A Simple Approach to Staggered Difference-in-Differences in the Presence of Spillovers

Mario Fiorini * Wooyong Lee † Gregor Pfeifer ‡

June 13, 2024

Abstract

We establish identifying assumptions and estimation procedures for the ATT in a Difference-in-Differences setting with staggered treatment adoption in the presence of spillovers. We show that the ATT can be estimated by a simple TWFE regression that extends the approach of Wooldridge [2022]'s fully interacted regression model. We broaden our framework to the non-linear case of count data, offering estimation of the ATT by a simple TWFE Poisson regression, and we revisit a corresponding application from the crime literature. Monte Carlo simulations show that our estimator performs competitively.

Keywords: Difference-in-Differences, staggered treatment adoption, spillovers, (non-)linear models.

*(corresponding author): University of Technology Sydney (mario.fiorini@uts.edu.au).
†University of Technology Sydney (wooyong.lee.econ@gmail.com)
‡University of Sydney, CESifo, and IZA (gregor.pfeifer@sydney.edu.au).

1 Introduction

The Difference-in-Differences (DiD) literature, particularly the one concerned with staggered treatment adoption, has experienced significant advances in the last few years, and papers by Roth et al. [2023] and de Chaisemartin and D'Haultfœuille [2021] have summarized these developments. Within this array of advances, one area still understudied is the one linked to spillovers—implying that the Stable Unit Treatment Value Assumption (SUTVA) assumption does not hold. However, as Roth et al. [2023] point out, spillover effects may be important in many economic applications, such as when a policy in one area affects neighboring areas, or when individuals are connected in a network.[1] Our work contributes to this area and links two active DiD literature strands.

The first focuses on estimation issues under staggered adoption and heterogeneous treatment effects across units and time. Borusyak et al. [2024], de Chaisemartin and D'Haultfoeuille [2020], Callaway and Sant'Anna [2020], Goodman-Bacon [2021], Sun and Abraham [2020], and Wooldridge [2022] highlight that the two-way fixed effect (TWFE) regression estimator may be biased for the average treatment effect on the treated (ATT), to the extreme of showing the opposite sign. The authors suggest alternative estimators that account for the variation in treatment timing, thereby providing a consistent estimator for the ATT. We contribute to this literature by extending it to the case of spillovers in both linear and non-linear models.

The second strand studies the identification of average treatment effects in the presence of spillovers. Berg et al. [2021], Butts [2023], Clarke [2017], and Huber and Steinmayr [2021] highlight two main challenges for identification of the ATT if the treatment also impacts units that are not formally treated.[2] First, untreated units are no longer valid controls. So far, proposed solutions mostly centre around ruling out spillovers for a given group of units, often based on some spatial or social network distance, allowing the researcher to use this latter group as a control. This structure is sometimes defined as partial interference. Alternatively, if sufficient information exists, one can parametrize how units are exposed to spillovers. Second, multiple definitions of the ATT are possible in the presence of spillovers. This is because a unit's treatment can lead to changes to its own outcome, but also to other units' outcomes. In this case, the researcher might be interested in examining the former effect, summarized by the ATT without interference (i.e., the one identified under SUTVA), or in a broader definition of the ATT that also accounts for the latter effect. Here, we contribute to this literature by providing conditions

[1] As another example, Minton and Mulligan [2024] use price theory to demonstrate that when treated and control units are in the same market, control units are indirectly affected by the treatment.

[2] Less close, but still related, is the literature that studies the role of spillovers in randomized controlled trials, see Sävje et al. [2021], Vazquez-Bare [2023], Han et al. [2024].

that allow for the identification of the ATT without interference, despite the presence of spillovers. Our setting also departs from this literature since we focus on the more complex staggered treatment adoption, which has the potential for cumulative spillovers. Nevertheless, our results also apply to the simpler simultaneous treatment case.

Specifically regarding contributions, we first establish the identifying assumptions for the ATT without interference given a staggered DiD setting in the presence of spillovers. We show that aside from the canonical i) treatment irreversibility, ii) no-anticipation, and iii) parallel trends assumption, identification requires that once a unit receives treatment, it is no longer influenced by spillover effects. This means the unit forfeits any spillovers it may have previously received and remains unaffected by spillovers from subsequently treated groups. This assumption also unifies the multiple definitions of the ATT, because they are the same with or without spillovers, simplifying policy evaluation and joining with the definition of ATT under SUTVA. We also assume that a set of never-treated units is not exposed to spillovers, in line with the existing literature. The combination of these assumptions allows for the identification of the ATT. Below, we argue that such a scenario applies to many contexts. Differently from Butts [2023], who is closest to our work, we directly focus on the staggered treatment scenario and, importantly, provide assumptions for the identification of the ATT without interference.[3]

Our second contribution regards estimation. We show that the extended TWFE model approach of Wooldridge [2022], which is numerically equivalent to the imputation-based approach of Borusyak et al. [2024], can be used to account for spillovers. Furthermore, we discuss identification and estimation in the non-linear case of count data, broadening the range of applications to which our approach can be applied, since the parallel trends assumption is sensitive to the functional form (Roth and Sant'Anna [2023]). For our empirical application, we revisit Gonzalez-Navarro [2013], who studied the effects of installing a stolen vehicle recovery device on car theft. Since car theft is a count variable, we implement the non-linear Poisson DiD adjusted for spillovers. Our correction leads to a larger effect of the policy relative to the original contribution's specification.

Finally, we perform a Monte Carlo analysis, highlighting the bias-variance trade-off implicit in the correction for staggered treatment and spillovers. Identification of time and group fixed effects can neither rely on the already treated units due to heterogeneous treatment effects, nor on the untreated units potentially exposed to spillovers. However, the benefit of excluding such units from estimation can be small if treatment effects are relatively homogeneous and if spillovers are small, while costing the researcher precision. We compare the traditional TWFE estimator, which ignores both staggered adoption and spillovers, the Wooldridge [2022] estimator, which accounts for staggered adoption but

[3] Butts [2023] is concerned with establishing identification of the sum of direct and spillover effects.

not for spillovers, and our estimator, which corrects for both. We do so under different sample sizes, degrees of staggered treatment, and degrees of spillovers, showing that our estimator performs competitively in many settings.

The remainder of the paper is organized as follows. Section 2 provides intuition alongside two motivating examples, after which Section 3 lays out the formal DiD setup with staggered treatment adoption. Section 4 establishes conditions for identifying the ATT, while Section 5 discusses estimation and inference considering the formerly established assumptions. Section 6 extends our model to the non-linear case, and Section 7 discusses a corresponding application. Section 8 provides Monte Carlo simulations, and Section 9 concludes.

2 Intuition and Motivating Examples

To illustrate our setting, consider panel data of units divided into three groups, A, B, and Z, observed over three periods, 1, 2, and 3. The timing of their respective treatment distinguishes these groups. Group A is treated in periods 2 and 3, group B in period 3, and group Z remains untreated. Each group consists of two units, denoted by a, a', b, b', z, z'. We use the indices i and j to refer to any unit within these groups. Figure 1 illustrates the potential treatment and spillover mechanisms in this setting, focusing on periods 2 and 3. Solid lines indicate treatment effects under no interference (β_{it}), equivalent to the conventional definition of the treatment effect under SUTVA. We call them direct effects. Dotted lines indicate spillovers from a treated unit j to a treated unit i (γ_{it}^j), and dashed ones represent spillovers from a treated unit j to an untreated unit i (η_{it}^j). The figure highlights a key challenge introduced by the presence of spillovers: there are no valid control units. Under SUTVA, only the direct effects represented by solid lines would exist.

We can visualize possible data patterns in this setting using a simple parametric model. Suppose that the outcome of interest is deterministic and given by:

$$Y_{it} = 1 + \delta_t + \beta_{it} \cdot D_{it} + D_{it} \cdot \sum_{j \neq i} \gamma_{it}^j \cdot D_{jt} + (1 - D_{it}) \cdot \sum_{j \neq i} \eta_{it}^j \cdot D_{jt}, \tag{1}$$

where D_{it} is a binary variable equal to 1 when unit i is treated. Equation (1) illustrates a scenario where unit i's outcome is not only influenced by its own treatment, represented by β_{it}, but also by the treatments of other units, as captured by γ_{it}^j and η_{it}^j. Assuming that the treatment effect in the absence of interference is homogeneous across units and time, we set $\beta_{it} = -0.5$ for all i and t, in which case the ATT without interference equals to -0.5. We also assume that the time effect is given by $\delta_t = 0.1 \cdot (t - 1)$.

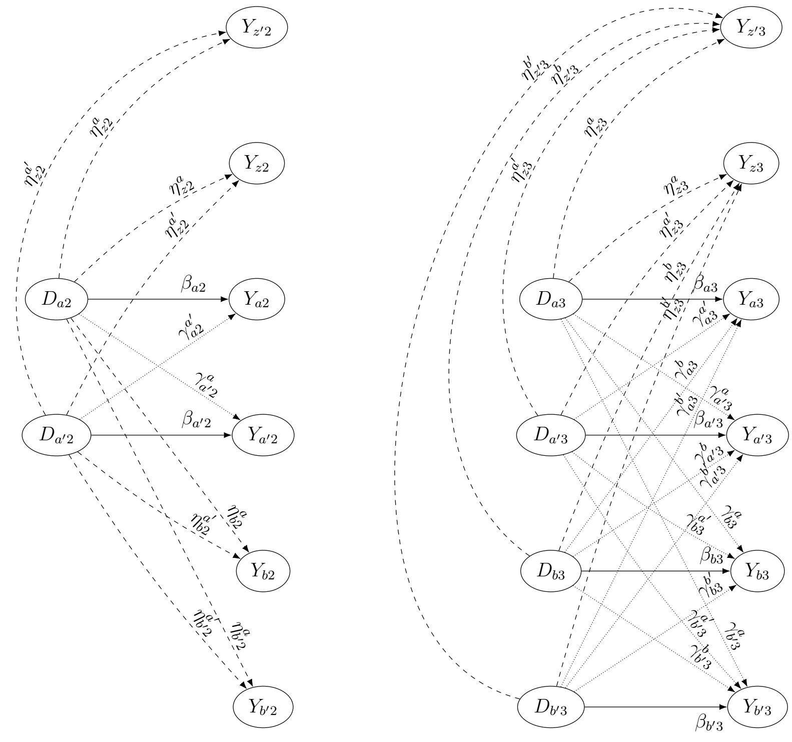


Figure 1: Illustration of the potential treatment and spillover paths

The left panel in Figure 2 illustrates a data pattern in scenarios without a spillover effect. Estimators that account for staggered adoption and heterogeneous treatment effects typically utilize the observations from the never-treated group Z and the not-yet-treated observations in group B at time 2 as the control group. These observations are used to estimate the time trend and are contrasted with treated observations to estimate the ATT without interference. For example, Wooldridge [2022] proposes the following variant of the TWFE regression model. Let G_i be the group membership of unit i, with $G_a = G_{a'} = A$, $G_b = G_{b'} = B$, and $G_z = G_{z'} = Z$. The model is given by:

$$
\begin{aligned}
Y_{it} = \alpha_i + \delta_t + \beta_{A2} \cdot \mathbf{1}(G_i = A, t = 2) \\
+ \beta_{A3} \cdot \mathbf{1}(G_i = A, t = 3) \\
+ \beta_{B3} \cdot \mathbf{1}(G_i = B, t = 3) + \varepsilon_{it},
\end{aligned}
\tag{2}
$$

where, abusing notation, $(\beta_{A2}, \beta_{A3}, \beta_{B3})$ are coefficients on group-period indicators, and α_i, δ_t are the unit and time fixed effects, respectively. Under suitable conditions, it can be shown that the estimate of β_{gt}, denoted by $\widehat{\beta}_{gt}$, is consistent for the ATT for each

