Generalized Difference-in-differences Models: Robust Bounds

Kyunghoon Ban and Désiré Kédagni
Iowa State University - Department of Economics
khban@iastate.edu and dkedagni@iastate.edu

Selected Paper prepared for presentation at the 2022 Agricultural & Applied Economics Association Annual Meeting, Anaheim, CA; July 31-August 2

Generalized Difference-in-differences Models: Robust Bounds

Abstract

The difference-in-differences (DID) method identifies the average treatment effects on the treated (ATT) under the so-called parallel trends (PT) assumption. The most common and widely-used approach to justify the PT assumption is the pre-treatment period examination. If a null hypothesis of the same trend for both treatment and control groups in the pre-treatment periods is rejected, researchers believe less in PT and the DID results. This paper fills this gap by developing a generalized DID framework that utilizes all the information available not only from the pre-treatment periods but also from multiple data sources. Our approach interprets PT in a different way using a notion of selection bias, which enables us to generalize the standard DID estimand by defining an information set that may contain multiple pre-treatment periods or other baseline covariates. Our main assumption states that the selection bias in post-treatment periods lies within the set of all selection biases in pre-treatment periods. Based on the baseline information set we construct, we first provide an identified set for ATT that always contains the true ATT under our identifying assumption, and also the standard DID estimand. Secondly, we propose a class of criteria on the selection biases from the perspective of policymakers that can achieve a point identification of ATT.

Key words: Differences-in-differences, average treatment effect on the treated, baseline information

JEL classification: C14, C31, C35, C36

Introduction

The difference-in-differences (DID) technique is one of the most popular methods in the social sciences where an experimental research design cannot be used. DID requires observational data consisting of two different groups (a treatment group and a control group) and two time periods of pre-treatment and post-treatment, and the average treatment effects on the treated (ATT) is identified as the DID estimand under a certain set of assumptions. The key identifying assumption of interest is the so-called parallel trends (PT) assumption. This assumption states that the potential outcome variable for the treatment group would have followed the same trend as that for the control group if they had not been treated. However, it is difficult to empirically verify the PT assumption because it restricts a hypothetical quantity that is not identifiable. Accordingly, convincing readers to approve the PT assumption has been the most vital and controversial part of the DID literature. For instance, Kearney and Levine's (2015) ambitious identification strategy on discovering the effects of an MTV reality show on teen childbearing provided insightful findings that would not have been discovered without the study, but there have been heated debates on the validity of its main PT assumption as well (Jaeger, Joyce, and Kaestner 2018; Kahn-Lang and Lang 2019). The most common and widely understood approach for justifying the PT assumption is the pre-treatment period examination; if a null hypothesis of the same trend for both groups in the pre-treatment periods cannot be rejected, then it is claimed that the PT assumption also holds for the post-treatment periods. Still, rigorously speaking, the evidence of pre-treatment PT is different from the true PT of interest, and thus additional argument should be established for the PT assumption separately. Hence, this paper aims to develop a generalized DID framework that can utilize all the information available not only from the pre-treatment periods but also from multiple data sources.

Our approach is unique in that we interpret the PT assumption in a completely different way using a notion of selection bias, which enables us to generalize the standard DID estimand by defining an information set that can represent a set of multiple pre-treatment periods or other baseline covariates. Our main assumption states that the selection bias in post-treatment periods lies within the set of all selection biases in pre-treatment periods. We define selection bias as the mean-difference of the untreated potential outcome between the treatment and control groups. Based on the baseline information set we construct, we first provide an identified set for ATT that always contains the true ATT under our identifying assumption, and also the standard DID estimand, given that we are using a weaker assumption than the PT assumption. We provide an illustrative example where the standard DID estimand does not identify ATT while our bounds cover it. Secondly, we propose a class of criteria on the selection biases from the perspective of policymakers that can achieve a point identification of ATT. Finally, we illustrate the empirical relevance of our methodology using the Cawley et al. (2021) and Cai (2016) datasets.

Analytical Framework

Consider the following two-period model:

$$\begin{cases} Y_0 &= Y_0(0) \\ Y_1 &= Y_1(1)D + Y_1(0)(1-D) \end{cases} \tag{1}$$

where the vector (Y_0, Y_1, D, I_0, X) represents the observed data, while the vector $(Y_1(0), Y_1(1))$ is latent. In this model, the variables $Y_0, Y_1 \in \mathscr{Y}$ are respectively the observed outcomes in the baseline period 0 and the follow-up period 1, while $D \in \{0,1\}$ is the observed treatment that occurred between periods 0 and 1, $Y_1(0)$ and $Y_1(1)$ are the potential outcomes that would have been observed in period 0 had the treatment D been externally set to 0 and 1, respectively. The variable $Y_0(0)$ is the potential outcome that is realized in the baseline period when no individual/unit was treated. As is common in the DID literature, model (1) assumes that there is no anticipatory effect of the treatment, so that $Y_0(1) = Y_0(0)$. The set $I_0 \in \mathscr{I}_0$ contains information on baseline data, while X is the vector of covariates. In this paper, we are interested in identifying the average treatment effect on the treated (ATT) defined as

$$ATT \equiv \mathbb{E}[Y_1(1) - Y_1(0)|D = 1].$$

Define the standard ordinary least squares (OLS) estimand as

$$\theta_{OLS} \equiv \mathbb{E}[Y_1|D = 1] - \mathbb{E}[Y_1|D = 0].$$

We have

$$\begin{aligned} \theta_{OLS} &= \mathbb{E}[Y_1(1)|D = 1] - \mathbb{E}[Y_1(0)|D = 0], \\ &= \mathbb{E}[Y_1(1) - Y_1(0)|D = 1] + \mathbb{E}[Y_1(0)|D = 1] - \mathbb{E}[Y_1(0)|D = 0], \\ &= ATT + SB_1, \end{aligned}$$

where $SB_1 \equiv \mathbb{E}[Y_1(0)|D = 1] - \mathbb{E}[Y_1(0)|D = 0]$ is the selection bias.

