Aalto-yliopisto
Aalto-universitetet
Aalto University

Bachelor's Programme in Economics

ADVANCING APPLICABILITY OF DIFFERENCE-IN-DIFFERENCE METHODS IN NON-EXPERIMENTAL SETTINGS

An empirical analysis of karaoke's impact on hospitality venues' return on assets in the UK

Ngoc Anh Pham

Bachelor’s thesis

2023

Author: Ngoc Anh Pham

Title of thesis: Advancing Applicability Of Difference-In-Difference Methods Applicability In Non-Experimental Settings: An empirical analysis of karaoke's impact on hospitality venues' return on assets in the UK

Program: Bachelor of Science in Economics and Business Administration

Major: Economics

Thesis supervisor: Prof. Miri Stryjan

Thesis advisor(s): Prof. Pablo Ernesto Warnes

Collaborative partner: Singa Oy

Date: 21.11.2023 **Number of pages:** 21 **Language** English

Abstract

This thesis aims to bridge the gap between theory and practice in employing the Difference-in-Differences (DiD) method within non-experimental settings. The first part of the thesis systematically explores DiD mechanisms, encompassing basic principles, assumptions, and recent advancements. Challenges such as (i) treatment timing variation, (ii) treatment effect heterogeneity, and (iii) potentially violated unconditional parallel assumptions are also addressed using the new CS estimator introduced by Callaway and Sant'Anna (2021). In the second part of the thesis, the author demonstrates the application of the CS estimator in an analysis of the impact of Karaoke rooms on hospitality venues' financial performance, offering a clear pathway for practitioners to employ and interpret DiD methodologies in their research endeavors effectively.

Keywords Difference-in-Difference, staggered treatment adoptions, parallel trend assumption, treatment effect heterogeneity

Table of contents

1. INTRODUCTION

1.1 Research objectives

According to Roth et al. (2023), in contemporary economics, the DiD (Difference-in-Differences) method stands as a prominent tool for estimating causal effects within non-experimental settings. Despite its widespread application, the nuances of applying DiD in real business scenarios pose considerable challenges, such as staggered treatment adoptions, violated unconditional parallel trends, and potential heterogeneous treatment effects (Roth et al. 2023).

This thesis aims to clarify the basic concepts and underlying assumptions of the Difference-in-Differences (DiD) approach within its preexisting framework. It also aims to explore the latest developments in the DiD technique, specifically tackling issues related to violated unconditional parallel trends, uneven treatment uptake, and inconsistent treatment outcomes. One important area of emphasis is the application of the DiD estimating method introduced by Callaway and Sant'Anna in 2021, the CS estimator, which provides a more robust method for dealing with those issues. Moreover, this thesis aims to provide future users of these techniques with a thorough understanding of CS setups and clear instructions for their use in practice. This is done through an empirical analysis using the advanced DiD framework and the CS estimator to examine the causal relationship between karaoke rooms and the financial performance of hospitality venues.

1.2 Research questions

The central research question is whether utilizing Callaway and Sant'Anna's estimator improves the robustness of estimating the causal effect in staggered treatment adoption cases where unconditional parallel trends might be violated and treatment effects are heterogeneous.

1.3 Scope of research

Firstly, the research methodology review part will go through the mechanism of the canonical framework, where certain problems exist when studying the causal effects in non-experimental settings, such as inconsistent treatment timing and treatment effect; and unheld unconditional parallel trend assumption. Subsequently, the advanced estimator introduced by Callaway and Sant'Anna (2021) is discussed as an improved method to tackle these problems.

Secondly, the research empirical evaluation will study the financial effect of karaoke room installation in hospitality venues in the UK by analyzing the financial data of 53 venues in the hospitality industry in the UK from 2018 to 2022. Contextually, the hospitality venues turn their vacant space into karaoke rooms that are open for customers' bookings. The venues continue operating as usual, except for the new karaoke rooms. To identify the potential causal impact (if any) of introducing karaoke rooms on the performance of hospitality venues in the UK, the author uses the venues' return on assets (ROA) as the metric, a profitability ratio that describes how much profit a company can generate from its assets.

1.4 Structure of the research

The rest of the thesis is structured as follows: In Section 2, the Difference-in-Differences (DiD) model's underlying mechanics are explained inside its canonical framework, and the advanced DiD estimator by Callaway and Sant'Anna (2021) is introduced. In section 3, the author studies the applicability of DiD in a business scenario by applying this estimator to examine the causal impact of karaoke installation in vacant spaces of UK hospitality venues on their financial performance. This section reveals a positive and statistically significant effect of the karaoke installation. In section 4, the limitation of this research caused by parallel trend pre-test bias and SUTVA violation will be discussed. Section 5 concludes the thesis by summarizing key insights derived from the empirical analysis.

2. DID METHODOLOGY REVIEW

To provide a solid foundation for future practitioners, the mechanism of DiD is explained under its well-known canonical model. Although the canonical DiD model is too simple and relies on various unrealistic assumptions that have low applicability in real-life non-experimental settings, understanding it would construct a necessary foundation for the future practitioners of more advanced estimators of DiD. The framework of the novel estimator suggested by Callaway and Sant'Anna (2021) will also be presented.

2.1 DiD mechanism: the canonical model

Let's first learn what Differences-in-Differences is. According to Goodman-Bacon (2019), Differences-in-Differences regression (DID) is used to assess the causal effect of a treatment by comparing the variations in outcome before and after the treatment happened (first difference) between the treatment group and the control group (second difference). It constructs these groups by observing their pre- and post-treatment behavior. As the canonical DiD framework, as well as the TWFE estimator, has been discussed comprehensively in a paper by Roth et al. (2023), hence the following sections will refer mostly to this paper.

2.1.1 Canonical DiD treatment assignment and timing:

The fundamental elements of DiD consist of two discrete time intervals and two cohorts. During the before-treatment period (t = 1), all units were not exposed to the treatment; during the after-treatment period (t = 2), they were divided into two groups: those who received the treatment, referred to as the treatment group; and those who did not, known as the comparison or control group. The individual i's treatment state is marked as follows:

- $D_i = 1$: individual i is treated between time t = 1 and t = 2
- $D_i = 0$: individual i is untreated in both periods.