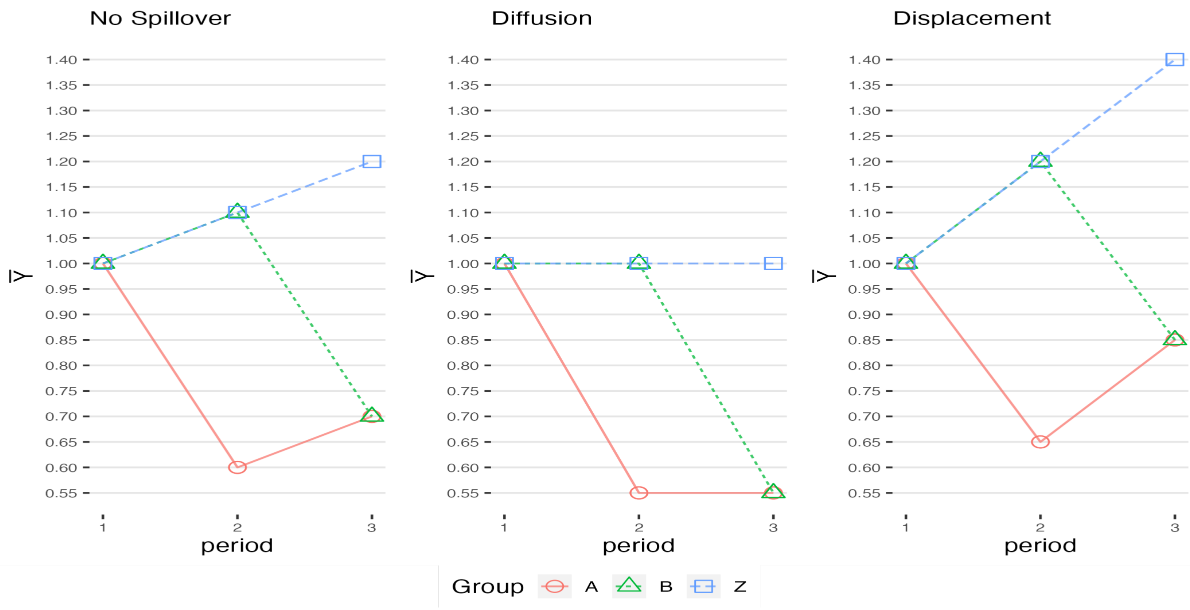


Figure 2: Possible data patterns under Equation (1)

group g at each time t, all equal to -0.5 in our example.[4]

We now introduce spillovers under two alternative scenarios. We will also use these so-called Examples 1 and 2 throughout the paper to motivate our assumptions and empirical application. In the first scenario, the spillover is in the form of a diffusion effect, meaning that the direct effect β_{it} and the spillover effect $(\gamma_{it}^{j}, \eta_{it}^{j})$ have the same sign. In the second scenario, the spillover is in the form of displacement, where β_{it} and $(\gamma_{it}^{j}, \eta_{it}^{j})$ have opposite signs.

Example 1 (installation of a water treatment plant)**.** Consider a scenario where we are interested in the effect of introducing a water treatment plant on the health outcomes of villages situated along a river. Suppose the nearby villages a and a', categorized as group A, are the first to adopt the plant. This adoption not only improves water quality in these villages but may also enhance the water quality of the not-yet-treated downstream villages, resulting in a spillover effect.

The middle panel in Figure 2 visualizes a possible data pattern of Example 1. For a numerical illustration, let's build upon Equation (1) and set the spillovers to also be homogeneous across units and time: $\gamma_{it}^{j} = \eta_{it}^{j} = \gamma = \eta = -0.05$ for all i, j and t. There are two key challenges arising from this setting. First, there is no valid control group because both never-treated and not-yet-treated observations are negatively affected by spillovers. To illustrate, note that the time-difference of Equation (1) for an untreated

[4]In fact, since the treatment is homogeneous, the standard TWFE regression would also consistently estimate the ATT.

unit is given by:

$$Y_{it} - Y_{i,t-1} = \delta_t - \delta_{t-1} + \sum_{j \neq i} \left(\gamma_{it}^{j} \cdot D_{jt} - \gamma_{i,t-1}^{j} \cdot D_{j,t-1} \right).$$

This equation reveals that the spillover effect introduces a bias term, disrupting the consistent estimation of the time trend $(\delta_t - \delta_{t-1})$. For instance, the bias term for the time-difference of an untreated unit z between periods 2 and 3 is calculated as $\gamma_{z3}^{a} + \gamma_{z3}^{a'} + \gamma_{z3}^{b} + \gamma_{z3}^{b'} - \gamma_{z2}^{a} - \gamma_{z2}^{a'} = -0.1$. Unit z' has the same bias term.

Second, even if we could correctly identify time and group fixed effects, it would not be possible to separately identify the direct and spillover effects. Estimators that account for staggered adoption and heterogeneous treatment effects, such as (2), would at best identify the average sum of the direct and spillover effects. For example, such an approach would estimate $\widehat{\beta}_{A2} = (\beta + \gamma) = -0.55$ and $\widehat{\beta}_{A3} = \widehat{\beta}_{B3} = (\beta + 3 \times \gamma) = -0.65$.

Example 2 (installation of stolen vehicle recovery devices)**.** Gonzalez-Navarro [2013] studied the effect of installing a stolen vehicle recovery device on car theft incidents. The introduction of this treatment was staggered across different states within a country and was limited to specific car models. In this scenario, car theft could potentially be displaced to other unprotected models within treated states or to the same models in states that had not yet adopted the device. Gonzalez-Navarro [2013] found a 52% increase in thefts for the same models in states without the installed device.

The right panel in Figure 2 visualizes a possible data pattern of Example 2, where we set $\gamma_{it}^{j} = \eta_{it}^{j} = 0.05$ for all i, j and t. Example 2 face the same key challenges we discussed: the absence of a valid control group and the difficulty in separately estimating direct and spillover effects. Note that, especially in the case of displacement, spillover effects could intensify over time as more and more treated units spill on an increasingly narrower pool of untreated units.

While the sum of direct and spillover effects might be of interest in some cases, identifying the direct effect separately should be of prime importance in most contexts. For example, when a unit decides whether to participate in a policy or treatment, its main concern probably is the direct effect, because other units' decisions are out of its control. Policymakers whose jurisdiction spans all units might also want to understand the distinct impact of each channel. In addition, it should be noted that the sum of direct and spillover effects might have limited external validity, as this sum is specific to the observed treatment histories of all units, while there is a vast array of counterfactual treatment histories that all of these units might experience.

Figure 3 visualizes our key assumptions that allow the identification of the direct effect. They assume that units are not influenced by spillovers once treated and that

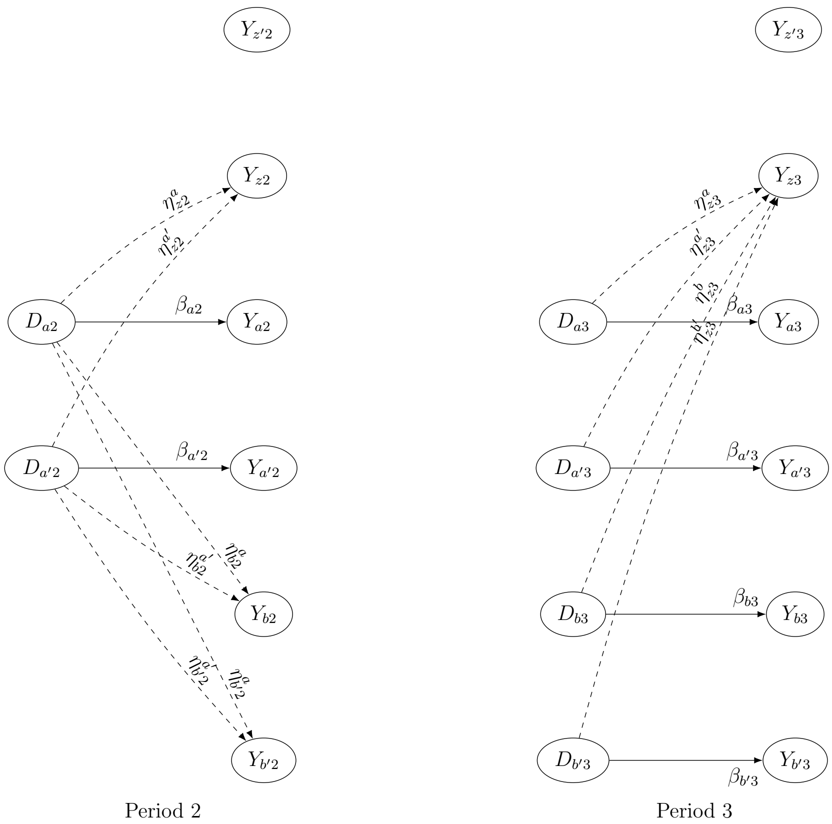


Figure 3: DAG under the key identification assumptions

a subset of never-treated units remains unaffected by spillovers as well. Consequently, in this figure, there are no longer lines to treated observations and no lines to unit z', allowing for the identification of the time trend. In what follows, we detail how these assumptions are likely to hold in empirical applications, illustrated through Examples 1 and 2.

