contain the true effect in expectation, while the point estimate using synthetic controls need not be. A detailed example of such a case can be found in the supplement.

References

[1] T. D. Cook, D. T. Campbell, and W. Shadish, *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin Boston, 2002.

[2] B. D. Meyer, "Natural and quasi-experiments in economics," *Journal of Business & Economic Statistics*, vol. 13, no. 2, pp. 151–161, 1995.

[3] J. L. Bernal, S. Cummins, and A. Gasparrini, "Interrupted time series regression for the evaluation of public health interventions: a tutorial," *International journal of epidemiology*, vol. 46, no. 1, pp. 348–355, 2017.

[4] C. Wing, K. Simon, and R. A. Bello-Gomez, "Designing difference in difference studies: Best practices for public health policy research," *Annual Review of Public Health*, vol. 39, pp. 453–469, 2018.

[5] T. D. Cook and D. T. Campbell, *Quasi-experimentation: Design and analysis for field settings*. Rand McNally, 1979.

[6] K. D. Reynolds and S. G. West, "A multiplist strategy for strengthening nonequivalent control group designs," *Evaluation Review*, vol. 11, no. 6, pp. 691–714, 1987.

[7] D. T. Campbell, "Prospective: Artifact and control," in *Artifacts in Behavioral Research: Robert Rosenthal and Ralph L. Rosnow's Classic Books*. Academic, 1969, pp. 351–382.

[8] P. R. Rosenbaum, "The role of a second control group in an observational study," *Statistical Science*, vol. 2, no. 3, pp. 292–306, 1987.

[9] S. Athey and G. W. Imbens, "Identification and inference in nonlinear difference-in-differences models," *Econometrica*, vol. 74, no. 2, pp. 431–497, 2006.

[10] M. Shaked and J. G. Shanthikumar, *Stochastic Orders and Their Applications*. New York: Academic Press, 1994.

[11] R. Dix-Carneiro and B. K. Kovak, "Trade liberalization and regional dynamics," *American Economic Review*, vol. 107, no. 10, pp. 2908–46, 2017.

[12] A. Burstein and J. Vogel, "International trade, technology, and the skill premium," *Journal of Political Economy*, vol. 125, no. 5, pp. 1356–1412, 2017.

[13] K. G. Volpp, A. K. Rosen, P. R. Rosenbaum, P. S. Romano, O. Even-Shoshan, Y. Wang, L. Bellini, T. Behringer, and J. H. Silber, "Mortality among hospitalized medicare beneficiaries in the first 2 years following ACGME resident duty hour reform," *Journal of the American Medical Association*, vol. 298, no. 9, pp. 975–983, 2007.

[14] D. Webster, C. K. Crifasi, and J. S. Vernick, "Effects of the repeal of missouris handgun purchaser licensing law on homicides," *Journal of Urban Health*, vol. 91, no. 2, pp. 293–302, 2014.

[15] Centers for Disease Control and Prevention, National Center for Health Statistics. Compressed Mortality File on CDC WONDER Online Database, released December 2017. Data are from the Compressed Mortality File 1999-2016 Series 20 No. 2V, 2017, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Accessed at http://wonder.cdc.gov on Mar 19, 2018.

[16] K. Matthews, J. Shepherd, and V. Sivarajasingham, "Violence-related injury and the price of beer in england and wales," *Applied Economics*, vol. 38, no. 6, pp. 661–670, 2006.

[17] A. Wolf, R. Gray, and S. Fazel, "Violence as a public health problem: An ecological study of 169 countries," *Social Science & Medicine*, vol. 104, pp. 220–227, 2014.

[18] J. Shepherd and N. Page, "The economic downturn probably reduced violence far more than licensing restrictions," *Addiction*, vol. 110, no. 10, pp. 1583–1584, 2015.

[19] D. R. Brillinger, "A biometrics invited paper with discussion: the natural variability of vital rates and associated statistics," *Biometrics*, vol. 42, pp. 693–734, 1986.

[20] N. Keyfitz, "Sampling variance of standardized mortality rates," *Human Biology*, vol. 38, no. 3, pp. 309–317, 1966.

[21] U. D. of Health, H. Services *et al.*, "Vital statistics of the united states: mortality, 1999 technical appendix," *National Center for Health Statistics*, 2004.

[22] M. Bertrand, E. Duflo, and S. Mullainathan, "How much should we trust differences-in-differences estimates?" *The Quarterly Journal of Economics*, vol. 119, no. 1, pp. 249–275, 2004.

[23] S. G. Donald and K. Lang, "Inference with difference-in-differences and other panel data," *The Review of Economics and Statistics*, vol. 89, no. 2, pp. 221–233, 2007.

[24] T. G. Conley and C. R. Taber, "Inference with difference in differences with a small number of policy changes," *The Review of Economics and Statistics*, vol. 93, no. 1, pp. 113–125, 2011.

[25] A. Abadie, A. Diamond, and J. Hainmueller, "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program," *Journal of the American Statistical Association*, vol. 105, no. 490, pp. 493–505, 2010.

[26] C. K. Crifasi, M. Merrill-Francis, A. McCourt, J. S. Vernick, G. J. Wintemute, and D. W. Webster, "Association between firearm laws and homicide in urban counties," *Journal of urban health*, vol. 95, no. 3, pp. 383–390, 2018.

[27] D. T. Campbell and J. C. Stanley, "Experimental and quasi-experimental designs for research," *Handbook of research on teaching*, pp. 171–246, 1963.

[28] R. Hasegawa, S. Deshpande, D. Small, and P. Rosenbaum, "Causal inference with two versions of treatment," 2017, posted on arXiv, https://arxiv.org/pdf/1705.03918.pdf.

[29] K. E. Rudolph, E. A. Stuart, J. S. Vernick, and D. W. Webster, "Association between connecticuts permit-to-purchase handgun law and homicides," *American journal of public health*, vol. 105, no. 8, pp. e49–e54, 2015.