In periods t = 1, 2, the practitioners collect panel data by observing the treatment state D_i and the outcome variable Y_i across individuals i = 1, 2,..., N from both groups. The following categories describe the possible outcomes of units:

- $Y_{i,t}(0)$: individual i's potential outcome at time t if untreated in both periods
- $Y_{i,t}(1)$: individual i's potential outcome at time t if treated in the second period

2.1.2 Canonical DiD assumptions

a) Unconditional parallel trend assumption

The unconditional parallel trend assumption is the key assumption for the DiD estimator, which intuitively indicates that without the treatment, the treated and untreated units' average outcome would have progressed similarly. Even in the event of selection bias, which permits non-random treatment assignment, this parallel trend can persist.

$$E\,[Y_{i,2}(0) - Y_{i,1}(0) \mid D_i = 1] = E\,[Y_{i,2}(0) - Y_{i,1}(0) \mid D_i = 0] \quad (1)$$

To exemplify, according to this assumption, the average ROA would have followed a similar trend for all cohorts of venues, regardless of whether they implemented karaoke or not.

However, this demands that the treatment be adopted with mean independence from confounding factors that affect the trend of the outcome. For example, in our karaoke scenario, one potential confounding factor is that the financial performance of venues may differ depending on how long they were established. Assessments get more challenging if the choice to add karaoke rooms coincides with the number of years that the venue was established, going against mean independence. Karaoke may be introduced by newer establishments trying to stand out or by older ones looking for a refresh. Hence, this complexity challenges attributing financial changes solely to karaoke, as venue evolution could also play a role in revenue shifts, complicating the assessment of karaoke's exclusive impact on revenue trends. The author will elaborate on the potential violation of the unconditional assumption and the corresponding solutions in the later section of the thesis.

b) Non-anticipatory effects assumption

The non-anticipatory effect assumption essentially states that the treatment group's outcome in the first period is unaffected by the group's treatment in the second. In other words, units do not behave in a way that implies they are aware of when they will be treated before it starts.

$Y_{i,1}(0) = Y_{i,1}(1)$ for all i with $D_i = 1$ (2)

In the karaoke scenario, this suggests that in years before karaoke installation, the ROA in venues that had karaoke was unaffected by the forthcoming karaoke installation.

c) Stable unit treatment value assumption (SUTVA)

SUTVA implies that the outcome of individual i is independent of the treatment effect of another individual k >< i. This makes it easier to rule out the effects of externalities caused by the treatment of other individuals. Additionally, SUTVA plays a crucial role in reducing the consequences of general equilibrium, which describes the intricate interdependencies that arise within a whole economic system when one component of the system changes.

d) Strong Overlap:

This implies that, based on the observed characteristics, the likelihood of being in the treatment group is consistently limited to a value that is not close to one. Additionally, the proportion of units receiving treatment is consistently limited to a value that is not close to zero. That is, for some $\epsilon > 0$, $P(D_i = 1|X_i) < 1 - \epsilon$, very likely, and $E[D] > 0$. This ensures a diverse distribution of units across treatment and control groups, essential for DID's assumptions to hold.

In our motivating example, if every venue consistently had a karaoke room installed (likelihood close to one), it would be challenging to establish a comparison with venues without the installation to accurately measure the effect. Conversely, if no venue ever opted for a karaoke room (likelihood close to zero), there wouldn't be a treated group to evaluate the impact of the installation. Hence, having a mix of hospitality venues with and without karaoke rooms allows for a more meaningful comparison, enabling Differences-in-Differences analysis to accurately estimate the causal effect of installing karaoke rooms on revenue.

2.1.3 Basic DID estimator:

The ultimate goal is to study the causal effect of the treatment on an individual by obtaining the average treatment effect on the treatment group in the second period:

$$ATT = E[Y_{i,2}(1) \mid T_i = 1] - E[Y_{i,2}(0) \mid D_i = 1] \quad (3)$$

In simple terms, this computes the expected difference in outcomes of a unit treated in period 2 and a unit that is not treated, both from the treatment group ($D_i = 1$).

However, it is impossible for practitioners to observe the untreated outcome of the treatment group in period 2, which is $Y_{i,2}(0)$ for $D_i = 1$, as only one potential outcome for each unit i is observable. Therefore, practitioners need to rely on the mentioned assumptions for the standard estimate to hold.

Under the unconditional parallel trend assumption, (Eq. 1), we obtain that:

$$E[Y_{i,2}(0) \mid D_i = 1] = E[Y_{i,1}(0) \mid D_i = 1] + E[Y_{i,2}(0) - Y_{i,1}(0) \mid D_i = 0]$$

Additionally, from non-anticipatory assumption, (Eq. 1), we obtain that:

$$E[Y_{i,1}(0) \mid D_i = 1] = E[Y_{i,1}(1) \mid D_i = 1]$$

Combine these two equations, it follows that:

$$E[Y_{i,2}(0) \mid D_i = 1] = E[Y_{i,1}(1) \mid D_i = 1] + E[Y_{i,2}(0) - Y_{i,1}(0) \mid D_i = 0]$$

Substitute this into the target estimator $ATT = E[Y_{i,2}(1) - Y_{i,2}(0) \mid D_i = 1]$ (Eq. 3) to get:

$$\begin{aligned} ATT &= E[Y_{i,2}(1) \mid D_i = 1] - (E[Y_{i,1}(1) \mid D_i = 1] + E[Y_{i,2}(0) - Y_{i,1}(0) \mid D_i = 0]) \\ &= E[Y_{i,2}(1) \mid D_i = 1] - E[Y_{i,1}(1) \mid D_i = 1] - (E[Y_{i,2}(0) - Y_{i,1}(0) \mid D_i = 0]) \\ &= \underbrace{E[Y_{i,2} - Y_{i,1} \mid D_i = 1]}_{\text{Change for treatment group}} - \underbrace{E[Y_{i,2} - Y_{i,1} \mid D_i = 0]}_{\text{Change for control group}} \quad (4) \end{aligned}$$

This intuitively means Difference - in – Difference.