Example 1 [continued]. Consider villages situated at the most upstream part of a river, none of which have water treatment plants. These upstream villages are not affected by the installation of water treatment plants in other villages along the river since all other villages are downstream relative to them. Therefore, in this context, these upstream villages represent untreated units that are not subject to spillover effects.

Next, consider the village located furthest downstream, which initially does not have a water treatment plant. When an upstream village installs a plant, the downstream village experiences spillover effects, benefiting from improved water quality resulting from the upstream water treatment. However, once the downstream village installs its own water treatment plant, the treatment status of the upstream village becomes irrelevant. The water quality in the downstream village is now only determined by its own treatment. Consequently, in this situation, treated units do not experience spillover effects.

Example 2 [continued]. Consider states that are distant from all states where stolen vehicle recovery devices have been installed in specific car models. These states might be unaffected by spillover effects because car thieves deterred from targeting models equipped with the device are likely to limit their alternative targets to those in areas within a manageable distance, for instance, due to their networks being more robust. Gonzalez-Navarro [2013] shows that the data supports the notion that geographical constraints limit displacement behavior.

Next, consider a car model without the device, located in a state adjacent to the one where the device had been installed. This car model is subject to spillover effects because installing the device in the neighbouring state prompts thieves to redirect their targets to models without the device in nearby areas. However, once the device is installed in these previously unprotected models, they no longer experience spillover effects, as thieves' attention turns to vehicles still lacking the device.

It might be argued that as the coverage of states and car models with the protection device expands to become almost universal, thieves could eventually revert to targeting protected cars, violating the assumption. This scenario might not be totally implausible unless thieves shift their focus to other, less protected assets or leave the criminal market entirely. Nevertheless, such an almost universal adoption of the protection device would be considered an extreme case and hard to evaluate due to a very small set of control units.

3 Setup

We consider a DiD model with staggered treatment adoption, which involves panel data of units observed over time periods $t \in \{1, \dots, T\}$. For each unit i at each time t, let D_{it} be the binary treatment status indicating whether the unit is treated (1) or not treated (0). We assume that the treatment is irreversible, meaning that once a unit undergoes treatment, it remains treated in all subsequent periods.

Assumption 1 (irreversibility). *For any $s < t$, $D_{is} = 1$ implies $D_{it} = 1$.*

Under Assumption 1, we can categorize units into groups according to the periods at which they receive treatment for the first time. Let G_i be the group label of unit i, defined by

$$G_i \equiv \begin{cases} \min\{t \mid D_{it} = 1\} & \text{if } D_{it} = 1 \text{ for some } t, \\ \infty & \text{if } D_{it} = 0 \text{ for all } t. \end{cases}$$

We let $\mathcal{G}$ be the support of G_i. Under this group label, let $D_i \equiv (D_{i1}, \dots, D_{iT})$ be the entire treatment history of unit i. It follows that:

$$D_i = (\underbrace{0, \ldots, 0}_{t<g}, \underbrace{1, \ldots, 1}_{t \geq g}) \quad \text{if} \quad G_i = g, \tag{3}$$

where D_i equals to a vector of zeros if $G_i = \infty$.

Let Y_{it} be the observed outcome of unit i at time t. Under SUTVA, it is standard to define the potential outcome by $Y_{it}(d_i)$, which is the outcome value when the own treatment history D_i is set to d_i. In the presence of spillover effects, this outcome is affected by the treatment histories of all units. We define the potential outcome by

$$Y_{it}(d_i, \mathbf{d}_{-i})$$

where d_i is the treatment history of unit i and $\mathbf{d}_{-i}$ is the treatment histories of all units other than i in the population. In the case of finite population, $\mathbf{d}_{-i}$ will be a $(N-1) \times T$ matrix, and in the case of infinite population, $\mathbf{d}_{-i}$ will be a mapping from an index to a $T \times 1$ vector. This notation does not impose any restrictions on the structure of the spillover effects.

Similarly to the case of SUTVA, we define treatment effects by comparing the observed potential outcome against various counterfactual outcomes. Let 0 and $\mathbf{0}_{-i}$ be the treatment histories where unit i and other units are untreated across all periods, respectively. Abusing notation, redefine d_i and $\mathbf{d}_{-i}$ to be the observed treatment histories in the data. The following four types of potential outcomes are relevant to our discussion:

- $Y_{it}(d_i, \mathbf{d}_{-i})$ represents the observed outcome where unit i is treated according to its group label, and other units are treated according to their group labels.
- $Y_{it}(d_i, \mathbf{0}_{-i})$ represents the counterfactual outcome where unit i is treated according to its group label, but all the other units are untreated.
- $Y_{it}(0, \mathbf{d}_{-i})$ represents the counterfactual outcome where unit i is untreated, but other units are treated according to their group labels.
- $Y_{it}(0, \mathbf{0}_{-i})$ represents the counterfactual outcome where both unit i and all the other units are untreated.

We assume no anticipatory effect exists for these four types of potential outcomes, a standard assumption in DiD analyses. Let $D_i^t \equiv (D_{i1}, \ldots, D_{it})$ be the current history of D_i up to time t, and define other variables with superscript t similarly.

Assumption 2 (no anticipation)**.** $Y_{it}(d, \mathbf{d}) = Y_{it}(d^t, \mathbf{d}^t)$ *for all* $d \in \{d_i, 0\}$ *and* $\mathbf{d} \in \{\mathbf{d}_{-i}, \mathbf{0}_{-i}\}$.

Next, we introduce the parallel trends assumption for a linear DiD model, following the assumption in the absence of spillovers in Borusyak et al. [2024] and Wooldridge [2023]. Later, we will extend our discussion to a nonlinear DiD model.

Assumption 3 (parallel trends, linear model)**.** *For every group g and time t,*

$$\mathbb{E}(Y_{it}(0, \mathbf{0}_{-i}) - Y_{i1}(0, \mathbf{0}_{-i})|G_i = g) = \mathbb{E}(Y_{it}(0, \mathbf{0}_{-i}) - Y_{i1}(0, \mathbf{0}_{-i})|G_i = \infty).$$

Remark 1. Assumption 3 can be written equivalently as

$$Y_{it}(0, \mathbf{0}_{-i}) = \alpha_i + \delta_t + \varepsilon_{it}, \tag{4}$$

where $\alpha_i = Y_{i1}(0, \mathbf{0}_{-i})$, $\delta_t = \mathbb{E}(Y_{it}(0, \mathbf{0}_{-i}) - Y_{i1}(0, \mathbf{0}_{-i})|G_i = \infty)$, and $\mathbb{E}(\varepsilon_{it}|G_i = g) = 0$ for all $g \in \mathcal{G}$. We refer to δ_t as the common time effect.

Remark 2. When $t = 1$ is the only pre-treatment period, Assumption 3 is necessary for the identification of the ATT. When there are multiple pre-treatment periods, Assumption 3 can be relaxed to:

$$\mathbb{E}(Y_{it}(0, \mathbf{0}_{-i}) - Y_{i,q-1}(0, \mathbf{0}_{-i})|G_i = g) = \mathbb{E}(Y_{it}(0, \mathbf{0}_{-i}) - Y_{i,q-1}(0, \mathbf{0}_{-i})|G_i = \infty),$$

where $q \equiv \min\{t \mid t \in \mathcal{G}\}$ is the first period in which any treated unit exists, meaning that $q - 1$ is the last period in which all units are untreated. Note that, in the absence of spillovers, it is sufficient to assume

$$\mathbb{E}(Y_{it}(0, \mathbf{0}_{-i}) - Y_{i,g-1}(0, \mathbf{0}_{-i})|G_i = g) = \mathbb{E}(Y_{it}(0, \mathbf{0}_{-i}) - Y_{i,g-1}(0, \mathbf{0}_{-i})|G_i = \infty),$$

implying that the parallel trend only needs to hold up to the group's last pre-treatment period $(g - 1)$, rather than the last universal pre-treatment period $(q - 1)$.

In the presence of spillover effects, multiple definitions of the ATT arise. We first introduce the ATT without interference:

$$ATT_0(g, t) \equiv \mathbb{E}(Y_{it}(d_i, \mathbf{0}_{-i}) - Y_{it}(0, \mathbf{0}_{-i})|G_i = g).$$

This definition of $ATT_0(g, t)$ captures the expected treatment effect at time t when unit i is the only treated unit in the population, thereby excluding any spillover effects from the other units. In other words, $ATT_0(g, t)$ captures the direct effect from the treatment, illustrated by the solid edges in Figure 1. This aligns with the conventional definition of the ATT under SUTVA and is the estimand of interest in our paper. We can then define an aggregate ATT by $ATT_0 = \sum_{g,t} w_{gt} ATT_0(g,t)$, where w_{gt} is a weight chosen by the econometrician (see, e.g., Callaway and Sant'Anna [2020]).[5]

We can also consider an alternative definition of the ATT:

$$ATT_S(g,t) \equiv \mathbb{E}(Y_{it}(d_i, \mathbf{d}_{-i}) - Y_{it}(0, \mathbf{0}_{-i})|G_i = g).$$

This definition differs from $ATT_0(g,t)$ in that it incorporates the spillover effects from other treated units, namely the units with group labels $g \leq t$.

We refer to the difference $ATT_S(g,t) - ATT_0(g,t)$ as the average spillover effect on the treated:

$$AST(g,t) \equiv \mathbb{E}(Y_{it}(d_i, \mathbf{d}_{-i}) - Y_{it}(d_i, \mathbf{0}_{-i})|G_i = g).$$

Lastly, it is useful to define another estimand, which we refer to as the average spillover effect on the untreated:

$$ASUT(g,t) \equiv \mathbb{E}(Y_{it}(0, \mathbf{d}_{-i}) - Y_{it}(0, \mathbf{0}_{-i})|G_i = g).$$

4 Identification

The discussion on the identification of $ATT_0(g,t)$ is structured into two steps. We first show that identifying $ATT_0(g,t)$ is equivalent to identifying the sum of the time effect and the spillover effect on the treated. The second step then introduces conditions that allow the identification of this sum. An implication of our assumptions is that it unifies the definitions of the ATT by implying that $ATT_0(g,t) = ATT_S(g,t)$.

We first present the necessary and sufficient condition for identifying $ATT_0(g,t)$ when spillovers are present.