Question: How to obtain the selection bias SB_1?

1. Randomized Experiments Selection bias $SB_1 = 0$.

2. Parellel Trends Selection bias $SB_1 = SB_0 \equiv \mathbb{E}[Y_0|D = 1] - \mathbb{E}[Y_0|D = 0]$ (identified).

The parellel trends assumption is a *bias stability* assumption.

3. Our approach We know the range of the selection bias using baseline information $\mathscr{I}_0$.

Assumption 1 (Known bounds on selection bias)**.**

$$SB_1 \in \left[\inf_{\iota_0 \in \mathscr{I}_0} SB_0(\iota_0), \sup_{\iota_0 \in \mathscr{I}_0} SB_0(\iota_0) \right] \equiv \Delta_{SB_0},$$

where $SB_0(\iota_0) \equiv \mathbb{E}[Y_0|D = 1, I_0 = \iota_0] - \mathbb{E}[Y_0|D = 0, I_0 = \iota_0]$ *is the selection bias in the baseline period conditional on the information* $\{I_0 = \iota_0\}$.

Assumption 1 is weaker than the standard "parallel/common trends" assumption. Indeed, if $\mathscr{I}_0$ is the singleton of a single baseline data, then Assumption 1 is equivalent to

$$SB_1 = \mathbb{E}[Y_0|D=1] - \mathbb{E}[Y_0|D=0] \equiv SB_0,$$

which is equivalent to the parallel trends assumption:

$$\mathbb{E}[Y_1(0) - Y_0(0)|D=1] = \mathbb{E}[Y_1(0) - Y_0(0)|D=0].$$

Note that the set $\mathscr{I}_0$ could contain all pre-treatment periods, information from other data sources, or observed baseline characteristics. For example, suppose that $\mathscr{I}_0$ contains gender. The standard parallel trends assumption conditional on gender states that the selection bias for males in period 0 would be the same for males in period 1, and similarly for females. Our assumption 1 allows the selection bias for females in period 1 to be equal to that of males in period 0, and vice versa. We believe that this latter assumption is more flexible.

Proposition 1. *Suppose that model (1) along with Assumption 1 holds. Then, the following bounds hold for the ATT:*

$$ATT \in \left[\theta_{OLS} - \sup_{\iota_0 \in \mathscr{I}_0} SB_0(\iota_0), \theta_{OLS} - \inf_{\iota_0 \in \mathscr{I}_0} SB_0(\iota_0)\right] \equiv \Theta_I.$$

These bounds are sharp, and Θ_I is the identified set for the ATT.

The bounds in Proposition 1 are never empty, as they always contain the standard DID estimand under the parallel trends assumption. However, they may not contain θ_{OLS}, as 0 may not lie within Δ_{SB_0}.

Comparison with Rambachan and Roth's (2020) approach

First, suppose the information set I_0 contains two pre-treatment periods -1 and 0, such that $\mathscr{I}_0 = \{-1, 0\}$. Define $\delta \equiv (\delta_{-1}, \delta_1)'$, where:

$$\delta_1 \quad = \quad \mathbb{E}[Y_1(0) - Y_0(0)|D=1] - \mathbb{E}[Y_1(0) - Y_(0)|D=0],$$

$$\delta_{-1} = \mathbb{E}[Y_{-1}(0) - Y_0(0)|D=1] - \mathbb{E}[Y_{-1}(0) - Y_(0)|D=0].$$

Notice that $\delta_1 = SB_1 - SB_0$, and $\delta_{-1} = SB_{-1} - SB_0$.

Smoothness restrictions The differential trends evolve smoothly over time with slope changing by no more than M between consecutive periods:

$$\Delta^{SD}(M) \equiv \{\delta : |(\delta_1 - \delta_0) - (\delta_0 - \delta_{-1})| \leq M\},$$

where δ_0 is normalized to be equal to zero. We then have $\Delta^{SD}(M) \equiv \{\delta : |\delta_1 + \delta_{-1}| \leq M\}$. The parameter $M \geq 0$ is like a sensitivity parameter and governs the amount by which the slope of the differential trends can change between consecutive periods.

Under the smoothness restriction, we obtain the following bounds on the selection bias SB_1:

$$2SB_0 - SB_{-1} - M \leq SB_1 \leq 2SB_0 - SB_{-1} + M$$

Our bounding approach yields the following bounds on SB_1:

$$\min\{SB_{-1}, SB_0\} \leq SB_1 \leq \max\{SB_{-1}, SB_0\}.$$

In Appendix , we show that if $SB_{-1} \neq SB_0$, there exists no value of M such that the above two sets of bounds on SB_1 coincide. Furthermore, we show that there exist no values of M for which Rambachan and Roth's (2020) bounds are tighter than ours, while there exist values of M for which our bounds are tighter than theirs.