2.1.4 Two-way Fixed effect (TWFE) estimation and inference

The above expression of ATT, Eq. 4, is interpreted as the population expectation of Difference-in-Difference. Replacing the expectations of outcomes with their sample mean was suggested by Roth et al. (2023) as a natural way to estimate the ATT.

$$\text{ATT} = \left(\bar{Y}_{t=2,D=1} - \bar{Y}_{t=1,D=1}\right) - \left(\bar{Y}_{t=2,D=0} - \bar{Y}_{t=1,D=0}\right),$$

where $\bar{Y}_{t=t',D=d'}$ is the sample mean of Y of treatment group d' in period t′.

In order to simplify the computation, the two-way fixed effects (TWFE) regression estimator is employed as an alternative approach rather than manually computing the sample mean.

$$Y_{i,t} = \alpha_i + \varphi_t + (1[t = 2] \cdot D_i)\beta + \epsilon_{i,t}, \quad (5)$$

which illustrates the regression of the outcome $Y_{i,t}$ on an interaction between a treatment status and a post-treatment indicator, as well as a time-fixed effect. It is intuitive to demonstrate that the coefficient $\hat{\beta}$ is equivalent to ATT in this canonical DiD setup. When all canonical framework assumptions are combined with the independent sampling assumption, the estimates of $\hat{\beta}$ from Eq. 5 yield consistent estimates for ATT.

2.1.5 Limitation of TWFE and recent advancement in DiD estimator

Many recent papers including Athey and Imbens (2018), Goodman-Bacon (2019), and Sun and Abraham (2020) pointed out that the underlying assumptions and set-ups for the standard TWFE method are ideal yet not plausible in non-experimental settings. Discussing the limitation of its applicability in real scenarios where these assumptions may not hold would create a profound understanding of the current corresponding advances in DiD for a more robust causal estimator. The author narrows down the scope to study only the three main strands of development in recent literature including a) conditional parallel trend; b) staggered treatment adoptions; and c) heterogeneity in treatment effects.

a) The relaxation of unconditional parallel trend

Callaway and Sant'Anna (2021) argue that failing to account for covariate-specific trends and assuming unconditional parallel trends can introduce errors in evaluating causal effects. One explanation is the concern that the remaining confounding factors are time-invariant.

In order to strengthen the reliability of the parallel trends assumption, it is recommended by Callaway and Sant'Anna (2021) to verify that it applies exclusively when considering confounders. When considering a comprehensive set of covariates X_i, it is possible to assume that the treatment assignment is almost random given X_i.

$$E\,[Y_{i,2}(0) - Y_{i,1}(0) \mid D_i = 1, X_i] = E[Y_{i,2}(0) - Y_{i,1}(0) \mid D_i = 0, X_i]$$

By focusing solely on parallel trends based on Xi, a set of observable factors before treatment, the robustness of the study can be enhanced (Roth et al. 2023).

b) The extension of TWFE to staggered treatment assignment settings

In many empirical DiD applications, more than two periods are involved, and the treatment assignment occurs at different times for each treated unit. For instance, certain UK hospitality venues adopted karaoke installations in 2019, while others did so in 2020, and the pattern continues. Numerous discussions have centered on relaxing constant treatment timing and the two-period structure.

Even in staggered settings, adherence to SUTVA, parallel trend, non-anticipatory, and strong overlap assumptions remains essential. Extending from the fundamental TWFE assumptions previously outlined, we can readily apply these principles to staggered settings. The author will provide further elaboration on these staggered-setting assumptions in the new estimator section.

However, ensuring the efficiency of the TWFE extension in staggered settings hinges on assuming homogeneous treatment effects. This assumption implies constancy of treatment effects across time and cohorts, denoted as $ATT_{i,t}$ (g) ≡ ATT. Otherwise, as shown by Sun and Abraham (2021), there will be potentially large biases caused by TWFE estimates. This happens because, for instance, past-treated units (venues that installed karaoke earlier) are now serving as comparison units for the newly treated units (venues that installed karaoke later). However, due to heterogeneous treatment effects, these past treated units might not be ideal comparisons as the treatment effects have likely changed over time.

c) Allowance for heterogeneous treatment effect

As presented by Bernheim et al. (2021), one of the concerns when trying to capture the causal effects is that the treatment effects commonly vary across units, also known as heterogeneity in treatment effects. Hence, while the TWFE staggered requires the homogeneity in treatment effect assumption, the DiD method practitioners researchers often hesitate to restrict treatment effect heterogeneity (Roth, 2023).

According to Callaway and Sant'Anna (2021), one way to identify the heterogeneity in causal effect across groups with different treatment adoption timing is to combine the impacts of group-time average treatments. To do this, a parameter is considered:

$$\theta_{sel}(\tilde{g}) = \frac{1}{\mathcal{T} - \tilde{g} + 1} \sum_{t=\tilde{g}}^{\mathcal{T}} ATT(\tilde{g}, t).$$

where $\theta_{sel}(\tilde{g})$ is the average treatment effect for all units in group $\tilde{g}$, throughout their post-treatment periods.

This parameter is even insightful when we want to learn about the variation in the weight of the treatment effects between the early and late-adoption groups.

2.2 Staggered Callaway Sant Anna estimator

As synthesized by Roth (2023), there are at least 10 different robust estimators for staggered treatment adoption and heterogeneous treatment effects. However, in this research, the author only focuses on one of the latest estimators presented by Callaway and Sant'Anna in 2021, or the CS estimator, as it allows for staggered adoption settings, heterogeneity in causal effects, as well as conditional parallel trend implementation. These are the limitations of the standard TWFE framework as a causal effect estimator discussed by de Chaisemartin and D'Haultfœuille (2020), and the CS estimator is claimed to overcome those issues.

2.2.1 Set up:

Treatment periods:

A binary treatment of interest can be applied to units in any of the T periods (t = 1,..., T) when t > 1. During the panel, a unit that has been treated remains in an absorbing condition, meaning that once the unit receives treatment, it will continue to be treated during the rest intervals. According to Callaway and Sant'Anna (2021), this is known irreversibility of treatment.

Treatment status

$D_{i,t}$: unit i's treatment status in period t

Treatment assignment

$G_i = \min\{t: D_{i,t} = 1\}$: the earliest period in which unit i got treated

$G_i = \infty$ if i is not treated at all throughout T periods.