Theorem 1. *Suppose that Assumptions 1 to 3 hold, and that all units are untreated at $t = 1$. Then, for every group $g \in \mathcal{G}$ such that $2 \leq g < \infty$ and time $t \geq g$, the parameter $ATT_0(g,t)$ is identified if and only if $\delta_t + AST(g,t)$ is identified.*

Proof. Refer to the Appendix for the proof of this theorem and others that follow. □

The proof of Theorem 1 shows that, for every (g,t) satisfying $t \geq g$:

$$\mathbb{E}(Y_{it}) = \mathbb{E}(\alpha_i) + \delta_t + ATT_0(g,t) + AST(g,t).$$

[5]Note that $ATT_0(g,t)$ is typically defined only for pairs (g,t) satisfying $t \geq g$. In this paper, we extend its definition to also include pairs satisfying $t < g$ with a trivial definition of $ATT_0(g,t) = 0$. We adopt this extension as it simplifies the notation in the proofs of our results.

The intuition for Theorem 1 is that since $\mathbb{E}(\alpha_i)$ is identified from the data for group g at $t = 1$, it follows that identification of $ATT_0(g,t)$ requires knowledge of δ_t (the time effect) and $AST(g,t)$ (the average spillover effect on the treated). In general, Assumptions 1 to 3 are not sufficient for the identification of these two parameters. Note that $AST(g,t) = 0$ in the absence of spillover effects, in which case the identification of the $ATT_0(g,t)$ only requires knowledge of the time effect.

In what follows, we propose two additional assumptions that enable identification of $ATT_0(g,t)$. We state the first assumption below.

Assumption 4 (no spillover effects on treated units)**.** *For every (g,t) such that $t \geq g$,*

$$\mathbb{E}(Y_{it}(d_i, \mathbf{d}_{-i})|G_i = g) = \mathbb{E}(Y_{it}(d_i, \mathbf{0}_{-i})|G_i = g).$$

This assumption holds if $Y_{it}(d_i, \mathbf{d}_{-i}) = Y_{it}(d_i, \mathbf{0}_{-i})$, implying that once a unit receives treatment, it is no longer influenced by spillover effects. This means the unit forfeits any spillovers it may have previously received and remains unaffected by spillovers from subsequently treated groups. Recall that we have previously discussed the plausibility of this assumption in Section 2, illustrated through Examples 1 and 2. Note that Assumption 4 is equivalent to $AST(g,t) = 0$, in which case $ATT_0(g,t) = ATT_S(g,t)$, unifying the definition of $ATT(g,t)$.

Next, we state the second assumption.

Assumption 5 (existence of never-treated units unaffected by spillovers)**.** *There exists a positive mass of units within group $G_i = \infty$, denoted by $H_i = 1$ for these units and $H_i = 0$ for all other units including those with $G_i \neq \infty$, such that:*

$$\begin{aligned} \mathbb{E}(Y_{it}(0, \mathbf{d}_{-i})|G_i = \infty, H_i = 1) &= \mathbb{E}(Y_{it}(0, \mathbf{0}_{-i})|G_i = \infty, H_i = 1), \\ \mathbb{E}(Y_{it}(0, \mathbf{0}_{-i}) - Y_{i1}(0, \mathbf{0}_{-i})|G_i = \infty, H_i = 1) &= \mathbb{E}(Y_{it}(0, \mathbf{0}_{-i}) - Y_{i1}(0, \mathbf{0}_{-i})|G_i = \infty). \end{aligned} \tag{5}$$

Equation (5) has two components. The first equality states that there exists a subgroup of never-treated units that are not affected by spillovers. The second equality states that Assumption 3 (parallel trends) extends to said subgroup, allowing for the identification of the time effect δ_t.

Remark 3. In practice, the researcher may take a conservative approach by setting $H_i = 1$ only for units strongly believed to satisfy Equation (5), and $H_i = 0$ otherwise.

Remark 4. The researcher may also have knowledge of units believed to be unaffected by spillovers for $G_i \neq \infty$. In this case, we could modify Assumption 5 such that:

$$\mathbb{E}(Y_{it}(0, \mathbf{d}_{-i})|G_i = g, H_i = 1) = \mathbb{E}(Y_{it}(0, \mathbf{0}_{-i})|G_i = g, H_i = 1),$$
$$\mathbb{E}(Y_{it}(0, \mathbf{0}_{-i}) - Y_{i1}(0, \mathbf{0}_{-i})|G_i = g, H_i = 1) = \mathbb{E}(Y_{it}(0, \mathbf{0}_{-i}) - Y_{i1}(0, \mathbf{0}_{-i})|G_i = \infty).$$

for all $g \in \mathcal{G}$.

For instance, in the study by Gonzalez-Navarro [2013] described in Example 2, the author takes a conservative approach using only the never-treated states that are farthest from the treated ones as controls (Remark 3). Alternatively, in a scenario where spillover effects occur only among adjacent states, we could set $H_i = 1$ for all states not adjacent to any treated one until they become treated (Remark 4). In this case, the control group is not fixed but shrinks over t as more states adopt treatment.

We conclude this section by showing that $ATT_0(g, t)$ is identified under these two additional assumptions, which is a direct consequence of Theorem 1.

Theorem 2. *Suppose that Assumptions 1 to 5 hold, and that all units are untreated at $t = 1$. Then, $ATT_0(g, t)$ is identified for every group $g \in \mathcal{G}$ such that $2 \leq g < \infty$ and time $t \geq g$. In particular,*

$$ATT_0(g, t) = \mathbb{E}(Y_{it} - Y_{i1}|G_i = g) - \mathbb{E}(Y_{it} - Y_{i1}|G_i = \infty, H_i = 1).$$

5 Estimation and Inference

In this section, we discuss estimation and inference of $ATT_0(g, t)$ under Assumptions 1 to 5. Consider a balanced panel of T periods, where all units are untreated at $t = 1$. For estimation, it is useful to introduce a binary variable S_{it} that indicates whether an observation (i, t) could be subject to spillover effects. To define this variable, recall that an observation is potentially influenced by spillovers under the following conditions:

- There exists a treated unit (i.e., $t \geq q$, where q is the first period that any unit enters treatment).
- The observation is not treated (i.e., $D_{it} = 0$), as otherwise treated observations are not influenced by spillover effects by Assumption 4.
- The observation has $H_i = 0$, as otherwise untreated units with $H_i = 1$ are considered unaffected by spillover effects by Assumption 5.

We define S_{it} as

$$S_{it} = \begin{cases} 1 & \text{if } t \geq q \text{ and } D_{it} = 0 \text{ and } H_i = 0 \\ 0 & \text{otherwise} \end{cases}.$$

We first discuss the estimation of $ATT_0(g,t)$ in the case where units with $H_i = 1$ exist only within group $G_i = \infty$. It is useful to introduce an extended group label that partitions $G_i = \infty$ into $(G_i = \infty, H_i = 1)$ and $(G_i = \infty, H_i = 0)$. Define $\tilde{\mathcal{G}}$ as the support of $(G_i, H_i) \in \mathcal{G} \times \{0,1\}$. For example, if $\mathcal{G} = \{q, q+1, \ldots, T, \infty\}$, then $\tilde{\mathcal{G}} = \{(q,0), (q+1,0), \ldots, (T,0), (\infty,0), (\infty,1)\}$. We propose the following extension of Wooldridge [2022] as the estimation procedure. We estimate a linear regression model where Y_{it} is the outcome variable, and the regressors are:

- indicators of (G_i, H_i) (the "extended group fixed effects"),
- indicators of t (the "time fixed effects"),
- interactions between D_{it} and indicators of (G_i, H_i, t), and
- interactions between S_{it} and indicators of (G_i, H_i, t).

In other words, we estimate the linear regression model:

$$\begin{aligned} Y_{it} = \alpha_{G_i H_i} + \delta_t &+ \sum_{(g',h') \in \tilde{\mathcal{G}}} \sum_{t'} \beta_{g'h't'} \cdot \mathbf{1}((G_i, H_i, t) = (g', h', t')) \cdot D_{it} \\ &+ \sum_{(g',h') \in \tilde{\mathcal{G}}} \sum_{t'} \gamma_{g'h't'} \cdot \mathbf{1}((G_i, H_i, t) = (g', h', t')) \cdot S_{it} + \varepsilon_{it}. \end{aligned} \tag{6}$$

Since this equation contains multicollinear terms, it can also be written as

$$\begin{aligned} Y_{it} = \alpha_{G_i H_i} + \delta_t &+ \sum_{g' \in \mathcal{G} \setminus \{\infty\}} \sum_{t'=g'}^{T} \beta_{g't'} \cdot \mathbf{1}((G_i, t) = (g', t')) \cdot D_{it} \\ &+ \sum_{(g',h') \in \tilde{\mathcal{G}} \setminus \{(\infty,1)\}} \sum_{t'=q}^{\min\{g'-1, T\}} \gamma_{g'h't'} \cdot \mathbf{1}((G_i, H_i, t) = (g', h', t')) \cdot S_{it} + \varepsilon_{it}. \end{aligned} \tag{7}$$

Note that Equation (7) involves the *group*-level coefficients $(\alpha_{gh}, \beta_{gt})$, as opposed to the *unit*-level coefficients (α_i, β_{it}).[6] The following result shows that, despite this simplification, the population regression of Equation (7) correctly identifies $ATT_0(g,t)$.

Theorem 3. *Suppose that the assumptions of Theorem 2 hold. Consider the population regression of Equation (7). Then,* $\beta_{gt} = ATT_0(g,t)$.

This result yields a simple and straightforward procedure for estimation and inference of $ATT_0(g,t)$. The estimate $\widehat{\beta}_{gt}$ and its standard error can be easily obtained by implementing Equation (7) using any standard statistical software package. Estimation and

[6] This will become important when comparing our estimation procedure with imputation-based procedures later in this section, and when discussing estimation in the non-linear case in Section 6.

inference of an aggregate ATT is also straightforward, because the estimate is given by $\sum_{g,t} w_{gt}\widehat{\beta}_{gt}$, and its standard error is given by

$$\text{Var}\left(\sum_{g,t} w_{gt}\widehat{\beta}_{gt}\right) = \sum_{g,t}\sum_{g',t'} w_{gt}w_{g't'}\text{Cov}(\widehat{\beta}_{gt}, \widehat{\beta}_{g't'}),$$

where $\text{Cov}(\widehat{\beta}_{gt}, \widehat{\beta}_{g't'})$ is available in any statistical software package, e.g., via e(V) in Stata.