Bounding relative magnitudes This approach bounds the worst-case post-treatment violation of parallel trends in terms of the worst-case violation in the pre-treatment period:

$$\Delta^{RM}(\bar{M}) \equiv \{\delta : |\delta_1 - \delta_0| \leq \bar{M}|\delta_0 - \delta_{-1}|\},$$

where $\bar{M} \geq 0$ behaves as a sensitivity parameter. This implies the following bounds on SB_1:

$$SB_0 - \bar{M}|SB_{-1} - SB_0| \leq SB_1 \leq SB_0 + \bar{M}|SB_{-1} - SB_0|.$$

In Appendix , we show that if $SB_{-1} \neq SB_0$, there exists no value of $\bar{M}$ such that the above bounds on SB_1 coincide with ours. When $\bar{M} > 1$, our bounds are tighter than Rambachan and Roth's (2020), and there exist no values of $\bar{M}$ for which their bounds are tighter than ours.

Second, our approach covers the case where no pre-treatment trend exists, but there are multiple elements (data sets) available in the information set at period 0. Their methodology is silent about such a case.

Third, our approach does not require the knowledge of a sensitivity parameter, while theirs does. How to choose the values of the sensitivity parameters M and $\bar{M}$ remains unclear in their approach.

Generalized DID estimand

The main question we are trying to answer is how to obtain the selection bias SB_1. Given the baseline information I_0, we are going to assume that the decision maker will choose the selection bias SB_1 in such a way that a loss function is minimized.

Assumption 2. *Let $\mathscr{L}(SB_1)$ be the decision maker's loss function when she assumes that the selection bias is SB_1 in the presence of the baseline information I_0. The decision maker chooses SB_1 to minimize the loss $\mathscr{L}(SB_1)$.*

In this paper, we consider the class of p-norm losses defined as:

$$\mathscr{L}_p(SB_1) = \left(\mathbb{E}_{I_0}\left[|SB_1 - SB_0(I_0)|^p\right]\right)^{1/p},$$

where $1 \leq p \leq \infty$. We are going to derive the optimal selection bias SB_1 for $p \in \{1, 2, \infty\}$.

Definition 1. *We define the generalized difference-in-differences (GDID) estimand as*

$$\theta_{GDID} \equiv \theta_{OLS} - \arg\min \mathscr{L}(SB_1). \tag{2}$$

$L1$ loss: Mean absolute error (MAE) $\mathscr{L}_1(SB_1) = \mathbb{E}_{I_0}\left[|SB_1 - SB_0(I_0)|\right]$

Given this $L1$ loss function, under Assumption 2, the decision maker solves the following optimization problem:

$$\min_{SB_1} \mathbb{E}_{I_0}\left[|SB_1 - SB_0(I_0)|\right].$$

The optimal decision is to set the selection SB_1 to be equal to the median selection bias in the baseline period, i.e., $SB_1 = Med_{I_0}(SB_0(I_0))$. In such a case, the ATT is given by

$$\theta_{GDID} = \theta_{OLS} - Med_{I_0}(SB_0(I_0)).$$

$L2$ loss: Root mean square error (RMSE) $\mathscr{L}_2(SB_1) = \left(\mathbb{E}_{I_0}\left[|SB_1 - SB_0(I_0)|^2\right]\right)^{1/2}$

Minimizing the RMSE is equivalent to minimizing the mean square error (MSE). Therefore, under Assumption 2, the decision maker solves the following optimization problem:

$$\min_{SB_1} \mathbb{E}_{I_0}\left[(SB_1 - SB_0(I_0))^2\right].$$

This yields an optimal decision for the selection SB_1 to be set equal to the average selection bias in the baseline period, i.e., $SB_1 = \mathbb{E}_{I_0}[SB_0(I_0)]$. Hence, we have

$$\theta_{GDID} = \theta_{OLS} - \mathbb{E}_{I_0}[SB_0(I_0)].$$

$L\infty$ loss: Maximal regret $\mathscr{L}_\infty(SB_1) = \operatorname{ess\,sup}_{\mathscr{I}_0} |SB_1 - SB_0(I_0)|$, where ess sup denotes essential supremum and is defined as follows: $\operatorname{ess\,sup}_{\mathscr{I}_0} f = \inf\{M : \mathbb{P}(\iota_0 \in \mathscr{I}_0 : f(\iota_0) \leq M) = 1\}$.

For simplicity, assume $\operatorname{ess\,sup}_{\mathscr{I}_0} |SB_1 - SB_0(I_0)| = \sup_{\iota_0 \in \mathscr{I}_0} |SB_1 - SB_0(\iota_0)|$. Then

$$\begin{aligned}
\mathscr{L}_\infty(SB_1) &= \sup_{\iota_0 \in \mathscr{I}_0} |SB_1 - SB_0(\iota_0)|, \\
&= \sup_{\iota_0 \in \mathscr{I}_0} \max\left\{SB_1 - SB_0(\iota_0), SB_0(\iota_0) - SB_1\right\},
\end{aligned}$$

$$= \max\left\{SB_1 - \inf_{\iota_0\in\mathscr{I}_0} SB_0(\iota_0), \sup_{\iota_0\in\mathscr{I}_0} SB_0(\iota_0) - SB_1\right\}.$$

Therefore, the minimum of $\mathscr{L}_\infty(SB_1)$ is obtained when the two arguments of the max function are equal, i.e., $SB_1 - \inf_{\iota_0\in\mathscr{I}_0} SB_0(\iota_0) = \sup_{\iota_0\in\mathscr{I}_0} SB_0(\iota_0) - SB_1$. This implies $SB_1 = \frac{1}{2}(\inf_{\iota_0\in\mathscr{I}_0} SB_0(\iota_0) + \sup_{\iota_0\in\mathscr{I}_0} SB_0(\iota_0))$, and $\mathscr{L}_\infty(SB_1) = \frac{1}{2}(\inf_{\iota_0\in\mathscr{I}_0} SB_0(\iota_0) + \sup_{\iota_0\in\mathscr{I}_0} SB_0(\iota_0))$.