G_{comp} = comparison group or control group

In particular, G_{comp} is allowed for two possibilities in Callaway and Sant'Anna (2021). In the first, all not-yet-treated units ($G_{comp} = \{g': g' > t\}$) are used, whereas only never-treated units ($G_{comp} = \{\infty\}$) are used in the second. According to Callaway and Sant'Anna (2021), it potentially makes sense to report ATT(g,t) for all pertinent (g,t) when there aren't many periods or treatment cohorts.

Potential outcomes framework:

Next, we construct the framework for the potential outcomes in staggered settings. In periods $t = 1, 2, ..., T$ the practitioners collect panel data by observing the treatment state D_i and the outcome variable Y_i across individuals $i = 1, 2, ..., N$ from both groups. The following categories describe the possible results of units:

$Y_{i,t}(0)$: individual i's potential outcome at time t if untreated in all periods

For $g = 1,2,...,T$

$Y_{i,t}(g)$: individual i's potential outcome at time t if first treated in the period g

2.2.2 Identifying assumptions for CS estimator:

a) Limited Non-anticipatory:

Let's define $\delta \geq 0$ as the anticipatory effect parameter. For all $t \in \{1,...,T\}$ and $g \in G$ such that $t < g - \delta$,

$$E[Y_t(g)|X; G_g = 1] = E[Y_t(0)|X; G_g = 1]$$

Essentially, this implies that if a unit is given treatment within a certain time period, their outcome is independent of the timing of their future treatment. In other words, units do not behave according to their knowledge of when they will be treated before the treatment begins.

b) Conditional parallel trend

Furthermore, specifications where the parallel trends assumption holds conditionally on covariates are also examined by Callaway and Sant'Anna (2021).

Assumption b1. Conditional Parallel Trends with "Never-Treated" Control Group

For $t \in \{2,...,T\}$ and each $g \in G$ satisfied that $t \geq g - \delta$,

$$E[Y_t(0) - Y_{t-1}(0) \mid X, G_g = 1] = E[Y_t(0) - Y_{t-1}(0) \mid X, G_i = \infty,]$$

Assumption b2. Conditional Parallel Trends with "Not-Yet-Treated" Control Group

For each $(s,t) \in \{2,...,T\} \times \{2,...,T\}$ and each $g \in G$ satisfied that $t \geq g - \delta$ and $t + \delta \leq s < g'$,

$$E[Y_t(0) - Y_{t-1}(0) \mid X; G_g = 1] = E[Y_t(0) - Y_{t-1}(0) \mid X; D_s = 0; G_g = 0]$$

Under the condition of covariates X, assumptions b1 and b2 expand the concept of parallel trends to encompass several time periods and treatment groups. When the untreated group is similar to the treated group and is a sizable group, Assumption b1 is preferred above Assumption b2. When a "never-treated" group is too small or unavailable, assumption b2 becomes advantageous since it allows for more comparison groups and maybe deeper inference. Nevertheless, selecting Assumption b2 over Assumption b1 has disadvantages, such as restricting pre-treatment trends, whereas Assumption b1 may be necessary when early-period economic conditions vary among groups. It is essential to comprehend these trade-offs to choose the best parallel trend assumption for a given investigation. These insights are suggested by Callaway and Sant'Anna (2021).

c) Overlap

This assumption is extended from the canonical setup overlap assumption discussed by Sant'Anna and Zhao (2020) to various treatment periods and multiple cohort settings. For each $g \in G$ and $t \in \{2,...,T\}$, there exist some $\epsilon > 0$ such that $P(G_g = 1) > \epsilon$ and $p_{g,t}(X) < 1 - \epsilon$. According to this, for all g and t, a significant proportion of the population obtains treatment during period g, and the generalized propensity score remains consistently below one.

2.2.3 Estimator by Callaway and Sant'Anna

Given the staggered setting of restricted anticipatory and the parallel trends assumptions, comparing the anticipated variations in the cohort g's outcome between periods $t - 1$ and t to that of a control group that has not yet received treatment at period t allows one to calculate the Average Treatment Effect on the Treated (ATT). To be more specific, the average treatment effect on each treated group at time t is:

$$ATT(g, t) = E[Y_{i,t}(g) - Y_{i,t}(\infty) \mid G_i = g'], \text{ for any } g > t$$

which can be viewed as the multi-period equivalent of the identification outcome in Eq.(4). Given that it applies to any comparison group $g' > t$, it also applies when we take the average over a collection of comparisons G_{comp} such that $g' > t$ for all $g' \in G_{comp}$:

$$ATT(g, t) = E[Y_{i,t} - Y_{i,g-1} \mid G_i = g] - E[Y_{i,t} - Y_{i,g-1} \mid G_i \in G_{comp}].$$

Next, by substituting sample analogs for expectations, we can estimate ATT(g, t).

$$\widehat{ATT}(g,t) = \frac{1}{N_g}\sum\nolimits_{i:G_i=g}\left[Y_{i,t} - Y_{i,g-1}\right] - \frac{1}{N_{Gcomp}}\sum\nolimits_{i:G_i \in G_{comp}}\left[Y_{i,t} - Y_{i,g-1}\right] \qquad (6)$$

This estimator allows for staggered adoption settings with variation in control group selection as well as heterogeneity in treatment effect.

3. EMPIRICAL APPLICATION

3.1 Methodology

In this section, the author aims to demonstrate the practical application of the DiD estimator by Callaway and Sant'Anna (2021) for future practitioners. To achieve this, the author shows clear steps on how to apply this method in a real business scenario to study the karaoke installation effects on hospitality venues.

3.2 Research approach:

This research is an empirical application of the estimator developed by Callaway and Sant'Anna (2021) to study the causal effect. This research aims to collect data on hospitality venues in the UK about their return on assets as unit outcomes, and years in corporate as another covariate. Data about their treatment timing, which is the first year they installed karaoke, is also collected. Within this framework, my analysis focuses on estimating the effect of karaoke rooms on the return on assets (ROA) for each venue, divided into two distinct parts when not conditioning or conditioning on the covariate.