It is worth noting that Equation (7) is numerically equivalent to the following extension of the imputation-based procedure of Borusyak et al. [2024]:

1. Estimate the linear model
$$Y_{it} = \alpha_i + \delta_t + \varepsilon_{it},$$
using observations (i, t) such that $D_{it} = 0$ and $S_{it} = 0$.

2. Let $\widehat{\alpha}_i$ and $\widehat{\delta}_t$ be the estimates of α_i and δ_t. Impute the baseline outcome for unit i at time t as
$$\widehat{Y}_{it}(0, \mathbf{0}_{-i}) = \widehat{\alpha}_i + \widehat{\delta}_t.$$

3. For each unit i treated at time t, compute
$$\widehat{\beta}_{it}^{imp} \equiv Y_{it} - \widehat{Y}_{it}(0, \mathbf{0}_{-i}),$$
which can be interpreted as the imputed treatment effect for unit i at time t.

4. For a treated group g at time $t \geq g$, estimate $ATT_0(g, t)$ by
$$\widehat{\beta}_{gt}^{imp} \equiv \frac{1}{N_g}\sum_{i=1}^{N_g} \widehat{\beta}_{it}^{imp},$$
where $N_g = \sum_{i=1}^{N} \mathbf{1}(G_i = g)$.

While it can be shown that $\widehat{\beta}_{gt}^{imp}$ equals to $\widehat{\beta}_{gt}$ in Equation (7), the use of group-level coefficients in Equation (7) eases the calculation of standard errors through the use of standard statistical software packages. Along these lines, Borusyak et al. [2024] highlight the challenge in estimating the standard error of the imputation estimate, which arises from computing $\widehat{\beta}_{it}^{imp}$ for each i and t.

Next, we consider the case where there are also units with $H_i = 1$ in groups other than ∞. Recall that the extended group label (G_i, H_i) partitions each group $G_i = g$ into subgroups $(G_i = g, H_i = 1)$ and $(G_i = g, H_i = 0)$. We define ATT_0 for these subgroups as follows:

$$ATT_0(g, h, t) = \mathbb{E}(Y_{it}(d_i, \mathbf{0}_{-i}) - Y_{it}(0, \mathbf{0}_{-i})|G_i = g, H_i = h).$$

We introduce this new definition because the ATTs for the subgroups $(G_i = g, H_i = 0)$ and $(G_i = g, H_i = 1)$ may differ.[7] The aggregate ATT can then be defined as $ATT_0 = \sum_{g,h,t} w_{ght} ATT_0(g, h, t)$, where w_{ght} is a weight chosen by the econometrician. The same estimation procedure described earlier applies to this case as well, where we run the full regression as described in Equation (6), and the coefficient β_{ght} equals to $\beta_{ght} = ATT_0(g, h, t)$.

Lastly, if the data is an unbalanced panel, the regression in Equation (7) is no longer consistent for the ATT_0. The imputation-based estimation procedure discussed above is still consistent, but the standard error will be asymptotically conservative in general (see Borusyak et al., 2024, Section 4.3). In contrast, for a balanced panel, the standard error computed from Equation (7) is asymptotically exact.

6 Extension to Nonlinear DiD Models

In this section, we extend our previous findings to the case where Y_{it} is a count variable, for which the linear parallel trends condition (Assumption 3) does not hold. This extension contributes to the literature on nonlinear DiD models [Wooldridge, 2023], expanding the applicability of our results to a wider array of empirical applications.

We introduce the following assumption regarding parallel trends in the context of count data, following Wooldridge [2023].

Assumption 3' (parallel trends, Poisson model)**.** *For every group g at time t,*

$$\begin{aligned} &\ln \mathbb{E}(Y_{it}(0, \mathbf{0}_{-i})|G_i = g) - \ln \mathbb{E}(Y_{i1}(0, \mathbf{0}_{-i})|G_i = g) \\ &= \ln \mathbb{E}(Y_{it}(0, \mathbf{0}_{-i})|G_i = \infty) - \ln \mathbb{E}(Y_{i1}(0, \mathbf{0}_{-i})|G_i = \infty) \end{aligned}. \tag{8}$$

Remark 5. Assumption 3' can be written equivalently as, conditional on $G_i = g$,

$$Y_{it}(0, \mathbf{0}_{-i}) = \exp\{\alpha_g + \delta_t\}\varepsilon_{it}, \tag{9}$$

where $\alpha_g = \ln \mathbb{E}(Y_{i1}(0, \mathbf{0}_{-i})|G_i = g)$, $\delta_t = \ln \mathbb{E}(Y_{it}(0, \mathbf{0}_{-i})|G_i = \infty) - \ln \mathbb{E}(Y_{i1}(0, \mathbf{0}_{-i})|G_i = \infty)$, and $\mathbb{E}(\varepsilon_{it}|G_i = g) = 1$ for all $g \in \mathcal{G}$.

Remark 6. More generally, for any strictly increasing function F, we could consider the parallel trend of the form

$$\begin{aligned} &F\left(\mathbb{E}(Y_{it}(0, \mathbf{0}_{-i})|G_i = g)\right) - F\left(\mathbb{E}(Y_{i1}(0, \mathbf{0}_{-i})|G_i = g)\right) \\ &= F\left(\mathbb{E}(Y_{it}(0, \mathbf{0}_{-i})|G_i = \infty)\right) - F\left(\mathbb{E}(Y_{i1}(0, \mathbf{0}_{-i})|G_i = \infty)\right) \end{aligned}.$$

[7] In the previous case, where units with $H_i = 1$ existed only within group $G_i = \infty$, there were no units with $(G_i = g, H_i = 1)$ when $g \neq \infty$, so $ATT(g, t) = ATT(g, 0, t)$.

For example, for binary data, F can be set to be the inverse of the Gaussian CDF (probit) or the inverse of the Logistic function (logit). Our results in this section extends straightforwardly to these other choices of F.

By replicating the arguments in Theorems 1 and 2, the following corollaries show that $ATT_0(g,t)$ is identified under assumptions similar to those in Theorem 2. In doing so, we adapt Assumption 5 to the case of count data.

Assumption 5' (existence of never-treated units unaffected by spillovers)**.** *There exists a positive mass of units within group $G_i = \infty$, denoted by $H_i = 1$ for these units and $H_i = 0$ for all other units including those with $G_i \neq \infty$, such that:*

$$\mathbb{E}(Y_{it}(0, \mathbf{d}_{-i})|G_i = \infty, H_i = 1) = \mathbb{E}(Y_{it}(0, \mathbf{0}_{-i})|G_i = \infty, H_i = 1)$$

and

$$\begin{aligned} &\ln \mathbb{E}(Y_{it}(0, \mathbf{0}_{-i})||G_i = \infty, H_i = 1) - \ln \mathbb{E}(Y_{i1}(0, \mathbf{0}_{-i})|G_i = \infty, H_i = 1) \\ &= \ln \mathbb{E}(Y_{it}(0, \mathbf{0}_{-i})|G_i = \infty) - \ln \mathbb{E}(Y_{i1}(0, \mathbf{0}_{-i})|G_i = \infty). \end{aligned} \tag{10}$$

Corollary 1. *Suppose that Assumptions 1 and 2 and assumption 3' hold, and that all units are untreated at $t = 1$. Then, for every group $g \in \mathcal{G}$ such that $2 \leq g < \infty$ and time $t \geq g$, the $ATT_0(g,t)$ is identified if and only if $\exp\{\alpha_g + \delta_t\} + AST(g,t)$ is identified.*

Corollary 2. *Suppose that Assumptions 1, 2 and 4 and assumptions 3' and 5' hold, and that all units are untreated at $t = 1$. Then, $ATT_0(g,t)$ is identified for every group $g \in \mathcal{G}$ such that $2 \leq g < \infty$ and time $t \geq g$. In particular,*

$$ATT_0(g,t) = \mathbb{E}(Y_{it}|G_i = g) - \frac{\mathbb{E}(Y_{it}|G_i = \infty, H_i = 1)}{\mathbb{E}(Y_{i1}|G_i = \infty, H_i = 1)} \mathbb{E}(Y_{i1}|G_i = g).$$

In terms of estimation, let S_{it} be defined as in the previous section, and consider a balanced panel of T periods where all units are untreated at $t = 1$. In the case of count data, the average treatment effect in terms of percentage changes is also often reported:

$$ATTP_0(g,t) = \frac{ATT_0(g,t)}{\mathbb{E}(Y_{it}(0, \mathbf{0}_{-i})|G_i = g)},$$

which can be aggregated to define an $ATTP_0 \equiv \sum_{g,t} w_{gt} ATTP_0(g,t)$.

The estimation and inference procedure discussed in Section 5 can be straightforwardly extended to the count data. For example, in the case where units with $H_i = 1$ are all within group $G_i = \infty$, we use the following simple estimation procedure that involves a parsimonious generalized linear model.

1. Estimate the Poisson regression model where Y_{it} is the outcome variable, and the regressors are:

   - indicators of (G_i, H_i) (the "extended group fixed effects"),
   - indicators of t (the "time fixed effects"),
   - interactions between D_{it} and indicators of (G_i, H_i, t), and
   - interactions between S_{it} and indicators of (G_i, H_i, t).

   In other words, we estimate the Poisson regression model:

$$
\begin{aligned}
\ln \mathbb{E}(Y_{it}|\mathbf{X}_{it}) =& \alpha_{G_iH_i} + \delta_t + \sum_{g' \in \mathcal{G} \setminus \{\infty\}} \sum_{t'=g'}^{T} \beta_{g't'} \cdot \mathbf{1}((G_i, t) = (g', t')) \cdot D_{it} \\
&+ \sum_{(g',h') \in \tilde{\mathcal{G}} \setminus \{(\infty,0)\}} \sum_{t'=q}^{\min\{g'-1,T\}} \gamma_{g'h't'} \cdot \mathbf{1}((G_i, H_i, t) = (g', h', t')) \cdot S_{it},
\end{aligned}
\tag{11}
$$

   where $\mathbf{X}_{it}$ represents the vector of regressors. Let $\widehat{\alpha}_{gh}$, $\widehat{\delta}_t$, and $\widehat{\beta}_{gt}$ be the estimates of α_{gh}, δ_t, and β_{gt} from this model, respectively.