This optimization problem with the $L\infty$ loss is equivalent to a minimax criterion, and yields the mid-point of the bounds on SB_1 stated in Assumption 1. Hence, the ATT is given by

$$\theta_{GDID} = \theta_{OLS} - \frac{1}{2}\left(\inf_{\iota_0\in\mathscr{I}_0} SB_0(\iota_0) + \sup_{\iota_0\in\mathscr{I}_0} SB_0(\iota_0)\right).$$

Note that in all cases, if the information in the baseline period is a singleton, then the optimal SB_1 is the selection bias in the baseline period SB_0, which is equivalent to the parallel trends assumption. Unlike the $L1$ and $L2$ loss functions, the $L\infty$ loss function does not require the knowledge of the distribution of $SB_0(I_0)$, and is easy to compute. However, when the distribution of $SB_0(I_0)$ is uniform over $\left[\inf_{\iota_0\in\mathscr{I}_0} SB_0(\iota_0), \sup_{\iota_0\in\mathscr{I}_0} SB_0(\iota_0)\right]$, then the optimal selection bias SB_1 is the same in all three cases.

Let Λ denote the set of possible distributions for $SB_0(I_0)$, and $SB_1(SB_0, \lambda)$ denote the optimal selection bias in period 1 given the distribution $\lambda \in \Lambda$ for $SB_0(I_0)$. Define $ATT_\lambda \equiv \theta_{OLS} - SB_1(SB_0, \lambda)$.

Definition 2. *We define the robust DID bounds as follows:*

$$ATT \in \left[\inf_{\lambda\in\Lambda} ATT_\lambda, \sup_{\lambda\in\Lambda} ATT_\lambda\right].$$

The following lemma holds.

Lemma 1. *The robust DID bounds coincide with the bounds in Proposition 1 for the $L1$, $L2$, and $L\infty$ loss functions.*

A sufficient condition for the generalized DID estimand to be equal to the ATT is that the potential outcome $Y_1(0)$ satisfies:

$$Y_1(0) = \mathbb{E}[Y_1|D=0] + SB_1 + \varepsilon,$$

where $\mathbb{E}[\varepsilon|D=1] = 0$.

Numerical illustration Consider the following data generating process (DGP):

$$(3) \qquad \begin{cases} Y_{it} & = & [Y_{1i}(1)D_i + Y_{1i}(0)(1-D_i)]\,\mathbb{1}\{t=1\} \\ & & \quad + \left[\sum_{\iota_0 \in \mathscr{I}_0} Y_{\iota_0 i}(0)\mathbb{1}\{I_{0i} = \iota_0\}\right]\mathbb{1}\{t=0\} \\ D_i|I_{0i} & \overset{i.i.d.}{\sim} & \mathscr{B}ernoulli(0.6) \\ Y_{1i}(1)|I_{0i} & \overset{i.i.d.}{\sim} & D_i\mathscr{U}_{[6,8]} + (1-D_i)\mathscr{U}_{[8,10]} \end{cases}$$

For simplicity, assume $\mathscr{I}_0 = \{-1, 0\}$, and $Y_1(0) = \lambda Y_0(0) + (1-\lambda)Y_{-1}(0)$, where $\lambda \in [0,1]$.

$$(4) \qquad \begin{cases} Y_{-1i}(0)|I_{0i} & \overset{i.i.d.}{\sim} & D_i\mathscr{U}_{[0,2]} + (1-D_i)\mathscr{U}_{[3,5]} \\ Y_{0i}(0)|I_{0i} & \overset{i.i.d.}{\sim} & D_i\mathscr{U}_{[3,5]} + (1-D_i)\mathscr{U}_{[4,6]} \\ I_{0i} & \overset{i.i.d.}{\sim} & \mathscr{B}ernoulli(0.5) - 1 \end{cases}$$

A sufficient condition for Assumption 1 to hold in model (3) is that $Y_1(0)$ belongs to the convex hull generated by the sets $\{(Y_{-1}(0), 0), (0, Y_0(0))\}$ in the space $\mathbb{R}^2$. This generalizes to the case where the cardinality of $\mathscr{I}_0$ is finite.

Suppose $SB_0(-1) \neq SB_0(0)$, and $\lambda = 1$, then the parallel trends assumption holds. If instead, $\lambda \neq 1$ then parallel trends fail, and the standard DID estimand does not identify the ATT. However, in these two scenarios, the bounds in Proposition 1 cover the ATT.

In the design above, $SB_0(-1) = \mathbb{E}[Y_{-1i}(0)|D_i = 1] - \mathbb{E}[Y_{-1i}(0)|D_i = 0] = 1 - 4 = -3$, while $SB_0(0) = \mathbb{E}[Y_{0i}(0)|D_i = 1] - \mathbb{E}[Y_{0i}(0)|D_i = 0] = 4 - 5 = -1$. The true ATT is $ATT = \mathbb{E}[Y_{1i}(1) - Y_{1i}(0)|D_i = 1] = 7 - (4\lambda + (1-\lambda))$, and $\theta_{OLS} = 7 - (5\lambda + 4(1-\lambda)) = 3 - \lambda$.