3.3 Setups:

There are two main groups of research interest, the treatment group - who installed karaoke, and the control group - who never or not yet installed karaoke. For all units in both groups, the author gathers the information on their:

- $Y_{i,t}(g)$: Outcome in the measured year t, which is the ROA of venue i at the end of year t ($2018 \leq t \leq 2022$).
- $X_{i,t}$: Covariate value of venue i in year t, which is the number of years incorporation of venue i in year t.
- $D_{i,t}$: the binary value of whether the venue i had karaoke in year t

- G_i: The time period when the venue i first installed karaoke. If this value is $G_i = \infty$, the venue i did not have karaoke throughout all periods t.
- g_t: Group of venues that first installed in year t

In an absorbing condition, $T_{i,t} = 1$ for all t >= G_i. To exemplify, the venue that started to have karaoke in 2019 would have $G_i = 2019$ and belong. Similarly, a venue that did so in 2020 would have $G_i = 2020$, and a venue that had not put karaoke into use throughout the 5-year period of interest would get $G_i = \infty$

3.4 Data Preparation

Data description:

This analysis uses the panel data on hospitality venues in the UK from 2018 to 2022. The data consists of key research elements, including venue names, addresses, business types, business years incorporation, financial accounts, and karaoke installation year (if any) on 53 venues in the UK in 5 years, from 2018 to 2022, which add up to 265 observations.

Data sources:

The data was obtained from two main data sources. The first database is the Customer Relation Management system (CRM) of Singa Oy - a Finnish company specializing in digital karaoke rooms. This data source allows access to the identifying information of the venues that installed karaoke including their name, address, and the beginning year of karaoke service. The second database is an open-access database for business statistics provided by a UK data provider, Endole. This data source provides the hospitality businesses' names, addresses, business types, financial numbers on a yearly basis, and their years of incorporation.

Data matching:

To match the data from two different data sources together, the author utilizes the name, address, and business type of each venue shown on both sources. When these elements are matched between the two sources, the data of interest for that specific venue from both sources will be combined for the panel data.

Treatment group selection:

Any venues that first installed karaoke from 2019 to 2022 will be allocated to the treatment group. For the irreversibility of treatment condition, all venues that once had karaoke continued to have karaoke in the following periods, and no venue was treated at time t = 2018. Hence, due to the availability of data, there are 20 units in the treatment group.

Control group selection:

Two control group options exist. First, the control group will only include the Never Treated venues. These venues are selected using Endole Explorer, a data provider that generates UK business lists from filtered properties. It filters the business sectors as Public houses and bars, Licensed restaurants, and/or Other amusement and recreation activities. The property Accounts Last Made is restricted to be from 1 January 2023 and ongoing. The above procedure produces a list of hospitality-related UK companies that fall under the same categories as the treated venues.

The author next selects the never-treated group from this list, ensuring they did not install karaoke during any of the studied years and disclose sufficient ROA data in years of interest. In this never-treated group design, 32 control units are adopted. The second version of the control group, namely not-yet treated group, includes both never-treated venues and those that will not receive treatment until a designated year. Table 1 shows the number of units in the treatment group and control group, both not-yet and never treated in each year from 2018 to 2022. We can observe that depending on how many sites are not yet treated in a given year, the not-yet comparison group changes annually. The units in the not-yet-treated group is the combination of the will-be-treated and never-treated units.

Table 1. Distribution of Units Across Groups by Year

Year	Treated Units	Will be treated	Never Treated	Not-yet Treated
2018	0	20	33	53
2019	8	13	33	46
2020	3	11	33	44
2021	3	6	33	39
2022	6	0	33	33

Tools used:

The data cleaning process could be done through different data composition tools, and this part will focus only on the analysis process that utilizes STATA for method application. The code to implement the method proposed by Callaway and Sant'Anna is available in the STATA csdid package which can be install directly on STATA for usage. For the sake of transparency and reproducibility, the implementation of the research code and data can be accessed here (github link).