2. Estimate $ATT_0(g,t)$ by

$$
\widehat{ATT}_0(g,t) = \exp\{\widehat{\alpha}_{g0} + \widehat{\delta}_t + \widehat{\beta}_{gt}\} - \exp\{\widehat{\alpha}_{g0} + \widehat{\delta}_t\},
$$

   or estimate $ATTP_0(g,t)$ by $\widehat{ATTP}_0(g,t) = \exp\{\widehat{\beta}_{gt}\} - 1$.

The validity of the population regression of Equation (11) can be shown by replicating the arguments in Theorem 3, and we omit the proof here. The estimation and inference of $\widehat{ATT}_0(g,t)$ can then be carried out by implementing Equation (11) using any standard statistical software package that runs Poisson regressions.

Note that most statistical software packages that run Poisson regressions calculate the standard errors of $(\widehat{\alpha}_{gh}, \widehat{\delta}_t, \widehat{\beta}_{gt})$ using the maximum likelihood. This assumes that the distribution of $Y_{it}(0, \mathbf{0}_{-i})$ follows a Poisson distribution (as opposed to only specifying its mean as in Assumption 3'), ruling out heteroskedasticity. To accommodate heteroskedasticity, standard errors can instead be derived using the quasi-maximum likelihood estimation (QMLE) method. Specifically, let θ be the vector of all coefficients in the Poisson regression (i.e., all of α_{gh}, δ_t, β_{gt}, and γ_{ght}), $\hat{\theta}$ be their maximum likelihood estimates (i.e., all of $\widehat{\alpha}_{gh}$, $\widehat{\delta}_t$, $\widehat{\beta}_{gt}$, and $\widehat{\gamma}_{ght}$), and $\mathbf{X}_{it}$ be the vector of all regressors. Let

$\{\Lambda^c\}_{c=1}^C$ be the partition of units according to which the units are clustered. Define

$$\mathcal{S} = \sum_{c=1}^{C} \left[\sum_{i \in \Lambda^c} \sum_{t=1}^{T} \mathbf{X}_{it}(Y_{it} - \widehat{Y}_{it}) \right] \left[\sum_{i \in \Lambda^c} \sum_{t=1}^{T} \mathbf{X}_{it}(Y_{it} - \widehat{Y}_{it}) \right]'$$

as the clustered outer product of the score function, where $\widehat{Y}_{it} = \exp\{X_{it}'\hat{\theta}\}$ is the fitted value of Y_{it} in the Poisson regression.[8] In addition, define

$$\mathcal{H} = \sum_{c=1}^{C} \sum_{i \in \Lambda^c} \sum_{t=1}^{T} \mathbf{X}_{it}\mathbf{X}_{it}' \widehat{Y}_{it}$$

as the negative Hessian function. Then, the variance-covariance matrix of $\hat{\theta}$ is given by

$$\widehat{\text{Var}}(\hat{\theta}) = \mathcal{H}^{-1}\mathcal{S}\mathcal{H}^{-1}.$$

This variance-covariance matrix can then be used to compute the standard errors of the ATT_0 and $ATTP_0$ estimates via the delta method.

7 Application to Auto Theft Prevention Policy

In this section, we apply our method to revisit the findings of Gonzalez-Navarro [2013], who studied the effects of installing an auto theft prevention device known as Lojack. This was a compact device installed in vehicles, allowing for tracking of the vehicle.

The policy was implemented in Mexico through an exclusive agreement between the Ford Motor Company and the Lojack company. Initially, the technology was introduced for a particular Ford car model (Ford Windstar) in a specific state (Jalisco) among the 2001 car models. Subsequently, the installation of Lojack expanded to include other *model* $\times$ *state* combinations, eventually encompassing 32 *model* $\times$ *state* combinations by 2004. The dataset of Gonzalez-Navarro [2013] provides comprehensive information on car theft for each *model* $\times$ *state* $\times$ *vintage* (the car model's year) combination, for each calendar year. For our analysis, we use the indices m, s, v, and t to represent car model, state, vintage, and the calendar year of the auto theft, respectively.

Gonzalez-Navarro [2013] points out two possible sources of spillover effects following the introduction of Lojack. The first potential source is within-state spillover to car models not equipped with Lojack. Given the public knowledge about specific car models and states where Lojack was installed, criminals may alter their target preferences, focusing on car models without Lojack within the same state. The second source is geographical

[8] We abuse notation and let $\widehat{Y}_{it}$ represent a different object from the linear case.

spillovers, where installing Lojack in certain models may prompt thieves, particularly those specializing in those models, to shift their operations to other states where these specific models remain unprotected by Lojack.

Because of the potential for such spillovers, Gonzalez-Navarro [2013] relies only on time-series variation for identification, illustrating the challenge in extending the DiD framework to spillovers:

> *"In the presence of spatial externalities, DiD estimation using observations from different geographical locations produces biased estimates of policy impact. The basic challenge is that whenever treatment in one geographical location also has effects in control locations, these are no longer valid counterfactual observations. Furthermore, DiD estimation precludes actual estimation of externalities unless there is a set of observations subject to externalities and a set of observations that is not, so that the latter can play the role of counterfactual. For these reasons I do not use DiD estimation. Instead, I use an interrupted time series strategy in which the counterfactual is given by observations occurring before the intervention."*

Nevertheless, as a robustness check, Gonzalez-Navarro [2013] also estimates a DiD model while attempting to control for spillover effects, but without accounting for the staggered adoption design. In this section, we apply our method to revisit this study and estimate the treatment effect across various combinations of groups and time periods, thereby revealing the heterogeneous effects of Lojack installation.

Once Lojack was installed in a particular combination of car model, state, and vintage, it continued to be installed in all subsequent vintages of that model in the same state. This setup allows us to treat the situation as a staggered adoption design, where the unit of analysis is defined as the combination of *model* (m) $\times$ *state* (s) $\times$ *age* (a). *age* refers to the number of years elapsed since the car's model year, calculated as the difference between the calendar year (t) and the vintage year (v), such that $a = t - v$. Under this framework, our analysis is based on a balanced panel subset derived from the original dataset, consisting of 1152 units observed over 6 years from 1999 to 2004.

We define the binary treatment indicator for a unit (m, s, a) at time t as D_{msat}. To illustrate, consider the Ford Windstar model in Jalisco. For this unit, Lojack has been installed in all newly released $(age = 0)$ vehicles starting in 2001. Thus, for a Ford Windstar model in Jalisco with $age = 0$, we have $D_{Windstar,Jalisco,0,t} = 1$ for every $t \geq 2001$.

Our method relies on Assumptions 4 and 5. Assumption 4 requires that once a *model* $\times$ *state* $\times$ *age* unit has Lojack installed, it is not influenced by spillover effects. Generally, when Lojack is installed in certain units, we can expect that thieves targeting those models will shift their focus towards vehicles without Lojack protection, rather than those already with Lojack. Thus, it is reasonable to assume that units already fitted with Lojack will not be subject to displacement effects from other units, satisfying Assumption 4. Assumption 5 requires that there exist units which are not affected by spillover effects, and Gonzalez-Navarro [2013] provides empirical support for this assumption, demonstrating that car models in states geographically distant from those where the treatment was applied do not experience spillover effects.[9]

Let Y_{msat} be the number of auto thefts for a *model* $\times$ *state* $\times$ *age* unit that occurred in a given calendar year t. We consider two kinds of empirical models for this outcome. First, we consider linear parallel trends:

$$\mathbb{E}(Y_{msat}(0, \mathbf{0}_{-(msa,t)})|\alpha_{msa}, G_{msa} = g) = \alpha_{msa} + \delta_t.$$

This is equivalent to Assumption 3, where the combination (m, s, a) plays the role of i. Second, we consider Poisson parallel trends:

$$\ln \mathbb{E}(Y_{msat}(0, \mathbf{0}_{-(msa,t)})|\alpha_{msa}, G_{msa} = g) = \alpha_{msa} + \delta_t,$$

which is equivalent to Assumption 3'. The second model is particularly suitable when Y_{msat} is a count variable with a high frequency of zeros, in which case a Poisson regression model is more appropriate.

We define H_{msa} as a binary variable that is equal to 1 if unit (m, s, a) is such that s is a state that is not adjacent to any state with treated units throughout the rollout of Lojack. We then define S_{msat} as a binary indicator that is equal to 1 if $t \geq 2001$, $D_{msat} = 0$ and $H_{msa} = 0$. In addition, define $\overline{\mathcal{G}} = \{2001, 2002, 2003, 2004\}$ as the set of group labels for treated units, which are the periods when units enter treatment. Let N_g be the number of units in group $g \in \overline{\mathcal{G}}$ within the dataset, and let $\overline{N} \equiv \sum_{g=2001}^{2004} \sum_{t=g}^{2004} N_g = \sum_{g=2001}^{2004} (2005 - g) N_g$ be the total number of treated observations in the dataset. We

[9]The results of Gonzalez-Navarro [2013] using only time series variation vs. the DID approach are similar, suggesting that the units in states distant from the treated areas are unaffected by the installation of Lojack.

	Linear			Poisson		
	Estimate	Std Error	Reduction	Estimate	Std Error	Reduction
ATT_0	-6.1017	2.8893	-60%	-5.6349	2.5086	-66%
ATT_0^0	-3.9455	2.9166	-38%	-3.8738	2.4503	-50%
ATT_0^1	-6.7536	2.9801	-77%	-6.2742	2.5453	-77%
ATT_0^2	-16.9622	2.9691	-79%	-13.4790	4.2276	-85%

Table 1: Estimates of the aggregate ATT_0s. The standard errors are clustered at the *model* (m) $\times$ *state* (s) $\times$ *age* (a) level. The "Reduction" column stands for the reduction rate, which is calculated using the formula for computing $ATTP_0$.

estimate the following aggregate ATT_0s:

$$ATT_0 = \sum_{g=2001}^{2004} \sum_{t=g}^{2004} \frac{N_g}{\overline{N}} ATT_0(g,t),$$

$$ATT_0^0 = \sum_{g=2001}^{2004} \frac{N_g}{N_{2001} + \cdots + N_{2004}} ATT_0(g,g),$$

$$ATT_0^1 = \sum_{g=2001}^{2003} \frac{N_g}{N_{2001} + \cdots + N_{2003}} ATT_0(g,g+1),$$

$$ATT_0^2 = \sum_{g=2001}^{2002} \frac{N_g}{N_{2001} + N_{2002}} ATT_0(g,g+2).$$

In the above definitions, ATT_0 measures the overall effect of Lojack installation, computed as the weighted average of all $ATT_0(g,t)$ values across g and t. The ATT_0^k values, on the other hand, represent the weighted average of ATT_0 for the k-th year after installation of Lojack, measuring the temporal effects. For example, ATT_0^0 represents the immediate effect in the same year as the Lojack installation, ATT_0^1 represents the effect one year post-installation, and so forth.