Suppose $\lambda = 0.6$. Then $ATT = 6 - 3 * 0.6 = 4.2$, $\theta_{OLS} = 3 - 0.6 = 2.4$. The bounds in Proposition 1 are $[2.4-(-1), 2.4-(-3)] = [3.4, 5.4]$, which covers the true ATT. However, the standard DID estimand yields $\theta_{DID} = 2.4 - (-1) = 3.4$, which is biased.

Estimation and Inference

We can write the robust DID bounds Θ_I as the largest interval induced by the standard DID estimands as follows:

$$\begin{aligned} \Theta_I &= \left[\theta_{OLS} - \sup_{\iota_0 \in \mathscr{I}_0} SB_0(\iota_0), \theta_{OLS} - \inf_{\iota_0 \in \mathscr{I}_0} SB_0(\iota_0)\right], \\ &= \left[\inf_{\iota_0 \in \mathscr{I}_0} \{\theta_{OLS} - SB_0(\iota_0)\}, \sup_{\iota_0 \in \mathscr{I}_0} \{\theta_{OLS} - SB_0(\iota_0)\}\right]. \end{aligned}$$

We can then take the largest interval induced by the confidence intervals of all DID estimands $\theta_{OLS} - SB_0(\iota_0)$ to obtain valid confidence bounds for Θ. But, these confidence bounds could be too conservative.

Extension to multiple treatment periods

Consider the following multiple treatment periods model:

$$(5) \quad Y_{it} = \sum_{s=1}^{T} \left[Y_{is}(1) D_{is} + Y_{is}(0)(1 - D_{is})\right] \mathbb{1}\{t = s\} + \left[\sum_{\iota_0 \in \mathscr{I}_0} Y_{\iota_0 i}(0) \mathbb{1}\{I_0 = \iota_0\}\right] \mathbb{1}\{t = 0\}.$$

When we are interested in the static treatment effect in each period, we can compare each period's outcome to the baseline outcome. But, if instead we are interested in the dynamic treatment effect, then in order to obtain the ATT at period t, we will consider the information set $\mathscr{I}_{t-1} = \mathscr{I}_0 \cup \{1, \ldots, t-1\}$ as the updated baseline information set for period t.

Empirical application

We adopt Sant'Anna and Zhao's (2020) locally efficient doubly robust DID estimators to our approach considering multiple pre-treatment periods as the information set $\mathscr{I}_0$. To be specific, for each $\iota_0 \in \mathscr{I}_0$, we obtain $\hat{\theta}_{DID}(\iota_0)$ as an estimate for $\theta_{OLS} - SB_0(\iota_0)$ (with or without covariates) and its confidence interval $CI_{DID}(\iota_0)$ following Sant'Anna and Zhao (2020). Then we construct an estimate of the set Θ_I using the doubly robust DID estimates as:

$$\Theta_I \quad = \quad \left[\inf_{\iota_0 \in \mathscr{I}_0} \{\hat{\theta}_{DID}(\iota_0)\}, \sup_{\iota_0 \in \mathscr{I}_0} \{\hat{\theta}_{DID}(\iota_0)\} \right]. \tag{6}$$

Accordingly, the (conservative) confidence interval for the set Θ_I would be the convex hull of $CI_{DID}(\iota_0)$ for all $\iota_0 \in \mathscr{I}_0$. Note that this can be applied to any dataset with multiple pre-treatment periods.

Cawley et al. (2021) They examine the pass-through of a tax of two cents per ounce on sugar-sweetened beverages (SSB tax) enacted in Boulder, Colorado, using the standard DID framework. They considered both store and restaurant prices and collected two different datasets for each of them: hand-collected data and Nielsen retail scanner data for the store prices, and hand-collected data and web-scrapped (OrderUp.com) data for the restaurant prices. Hence, this exercise could have been the best example for us to explore the information set consisting of the multiple datasets, but we focus on utilizing multiple pre-treatment periods of the hand-collected datasets in this subsection due to the data limitation.[1]

Each dataset is bimonthly-collected and has four periods April, June, August, and October, where the tax was imposed on July 1st of the same year. Thus, our information set has two elements April and June. Moreover, we can implement the event-study type DID analysis (static treatment effects in multiple treatment periods model as in Equation (5)) to capture the non-parametric evolution of the treatment effects over the post-treatment periods. The control community is Fort Collins, Colorado, which is geographically close to Boulder and similar in demographic characteristics as well. Hence, the standard PT assumption states that the average equilibrium beverage price differences between Boulder and Fort Collins in the post-treatment periods would have been the same as the average equilibrium price differences in the pre-treatment periods if there had not been the SSB tax in Boulder. On the other hand, our GDID model assumes that the average equilibrium price differences without the tax in the post-treatment periods would have lain between the average equilibrium price differences in April and June between the two cities.

Tables 1 shows the standard DID results and our GDID results together, where the first three rows (*post_tax*, *reg_tax*, and *untax*) are obtained from the hand-collected data for store prices, and the fourth row (*fount*) is for ATT estimates of restaurant fountain drinks. In particular, *post_tax* uses post prices on the shelves, *reg_tax* uses prices at the register,[2] and *untax* uses prices of products irrelevant to SSB tax (e.g., diet soda, products in which milk is the primary ingredient, alcoholic mixers, or coffee drinks) for the blind test. The first column presents the locally efficient doubly robust DID estimates (Sant'Anna and Zhao 2020) for ATT, and the second and third columns are corresponding 95% confidence intervals. The fourth and fifth columns show lower and upper bounds of our identified for ATT in Proposition 1, and the corresponding 95% confidence intervals are given in the sixth and seventh columns. Note that we cannot reject the null hypothesis that the effect on the post prices is not different from zero under a significance level of 5% from our GDID model whereas the null hypothesis would be rejected under the standard DID model. On the other hand, we can reject the null of no effect on register prices and restaurant fountain drink prices regardless of whether we adopt the GDID model or the standard DID model, implying that the same quantitative conclusion can be drawn from the GDID model with a set of weaker assumptions. Lastly, it has to be noted that the pass-through rate higher than 100% cannot be rejected from the GDID model whereas the standard DID estimates rule out the case; the market could be imperfectly competitive (Anderson, Palma, and Kreider 2001).