3.5 Analysis

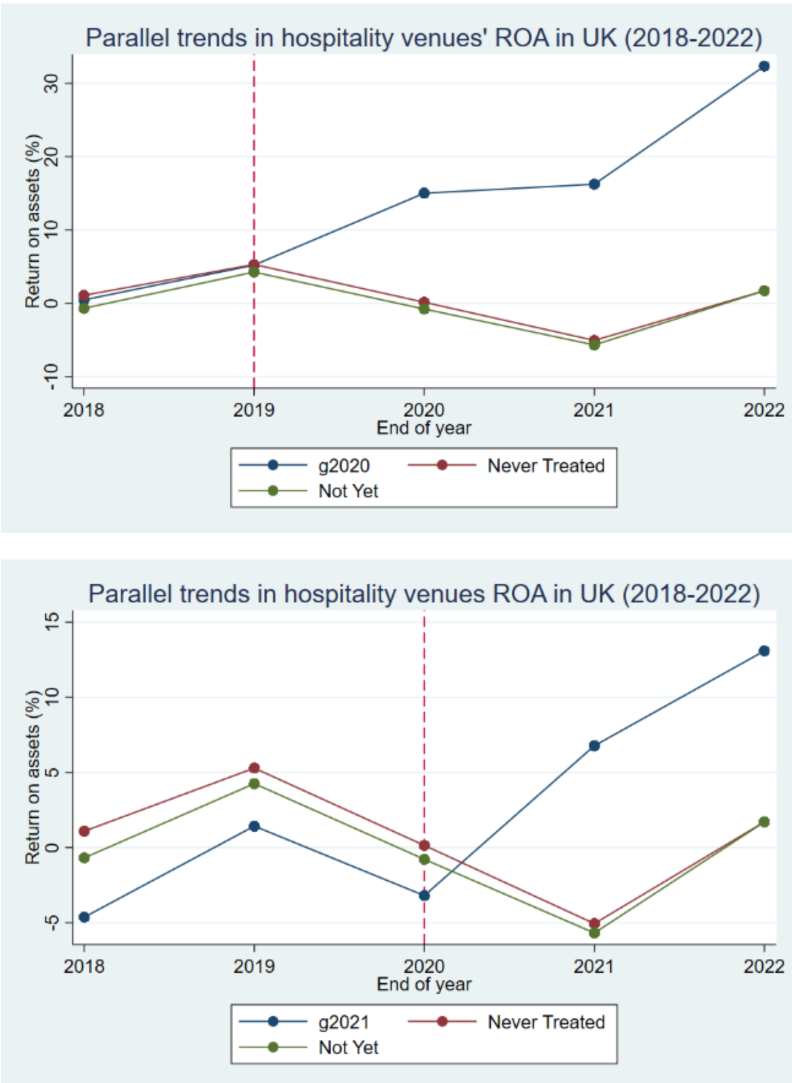
3.5.1 Assumptions justifying

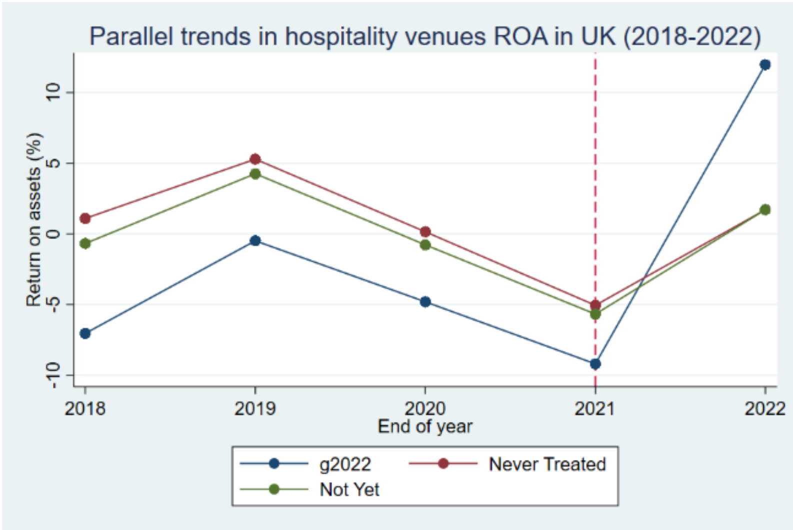
Before actually utilizing this method, it is necessary to ensure that the identifying assumptions hold. This research provides the justification of the feasibility of the two main identifying assumptions, parallel trend assumptions and non-anticipatory assumptions.

a) Parallel trend testing

Even when we are conditioning the parallel trend on covariate, it is still essential to condut the pre-test teatment. One common way to test for the parallel pre-treatment trend is to visualize the trends. We visualize the ROA of the treated group and compare it to that of the never-treated and the not-yet-treated group to see whether they share a common trend before the treatment is introduced to that cohort.

Figure 2. Parallel trends in hospitality venues ROA in the UK of treated cohorts to Never and Not-yet comparison group (2018-2019)





It can be observed from Figure 2 that the parallel trends hold for the pre-treatment period(s) of each cohort that first had karaoke installation from 2020 to 2022. Although this parallel trend testing is essential, there are limitations of this method in this case. That is the pre-trend of venues that installed karaoke in 2019 cannot be observed due to data missing problems. Furthermore, for evaluating the impact on business financial performances, the distribution of observed covariates such as venues' years of incorporation is often quite different between venues that install karaoke and those that do not. When the path of ROA growth (in the absence of karaoke rooms) depends on this covariate, a conditional parallel trend becomes more feasible compared to the unconditional parallel trend assumption. Ignoring the covariate-specific trends can lead to significant biases when assessing the causal inference of treatment employing unconditional trend approaches. Therefore, the author also extends this assumption to the conditional parallel trend assumption taking the venues' years of incorporation as covariate.

b) Limited anticipation assumption justification:

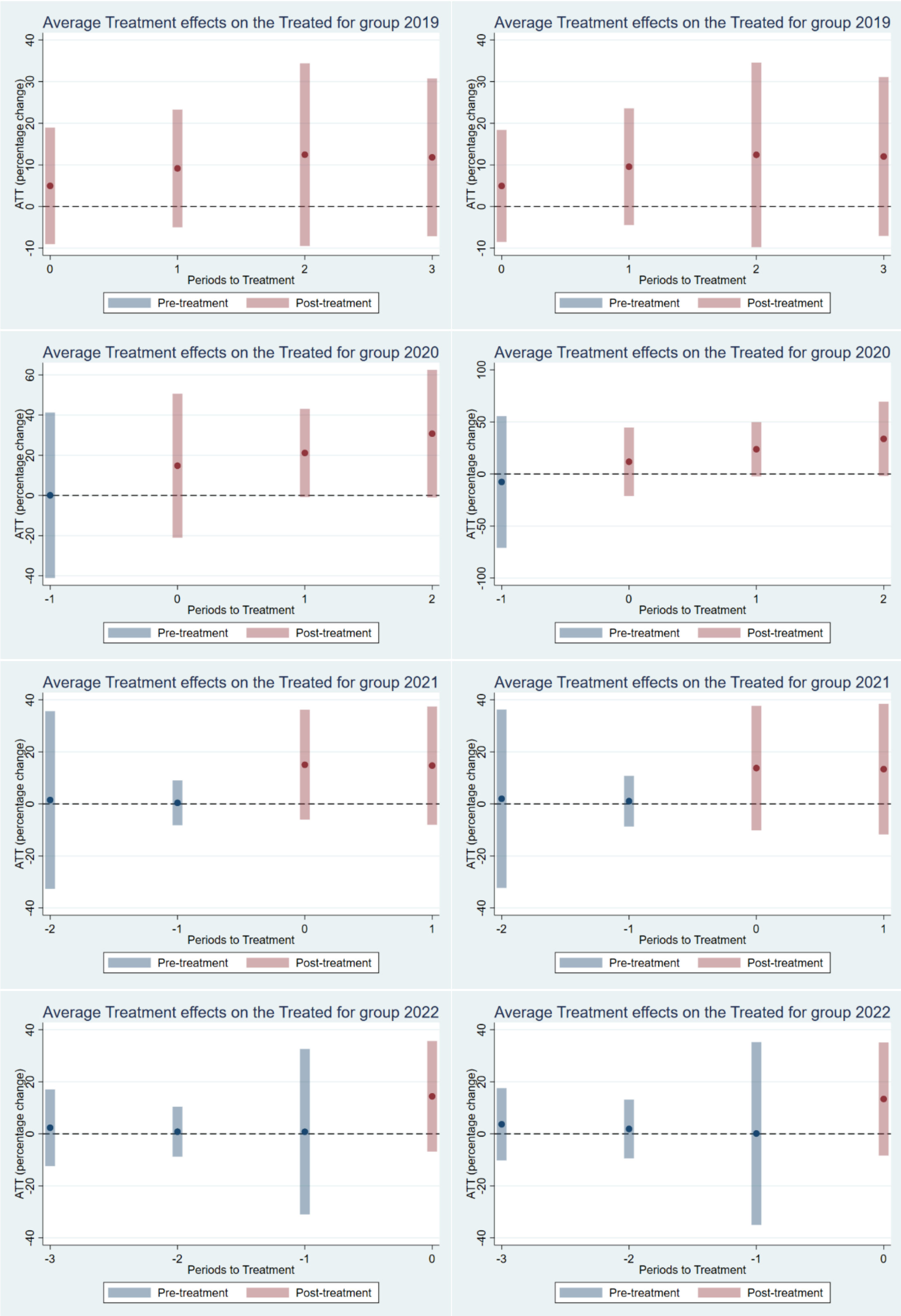
All studied venues are assumed to be profit-maximizing entities which means in all time periods the studied venues would always try to enhance their return on assets regardless of their knowledge of future installation of karaoke. In the scenario being used here, this suggests that in years before karaoke installation, the ROA in venues that installed karaoke was unaffected by the forthcoming karaoke installation.