Table 1 presents the estimated ATT_0 values obtained from both linear and Poisson model specifications, with standard errors clustered at the unit level. The analysis reveals a notable average reduction in thefts of 60% for the linear model and 64% for the Poisson model, highlighting Lojack's substantial deterrent effect. Moreover, the results from both models indicate that the rate of theft reduction becomes more pronounced over time, where the effect becomes statistically significant starting one year after installation. This highlights the increasing effectiveness of Lojack in preventing auto thefts over time.

For comparison, we also report the estimated ATT_0s from two misspecified models. First, we consider the TWFE specification that incorporates spillover effects but overlooks the staggered adoption nature of the treatment. Second, we consider the specification of

Wooldridge [2022] and Borusyak et al. [2024] that accounts for staggered adoption but does not include spillover effects. The results from these models are presented in Table 2. We find that the TWFE regression estimate closely aligns with the estimates presented in Table 1. However, the estimates that neglect spillover effects exhibit an upward bias relative to the correctly specified estimates in Table 1. This is what we would expect in the presence of displacement effects, where installing Lojack in a treated unit increases theft for units without Lojack.

	TWFE-Linear		WB-Linear	
	Estimate	Reduction	Estimate	Reduction
ATT_0	-7.8595	-69%	-7.8335	-72%
ATT_0^0	N/A		-5.6375	-58%
ATT_0^1	N/A		-8.4526	-82%
ATT_0^2	N/A		-19.1385	-88%
	TWFE-Poisson		**WB-Poisson**	
	Estimate	Reduction	Estimate	Reduction
ATT_0	-5.4990	-61%	-5.8514	-62%
ATT_0^0	N/A		-3.9569	-43%
ATT_0^1	N/A		-6.4894	-73%
ATT_0^2	N/A		-14.5736	-94%

Table 2: Estimates of the aggregate ATT_0s using the TWFE specification (the "TWFE" columns), and the specification of Wooldridge [2022] and Borusyak et al. [2024] (the "WB" columns), for each of linear and Poisson specifications. The "Reduction" columns stand for the reduction rate, which is calculated using the formula for computing $ATTP_0$.

8 Monte Carlo Simulation

In this section, we study the finite sample properties of our estimator in a simulated dataset, highlighting the bias-variance trade-off of our approach. We consider a balanced panel dataset over T periods, with either a simultaneous or staggered adoption design, starting with a pre-treatment period of $t = 1$. We consider M units in each group $g \in \mathcal{G} \equiv \{2, \ldots, T, (\infty, 0), (\infty, 1)\}$, meaning that we have a total of $N = (T + 1)M$ units in the dataset. In the absence of spillover effects, our estimator is less efficient than conventional estimators that rule out spillover effects. However, in the presence of spillovers, the conventional estimators become biased. Given this bias-variance trade-off, when the sample size is small, the improvement in bias may not sufficiently offset the loss in precision.

Specifically, we consider the following data generating process (DGP) that embeds Assumptions 1 to 5. Depending on the specification of the outcome model—linear or

Poisson—we adapt the relevant assumption, replacing Assumption 3 with 3' as necessary. The DGP is given by:

$$\mathbb{E}(Y_{it}|\alpha_i, G_i = g) = F\left(\alpha_i + \delta_t + \beta_{it} D_{it} + (1 - D_{it}) \cdot \sum_{h \in \mathcal{G} \setminus \{g\}} \sum_{j=1}^{M} \gamma_{it}^{j} \cdot D_{jt}\right),$$

where the function F is $F(x) = x$ for the linear model or $F(x) = \exp(x)$ for the Poisson model, and γ_{it}^{j} represents the spillover effect from unit j to unit i. We parametrize the DGP as follows.

- The M units in each group $g \in \{2, \dots, T, (\infty, 0), (\infty, 1)\}$ are homogeneous, implying that $\alpha_i = \alpha_{G_i}$, $\beta_{it} = \beta_{G_i t}$ and $\gamma_{it}^{j} = \gamma_{G_i t}^{G_j}$.
- Unit fixed effects are set to $\alpha_g = 26 - g + 1$ for all groups except for $(\infty, 0)$ and $(\infty, 1)$, where $\alpha_{(\infty,0)} = \alpha_{(\infty,1)} = 26 - T + 1$. This reflects selection into treatment because the units with earlier treatment have larger unit fixed effects. In the case of the Poisson model, we instead set $\alpha_g = \log(26 - g + 1)$.
- Common time effects are set to $\delta_t = \bar{\alpha} \times 0.1 \times ((t - 1) + \sin(t))$, where $\bar{\alpha}$ is the average of the unit fixed effects across all groups. This specification involves a linear upward trend $(t - 1)$ and a period-specific fluctuation modeled through $\sin(\cdot)$.
- The treatment effect is set to $\beta_{gt} = 0.5\alpha_g/t$. This effect is heterogeneous across groups and time periods, but homogeneous within a group. The effect gradually diminishes over time, with β_{gt} decreasing in t for each group g. The immediate effect β_{gg} is largest for group $g = 2$ with the highest α_g. This parametrizes sorting on gain since α_g also correlates with treatment timing.
- Spillover effects are set to $\eta_{gt}^{h} = -\rho \cdot \beta_{gt}/U_t$, representing displacement effects, where U_t is the number of untreated units at time t except for those in $(\infty, 0)$. That is, for each treated unit i, we consider a total spillover effect of $-\rho \cdot \beta_{G_i t}$, where $\rho \in [0, 1]$ denotes the spillover intensity. This total effect is then evenly spread among all untreated units excluding those in $(\infty, 0)$. As a result, each untreated unit receives a spillover effect of $-\rho \cdot \beta_{G_i t}/U_t$ from the treated unit i.

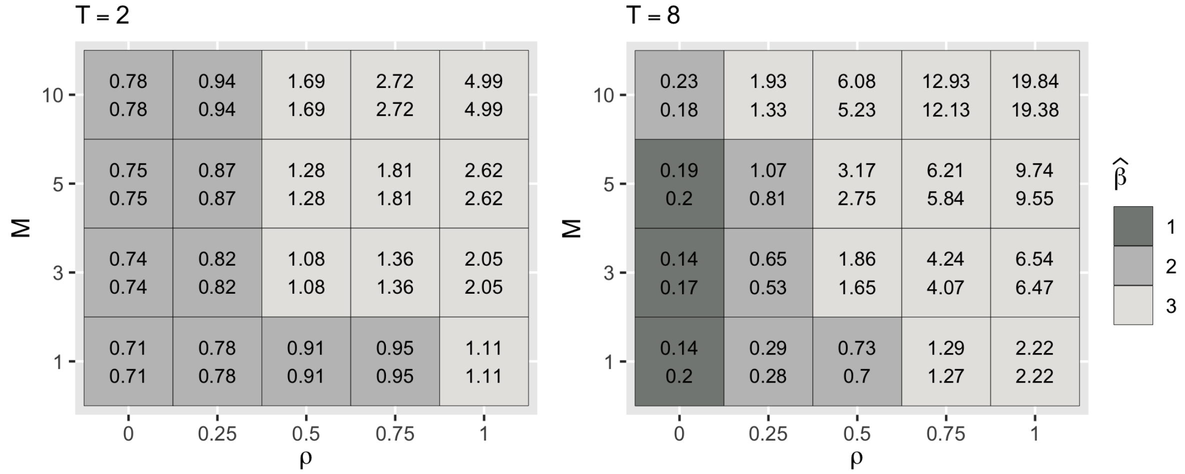
With this parametrization, Y_{it} is generated with an independent additive error term $\epsilon_{it} \sim N(0, \max(\alpha_{G_i})/10)$ for the linear model, and according to Poisson distribution for the Poisson model. We then estimate the aggregate ATT_0 defined as in Section 7, namely $ATT_0 = (1/\overline{G}) \sum_{g=2}^{T} \sum_{t=g}^{T} ATT_0(g, t)$, where $\overline{G} \equiv T(T-1)/2$ is the total number of treated group-time pairs in the dataset. We compare the mean Absolute Bias and the Mean Squared Error (MSE) across the following estimators:

($\widehat{\beta}_1$) The TWFE estimator, which neither accounts for staggered treatment adoption nor for spillovers.

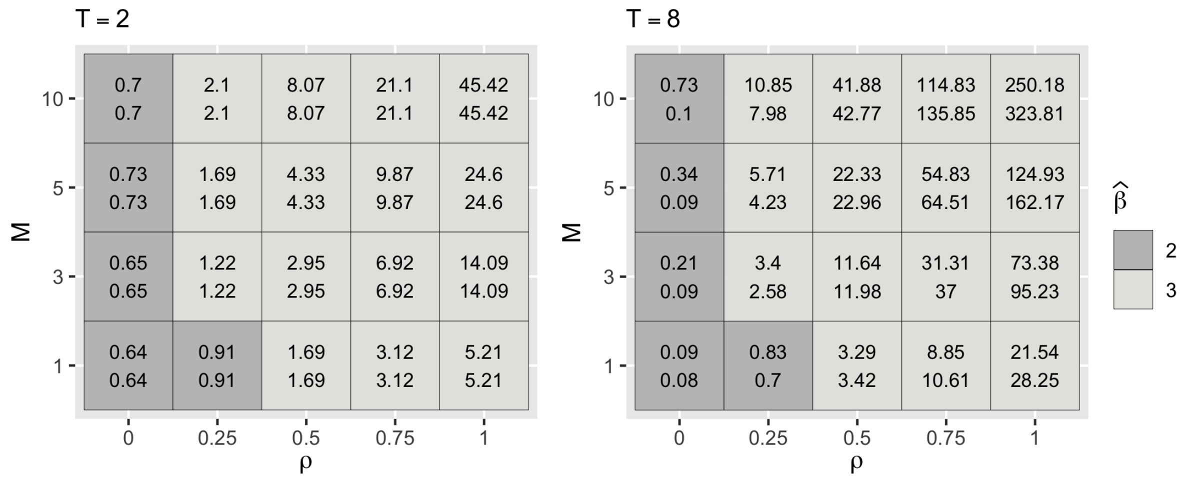
($\widehat{\beta}_2$) The extended TWFE estimator by Wooldridge [2022], which accounts for staggered treatment adoption but does not account for spillovers. This estimator is numerically equivalent to the imputation estimator by Borusyak et al. [2024].