Figure 1 shows the event-study type GDID estimates over the post-treatment periods. The red and dark blue dashed lines are the upper and lower bound of the dynamic treatment effects, and 95% confidence regions are depicted as gray areas with dotted lines. Although we have only two post-treatment periods, we observe the following remarks. First, the pass-through rates of SSB tax on store prices seem relatively stable over time compared to the restaurant fountain drink prices. Second, given the increasing pass-through rates on restaurant drinks, especially with the 100% pass-through rate within the bound estimates in Oct, it would be interesting to examine further whether or not there is any excessive market power exercised through the restaurant drink prices in later periods. Finally, the figure for untaxed product prices shows that the impact of SSB tax seems to be transmitted to the other drinks in Boulder city over time, but it is not statistically significant.

Cai (2016) Cai (2016) examine the impact of insurance provision on tobacco production using a household-level panel dataset provided by the Rural Credit Cooperative (RCC), the main rural bank in China. The regression equation used in Cai (2016) is as follows:

$$(7) \quad Y_{irt} = \alpha_0 + \alpha_1 After_t + \alpha_2 Insurance_{ir} + \alpha_3 After_t \times Insurance_{ir} + \beta X + \varepsilon_{irt},$$

where i, r, t are household, region, and year indices, respectively, and Y is the outcome variable (*area_tob*: area of tobacco production measured in mu,[3] *tobshare*: share of tobacco production in total area of agricultural production). The covariates X linearly enters the equation to be controlled for and consist of the household size, education level, and age of the household head. Note that α_3 in Equation (7) can identify ATT under the standard parallel trend assumption.

Tables 2 shows the standard DID results and our GDID results together for each outcome variable *area_tob* and *tobshare* with or without the covariates (X). The first column presents the locally efficient doubly robust DID estimates (Sant'Anna and Zhao 2020) for ATT, and the second and third columns are corresponding 95% confidence intervals. The fourth and fifth columns show lower and upper bounds of our identified for ATT in Proposition 1, and the corresponding 95% confidence intervals are given in the sixth and seventh columns. Note that both the DID estimates and GDID region estimates can reject the null hypothesis that the effect is not different from zero under a significance level of 5%, but the latter can draw the same conclusion from the weaker assumption than the standard parallel trend assumption.

Figure 2 shows the event-study type DID analysis (static treatment effects in multiple treatment periods model as in Equation (5)) to capture the non-parametric evolution of the treatment effects over the post-treatment periods. The red and dark blue dashed lines are the upper and lower bound of the dynamic treatment effects, and 95% confidence regions are depicted as gray areas with dotted lines. From this analysis, we observe that the initial impact of the insurance provision on the tobacco production area is relatively small that the null hypothesis cannot be rejected, but it becomes substantial as time goes.

Notes

[1]Nielsen retail scanner data are proprietary, and the unit price information is not available in the web-scrapped (OrderUp.com) data.

[2]Cawley et al. (2021) found that not all retailers included the tax in the posted (or shelf) prices; i.e., some retailers added the tax at the register making it less salient.

[3]1 mu corresponds to 1/15 ha.

References

Anderson, S.P., A.d. Palma, and B. Kreider. 2001. "Tax incidence in differentiated product oligopoly." *Journal of Public Economics* 81:173–192.

Cai, J. 2016. "The Impact of Insurance Provision on Household Production and Financial Decisions." *American Economic Journal: Economic Policy* 8:44–88.

Cawley, J., D. Frisvold, D. Jones, and C. Lensing. 2021. "The Pass-Through of a Tax on Sugar-Sweetened Beverages in Boulder, Colorado." *American Journal of Agricultural Economics* 103:987–1005.

Jaeger, D.A., T.J. Joyce, and R. Kaestner. 2018. "A Cautionary Tale of Evaluating Identifying Assumptions: Did Reality TV Really Cause a Decline in Teenage Childbearing?" *Journal of Business & Economic Statistics* 38:1–44.

Kahn-Lang, A., and K. Lang. 2019. "The Promise and Pitfalls of Differences-in-Differences: Reflections on 16 and Pregnant and Other Applications." *Journal of Business & Economic Statistics* 38:1–14.

Kearney, M.S., and P.B. Levine. 2015. "Media Influences on Social Outcomes: The Impact of MTV's 16 and Pregnant on Teen Childbearing." *American Economic Review* 105:3597–3632.

Rambachan, A., and J. Roth. 2020. "An Honest Approach to Parallel Trends." *Working Paper*, pp. .

Sant'Anna, P.H., and J. Zhao. 2020. "Doubly robust difference-in-differences estimators." *Journal of Econometrics* 219:101–122.

Proof of Proposition 1

Validity of the bounds Proven in the main text.