From Figure 3, except for group 2019, we can observe that all the group-time average treatment impacts of each cohort in their before-treatment intervals are comparatively close to zero. This indicates that there are no to few anticipated effects before the treatment occured proving the limited anticipatory assumption holds for this study.

Figure 3. Karaoke Group-Time Average Treatment Effects

a) Unconditional parallel trends

b) Conditional Parallel Trends



Notes: The impact of the karaoke installation on ROA evaluated using the unconditional parallel trends assumption and the conditional parallel trends assumption is illustrated in Panel (a) and Panel (b), respectively. The estimates in Panel (b) account for the conditional parallel trend on a covariate, specifically the years incorporation, estimator mentioned in the text. The blue bars provide point estimates and consistent 95% confidence intervals for the pre-treatment periods. If the null hypothesis of the parallel trends assumption is valid for all periods, these values should be exactly 0. The red lines represent the estimated treatment effect of installing karaoke at the venue level. They also show the 95% confidence bands, which indicate the range within which we are confident the true treatment effect lies. Each row of venues is arranged chronologically from 2019 to 2022, indicating the year in which karaoke was introduced at each venue.

3.5.2 Different identifications of method usage

Four regressions are run with different identifications of the method to see how well each identification can apply to estimate the causal effect in such a real business scenario setting. The author specifically examines two scenarios: one where the parallel trends assumption holds in every circumstance, and another where it holds solely when accounting for observed features X, or the venues' years of incorporation. Both scenarios will consider two versions of the comparison group: the first version will consist of venues that have never been treated, where no anticipated effects are allowed (i.e., $\delta = 0$); the second version will include venues that have not yet been treated and allow for one-year anticipation.

3.5.3 Findings

In the following, the author analyzes various sets of outcomes by employing alternative assumptions regarding parallel trend identification and selection of control groups. The standard error and p-value of results generated by different specifications will be compared to each other in order to find the most robust estimator in this specific business setting.

To understand the treatment effect closely, the author goes through different estimates of Average treatment effect on all cohorts and each cohort's aggregated treatment effect using Table 4.

Table 4. Karaoke Installation Aggregated Treatment Effect Estimates

	a) Unconditional Parallel		b) Conditional Parallel	
	Never	Not Yet	Never	Not Yet
g2019	10.012	9.59	10.635	9.733
	(8.38)	(8.10)	(8.29)	(8.01)
g2020	22.35	22.21	22.968	23.10
	(14.9)	(14.95)	(15.73)	(15.67)
g2021	14.95	14.88	13.87	13.56
	(11.15)	(11.00)	(12.29)	(12.3)
g2022	14.41*	14.41*	13.38	13.38
	(10.84)	(10.84)	(11.09)	(11.09)
ATT	12.61*	12.32	12.83	12.23*
	(6.2)	(6.19)	(6.188)	(6.186)

* p<0.05, ** p<0.01, *** p<0.001

gt: Group of venues that first installed karaoke in year t

Note: P-values, which quantify statistical significance, indicate the rejection of the zero-effect hypothesis. Assigned stars in tables are as follows by Imbens (2021): one represents 10%, two 5%, and three 1%. P-values with smaller values signify greater significance, whereas stars serve as a convenient indicator of the level of confidence in the results of a study.

From Table 4, let's first focus on the ATT values under different parallel assumptions and comparison group selections. Under the same parallel trend assumption, the not-yet comparison group design generates results at lower standard error compared to the never-treated comparison group design, indicating more robust estimates. This is due to the small size of the "never-treated" group, as mentioned in section 2.2.2., which makes the not-yet comparison group design more useful because it enables the use of additional comparison groups and possibly deeper inference. In other terms, under the same parallel trend assumption, the same-year cohort aggregated treatment effect using not-yet control group design is more robust than that of never control group design.

Furthermore, for the same control group set-up, the ATTs estimated under the b) panel of Table 4 have lower standard error compared to the ATTs under the a) panel, which indicates that when assuming parallel trend conditioning on the number of years of incorporation, the results have lower standard error than the unconditional parallel trend.

From these insights, it can be inferred that under this research setting, where the unconditional parallel trend assumption is unlikely to hold when covariate-specific trends exist in results throughout periods, the conditional parallel trend can give out more robust results. Additionally, when the same parallel trend assumptions are made, the not-yet comparison group design brings about more robust results compared to the never-treated comparison design. This is because the never-treated group size of this research is relatively small, and using the not-yet comparison group design allows for more groups to serve as reliable control units, which potentially results in more informative causal effects.