Sharpness of the bounds

Proof. Suppose $\mathscr{I}_0$ is finite. Then the lower and upper bounds for ATT are attained when

$$Y_1(0) = \mathbb{E}[Y_1|D=0] + \min_{\iota_0 \in \mathscr{I}_0} SB(\iota_0) + \varepsilon_\ell,$$

and

$$Y_1(0) = \mathbb{E}[Y_1|D=0] + \max_{\iota_0 \in \mathscr{I}_0} SB(\iota_0) + \varepsilon_u,$$

respectively, where $\mathbb{E}[\varepsilon_\ell|D=1]=0$, and $\mathbb{E}[\varepsilon_u|D=1]=0$. Any point θ within Θ_I can be written as

$$\theta = \theta_{OLS} - \left(\lambda \min SB(\iota_0) + (1-\lambda)\max SB(\iota_0)\right),$$

where $\lambda \in (0,1)$. Therefore, θ is achieved when

$$Y_1(0) = \mathbb{E}[Y_1|D=0] + \lambda \min_{\iota_0 \in \mathscr{I}_0} SB(\iota_0) + (1-\lambda) \max_{\iota_0 \in \mathscr{I}_0} SB(\iota_0) + \varepsilon,$$

where $\mathbb{E}[\varepsilon|D=1]=0$.

We need to define a joint distribution of the vector $\left(\{\tilde{Y}_{\iota_0}(0)\}_{\iota_0 \in \mathscr{I}_0}, \tilde{Y}_1(0), \tilde{Y}_1(1), \tilde{D}\right)$ that will yield any value in the identified set Θ_I. We define $\tilde{Y}_{\iota_0}(0) = Y_{\iota_0}$ for all $\iota_0 \in \mathscr{I}_0$, $\tilde{Y}_1(0)$ is as previously defined for the lower/upper bound and any interior point of Θ_I, $\tilde{D} = D$, and $\tilde{Y}_1(1) = Y_1$. □

Comparison with Rambachan and Roth's (2020): Proofs

Smoothness restrictions. We have: $2SB_0 - SB_{-1} - M = \min\{SB_{-1}, SB_0\}$ implies $M = 2SB_0 - SB_{-1} - \min\{SB_{-1}, SB_0\}$, and $2SB_0 - SB_{-1} + M = \max\{SB_{-1}, SB_0\}$ implies

$M = \max\{SB_{-1}, SB_0\} - 2SB_0 + SB_{-1}$. Therefore $2SB_0 - SB_{-1} - \min\{SB_{-1}, SB_0\} = \max\{SB_{-1}, SB_0\} - 2SB_0 + SB_{-1}$ implies $SB_{-1} = SB_0$.

Rambachan and Roth's (2020) bounds are tighter than ours if and only $2SB_0 - SB_{-1} - M > \min\{SB_{-1}, SB_0\}$, and $2SB_0 - SB_{-1} + M < \max\{SB_{-1}, SB_0\}$, i.e.,

$$\begin{aligned} M &< \min\{\max\{-(SB_0 - SB_{-1}), -2(SB_0 - SB_{-1})\}, \max\{SB_0 - SB_{-1}, 2(SB_0 - SB_{-1})\}\}, \\ &= \min\{SB_0 - SB_{-1}, SB_{-1} - SB_0\} \leq 0. \end{aligned}$$

Our bounds are tighter than theirs if and only if

$$\begin{aligned} M &> \max\{\max\{-(SB_0 - SB_{-1}), -2(SB_0 - SB_{-1})\}, \max\{SB_0 - SB_{-1}, 2(SB_0 - SB_{-1})\}\}, \\ &= 2|SB_0 - SB_{-1}|. \end{aligned}$$

Bounding relative magnitudes. We have: $SB_0 - \bar{M}|SB_{-1} - SB_0| = \min\{SB_{-1}, SB_0\}$ and $SB_0 + \bar{M}|SB_{-1} - SB_0| = \max\{SB_{-1}, SB_0\}$ imply $\bar{M}|SB_{-1} - SB_0| = SB_0 - \min\{SB_{-1}, SB_0\}$, and $SB_0 + SB_0 - \min\{SB_{-1}, SB_0\} = \max\{SB_{-1}, SB_0\}$, that is, $2SB_0 = SB_{-1} + SB_0$, which implies $SB_{-1} = SB_0$.

Rambachan and Roth's (2020) bounds are tighter than ours if and only $SB_0 - \bar{M}|SB_{-1} - SB_0| > \min\{SB_{-1}, SB_0\}$, and $SB_0 + \bar{M}|SB_{-1} - SB_0| < \max\{SB_{-1}, SB_0\}$, i.e., $\bar{M}|SB_{-1} - SB_0| < \min\{\max\{SB_{-1} - SB_0, 0\}, \max\{SB_0 - SB_{-1}, 0\}\} = 0$. Our bounds are tighter than theirs if and only if $\bar{M}|SB_{-1} - SB_0| > |SB_{-1} - SB_0|$, i.e., $\bar{M} > 1$ if $SB_{-1} \neq SB_0$.

Figures

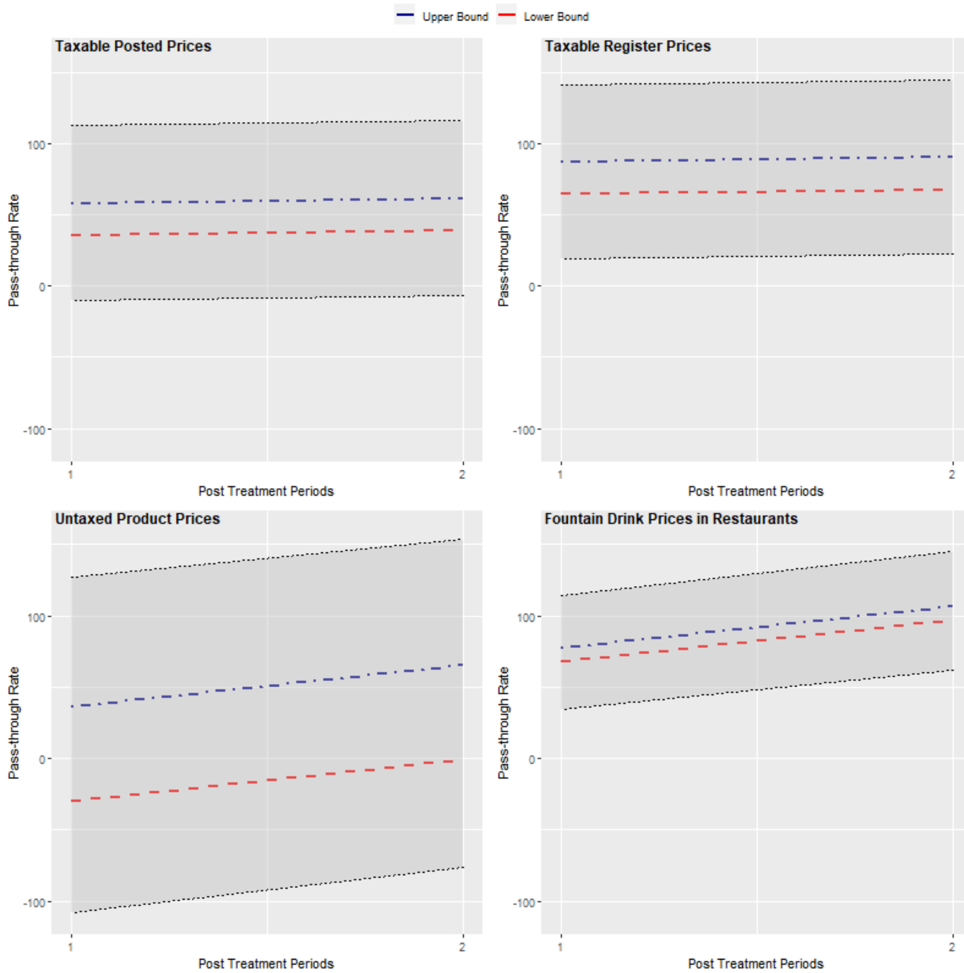

Figure 1. Treatment Effects Evolution over the Post-treatment Periods (SSB tax)

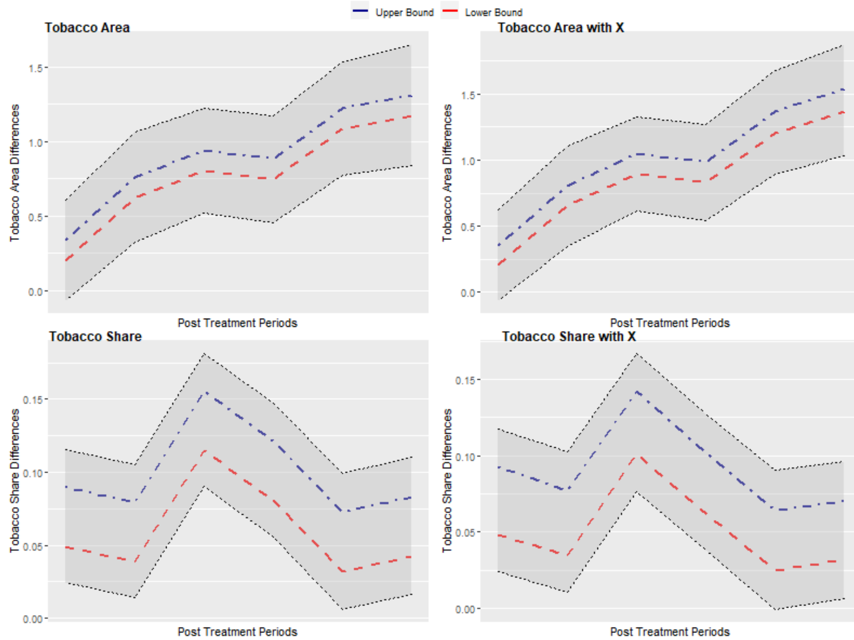

Figure 2. Treatment Effects Evolution over the Post-treatment Periods (Cai 2016)

Tables

Table 1. GDID Illustration on the Pass-through Rate (%) of SSB tax

	DID_PE	DID_CILB	DID_CIUB	GDID_PELB	GDID_PEUB	GDID_CILB	GDID_CIUB
post_tax	45.99	11.48	80.49	37.29	59.72	-1.91	108.99
reg_tax	74.61	40.07	109.14	65.91	88.34	26.68	137.63
untax	6.61	-51.30	64.53	-14.56	51.41	-80.84	132.07
fount	87.59	61.95	113.24	83.20	92.30	53.08	125.85

Table 2. GDID Illustration on the Insurance Provision and Tobacco Production

	DID_PE	DID_CILB	DID_CIUB	GDID_PELB	GDID_PEUB	GDID_CILB	GDID_CIUB
area_tob	0.84	0.69	0.99	0.77	0.91	0.55	1.13
area_tob with X	0.91	0.76	1.06	0.81	0.96	0.59	1.18
tobshare	0.09	0.07	0.10	0.06	0.10	0.04	0.12
tobshare with X	0.07	0.06	0.09	0.05	0.08	0.03	0.10