According to Table 1, we know that the units for the never-treated control group as well as the not-yet-treated control group in the year 2022 are similar, hence, from Table 4, we can see that the generated estimates for the two control groups are the same for the group treated in 2022. Based on the aggregated treatment effect by cohort shown in Table 4, the author analyses how the effect of employing karaoke changes by the duration that the installation has been conducted. Except for the group 2019 that is unable to be parallel-trend tested due to data insufficiency, the magnitude of the karaoke installation effect on venues' ROA appears to be increasing the longer venues are exposed to the karaoke treatment, under conditional parallel trend and not-yet comparison group design. In particular, in the year 2020 when a venue started to have karaoke, the aggregated treatment effect leads to a 23.10 increase in percentage points with a high statistical significance level. This number is higher than that of group 2021, which is a 13.56 percentage points increment. Similarly, the aggregated treatment effect of units that were exposed to the treatment in 2021 had a higher magnitude compared to that of group 2022 with a 13.38 increase in percentage points. Hence, the treatment effect is inconsistent between different cohorts and we can obtain an upward trend in treatment effect magnitude according to units's treatment duration after 2019. In other words, the treatment effect is heterogeneous across groups.

Most significantly, in all estimates made, the one used under the not-yet comparison group and the conditional parallel assumption generated the average treatment effect on the treated (ATT) of 12.23 change in percentage points, with the lowest standard error level and statistically significant result (p-value less than 0.05). All in all, the results show that the group-time average treatment effects estimation is positive strengthening the hypothesis that having karaoke installation led to the increment in the venue's ROA.

4. LIMITATION

In general, the findings indicate that the installation of karaoke at a venue enhances its ROA compared to what it would have been without the installation. Nevertheless, this application does have significant limitations. These limitations make it difficult to get a thorough grasp of the causal relationships, which forces one to critically evaluate the analysis's presumptions and potential biases.

According to Roth et al. (2023), previous research has questioned the validity of pre-treatment parallel trends, particularly in light of the possibility of parallel trend violations. A pre-test bias can be caused by relying just on pre-trend testing to guide analysis, which means that the results of this preliminary test could skew or selectively interpret the analysis outcome. Furthermore, conditioning on a single covariate to guarantee parallel trends could fail to yield the required parallel trends assumption because several time-varying unobserved confounding factors, rather than a single factor, may cause violations.

Another critical limitation arises from the potential violation of the Stable Unit Treatment Value Assumption (SUTVA). This violation raises concerns about spillover effects, where the installation of karaoke in one venue may impact neighboring venues through changes in customer rates, for instance. This violation introduces complexities in accurately assessing the true causal effects of karaoke installations, highlighting potential biases and complexities in traditional testing methodologies.

5. CONCLUSION

In conclusion, the estimator proposed by Callaway and Sant'Anna allowing for conditional parallel trend assumption and the not-yet control group can, all else being equal, possibly result in a more robust estimation of the true causal effect under the case of potentially violated unconditional parallel trend assumption, small never-treated comparison group and heterogeneous treatment effect.

To show future practitioners a clear method application guidance, the author applied this approach to analyze the effects of karaoke installations on hospitality venues headquartered in the UK. The research found evidence that having karaoke increases the venue ROA. The most robust estimate of the average treatment effect on the treated shows that treated venues have

12.23 percentage points higher ROA compared to the without-treatment scenario. Furthermore, with the exception of the untestable pre-treatment cohort (g2019), the data indicate a positive association between the duration of treatment exposure and the magnitude of the treatment effect.

However, the analysis highlights some shortcomings, such as incomplete data, possible spillover effects, and unresolved problems with parallel trend tests. These restrictions highlight the necessity of interpreting the data with caution. The study demonstrates that the selection of estimator can lead to different conclusions, as it considers factors such as variations in treatment effects, trends specific to certain variables, and the adoption of treatment at different times, while remaining the main assumptions constant.

6. FUTURE RESEARCH

Firstly, addressing violations in parallel trends offers promising areas for exploration. Conditioning on more covariates could limit the violation of the necessary parallel trends assumption. Recent studies propose modified pre-trend testing strategies and alternative approaches that accommodate post-treatment violations, enabling increased effectiveness in handling relevant deviations from parallel trends.

Secondly, the exploration of the Stable Unit Treatment Value Assumption (SUTVA) holds significant potential. Current work by Butts (2021) presents initial frameworks for accommodating local spatial spillovers within the DiD framework by extending the estimator developed by Callaway and Sant'Anna. Anticipating future endeavors, an increasing focus on DiD methodologies incorporating spillover effects is foreseeable.

Lastly, an exciting frontier involves extending the framework for assessing the causal effects of karaoke rooms to encompass richer panel data on a larger scale in terms of industry and geography. This expansion would allow for a more comprehensive understanding of the multifaceted impacts of karaoke installations on hospitality venues across diverse contexts and timeframes.

These potential areas for future research present compelling opportunities to enhance the applicability, scope, and precision of DiD methodologies in empirical studies.

Reference

Athey, S. and Imbens, G.W. (2022), “Design-based analysis in difference-in-differences settings with staggered adoption,” Journal of Econometrics, 226(1), 62–79.

Baker, A., Larcker, D.F. and Wang, C.C. (2021), “How much should we trust staggered difference-in-differences estimates?,” SSRN Electronic Journal.

Butts, K. (2023) ‘Difference-in-differences estimation with spatial spillovers’, arXiv.org.

Callaway, B. and Sant’Anna, P.H.C. (2021), “Difference-in-differences with multiple time periods,” Journal of Econometrics, 225(2), 200–230.

de Chaisemartin, C. and D’Haultfœuille, X. (2020), “Two-way fixed effects estimators with heterogeneous treatment effects,” American Economic Review, 110(9), 2964–2996.

Goodman-Bacon, A. (2020), “Difference-in-Differences with Variation in Treatment Timing,” NBER Working Paper.

Huber, M. and Steinmayr, A. (2019), “A framework for separating individual-level treatment effects from spillover effects,” Journal of Business & Economic Statistics, 39(2), 422–436.

Imbens, G.W. (2021), “Statistical significance, p-values, and the reporting of uncertainty,” Journal of Economic Perspectives, 35(3), 157–174.

Roth, J., Sant’Anna, P. H. C., Bilinski, A., & Poe, J. (2023), “What’s trending in difference-in-differences? A synthesis of the recent econometrics literature,” Journal of Econometrics, 235(2), 2218-2244.

Sun, L., Abraham, S. (2021), “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects,” Journal of Econometrics, 225 (2), 175–199.

Endole. Available at: https://www.endole.co.uk/ (Accessed: 25 November 2023).