($\widehat{\beta}_3$) Our estimator, which accounts for both staggered treatment adoption and spillovers.

Figure 4a and Table 3 present results from the linear DGP. The Figure visually contrasts the MSE across the three estimators to illustrate their relative performance under different scenarios, while the Table details their MSE and Absolute Bias values. Note that, when $T = 2$, the TWFE and the Wooldridge [2022] estimators are equivalent since treatment is not staggered. Overall, the relative performances of the estimators depend on the degree of spillovers, staggered treatment, and the number of units in each group. Intuitively, due to its efficiency, the TWFE has the lowest MSE in scenarios with no or little spillovers and with very few observations. As the number of observations increases and spillovers remain small, the Wooldridge [2022] estimator becomes the best-performing one, adjusting for staggered treatment without substantial bias. However, in scenarios where spillovers are not negligible and the number of units is large, our estimator achieves the lowest MSE, often by a large margin. Our estimator also performs better as treatment becomes more staggered ($T = 8$), highlighting our estimator's ability to accurately account for cumulative spillovers affecting the untreated units' outcomes. Furthermore, Figure 4b and Table 4 present results from the Poisson DGP, where our estimator performs even better relative to the TWFE and the Wooldridge [2022] ones.



(a) Linear



(b) Poisson

Figure 4: Comparison of the MSEs. Cell background color indicates the best-performing estimator. The numbers in cells represent the MSE ratios $\frac{MSE_1}{MSE_3}$ and $\frac{MSE_2}{MSE_3}$ respectively. The subscripts refer to: (1) TWFE estimator, (2) Wooldridge [2022] estimator, and (3) our estimator.

Table 3: MSE and Absolute Bias values - Linear

ρ	T	M	ATT	$\lvert Bias_1 \rvert$	$\lvert Bias_2 \rvert$	$\lvert Bias_3 \rvert$	MSE_1	MSE_2	MSE_3
0.00	2	1	-6.500	**3.589**	**3.589**	4.235	**19.642**	**19.642**	27.519
		3	-6.500	**2.060**	**2.060**	2.374	**6.613**	**6.613**	8.957
		5	-6.500	**1.615**	**1.615**	1.873	**3.991**	**3.991**	5.335
		10	-6.500	**1.159**	**1.159**	1.313	**2.134**	**2.134**	2.745
	8	1	-2.297	**0.898**	1.038	2.336	**1.244**	1.693	8.651
		3	-2.297	**0.554**	0.606	1.463	**0.477**	0.569	3.336
		5	-2.297	**0.463**	0.473	1.051	**0.334**	0.350	1.789
		10	-2.297	0.384	**0.334**	0.782	0.217	**0.172**	0.954
0.25	2	1	-6.500	**3.807**	**3.807**	4.321	**22.129**	**22.129**	28.292
		3	-6.500	**2.153**	**2.153**	2.391	**7.274**	**7.274**	8.856
		5	-6.500	**1.715**	**1.715**	1.824	**4.584**	**4.584**	5.280
		10	-6.500	**1.316**	**1.316**	1.342	**2.642**	**2.642**	2.805
	8	1	-2.297	1.403	**1.335**	2.456	2.776	**2.717**	9.587
		3	-2.297	1.396	**1.186**	1.509	2.314	**1.876**	3.561
		5	-2.297	1.355	1.116	**1.114**	2.054	**1.559**	1.917
		10	-2.297	1.342	1.069	**0.783**	1.928	1.324	**0.999**
0.50	2	1	-6.500	**3.870**	**3.870**	4.012	**23.616**	**23.616**	25.838
		3	-6.500	2.565	2.565	**2.395**	10.048	10.048	**9.274**
		5	-6.500	2.139	2.139	**1.877**	6.991	6.991	**5.448**
		10	-6.500	1.796	1.796	**1.338**	4.700	4.700	**2.787**
	8	1	-2.297	2.403	**2.284**	2.465	6.923	**6.688**	9.499
		3	-2.297	2.367	2.184	**1.440**	5.985	5.317	**3.220**
		5	-2.297	2.374	2.178	**1.087**	5.858	5.073	**1.847**
		10	-2.297	2.404	2.212	**0.782**	5.900	5.070	**0.970**
0.75	2	1	-6.500	**4.138**	**4.138**	4.162	**26.753**	**26.753**	28.210
		3	-6.500	2.905	2.905	**2.386**	12.336	12.336	**9.081**
		5	-6.500	2.652	2.652	**1.869**	9.928	9.928	**5.476**
		10	-6.500	2.405	2.405	**1.330**	7.524	7.524	**2.768**
	8	1	-2.297	3.448	3.346	**2.581**	13.089	12.935	**10.172**
		3	-2.297	3.452	3.352	**1.371**	12.300	11.811	**2.899**
		5	-2.297	3.417	3.290	**1.099**	11.884	11.162	**1.912**
		10	-2.297	3.424	3.308	**0.766**	11.839	11.106	**0.915**
1.00	2	1	-6.500	4.373	4.373	**4.148**	29.442	29.442	**26.621**
		3	-6.500	3.568	3.568	**2.355**	17.556	17.556	**8.570**
		5	-6.500	3.370	3.370	**1.906**	15.018	15.018	**5.723**
		10	-6.500	3.305	3.305	**1.270**	12.749	12.749	**2.555**
	8	1	-2.297	4.473	4.419	**2.453**	21.073	21.098	**9.491**
		3	-2.297	4.445	4.399	**1.394**	20.124	19.900	**3.076**
		5	-2.297	4.446	4.389	**1.134**	19.983	19.587	**2.051**
		10	-2.297	4.486	4.426	**0.802**	20.246	19.771	**1.020**

Note. Results over 1000 repetitions. Subscript refers to: (1) TWFE estimator, (2) Wooldridge [2022] estimator, and (3) our estimator. The lowest value across estimators is in bold.

Table 4: MSE and Absolute Bias values - Poisson

ρ	T	M	ATT	$\lvert Bias_1 \rvert$	$\lvert Bias_2 \rvert$	$\lvert Bias_3 \rvert$	MSE_1	MSE_2	MSE_3
0.00	2	1	-26.780	**9.701**	**9.701**	11.905	**153.431**	**153.431**	239.845
		3	-26.780	**5.540**	**5.540**	6.740	**47.762**	**47.762**	73.687
		5	-26.780	**4.245**	**4.245**	5.032	**29.219**	**29.219**	40.018
		10	-26.780	**3.071**	**3.071**	3.689	**14.968**	**14.968**	21.449
	8	1	-30.865	9.057	**8.085**	27.674	121.582	**105.769**	1347.861
		3	-30.865	8.280	**4.863**	15.381	85.855	**36.796**	404.870
		5	-30.865	8.044	**3.526**	11.869	75.668	**19.522**	224.123
		10	-30.865	8.125	**2.536**	7.890	71.494	**9.967**	98.496
0.25	2	1	-26.780	**12.067**	**12.067**	12.275	**248.172**	**248.172**	272.296
		3	-26.780	7.698	7.698	**6.889**	93.040	93.040	**75.998**
		5	-26.780	6.601	6.601	**5.072**	66.846	66.846	**39.553**
		10	-26.780	5.781	5.781	**3.857**	47.955	47.955	**22.876**
	8	1	-30.865	34.556	30.313	**28.662**	1266.576	**1058.689**	1518.343
		3	-30.865	33.742	28.917	**14.216**	1161.546	882.219	**341.538**
		5	-30.865	33.993	28.983	**11.307**	1169.549	866.469	**204.729**
		10	-30.865	33.858	28.888	**8.140**	1153.220	847.927	**106.282**
0.50	2	1	-26.780	15.622	15.622	**11.861**	390.349	390.349	**231.468**
		3	-26.780	12.878	12.878	**6.984**	231.040	231.040	**78.199**
		5	-26.780	11.946	11.946	**5.106**	184.497	184.497	**42.616**
		10	-26.780	12.066	12.066	**3.646**	167.718	167.718	**20.788**
	8	1	-30.865	67.194	67.826	**27.440**	4603.887	4782.310	**1398.162**
		3	-30.865	66.939	67.699	**15.029**	4508.284	4638.895	**387.225**
		5	-30.865	66.675	67.477	**11.203**	4462.200	4587.222	**199.809**
		10	-30.865	66.688	67.328	**8.195**	4456.398	4551.466	**106.420**
0.75	2	1	-26.780	22.597	22.597	**12.162**	762.129	762.129	**244.327**
		3	-26.780	20.320	20.320	**6.716**	503.553	503.553	**72.748**
		5	-26.780	19.418	19.418	**5.204**	430.610	430.610	**43.640**
		10	-26.780	20.115	20.115	**3.599**	430.218	430.218	**20.391**
	8	1	-30.865	108.579	118.317	**26.869**	11902.528	14259.553	**1344.294**
		3	-30.865	108.185	117.445	**15.289**	11742.232	13877.573	**375.027**
		5	-30.865	108.328	117.403	**11.473**	11757.889	13835.222	**214.460**
		10	-30.865	108.571	118.043	**7.979**	11799.580	13960.016	**102.759**
1.00	2	1	-26.780	31.420	31.420	**11.927**	1297.693	1297.693	**249.215**
		3	-26.780	30.005	30.005	**6.688**	1008.487	1008.487	**71.590**
		5	-26.780	30.104	30.104	**5.017**	972.364	972.364	**39.530**
		10	-26.780	29.547	29.547	**3.566**	906.943	906.943	**19.970**
	8	1	-30.865	164.055	187.416	**27.132**	27086.351	35529.966	**1257.582**
		3	-30.865	163.392	186.002	**15.000**	26748.599	34712.731	**364.521**
		5	-30.865	163.930	186.676	**11.684**	26902.365	34921.680	**215.341**
		10	-30.865	163.562	186.035	**8.031**	26768.157	34646.567	**106.997**

Note. Results over 1000 repetitions. Subscript refers to: (1) TWFE estimator, (2) Wooldridge [2022] estimator, and (3) our estimator. The lowest value across estimators is in bold.

9 Conclusion

We establish identifying assumptions and estimation procedures for the ATT without interference in a DiD setting with staggered treatment adoption and spillovers. Aside from the canonical DiD assumptions, identification requires that once a unit receives treatment, it is no longer influenced by the spillover effect. This means the unit forfeits any spillovers it may have previously received and remains unaffected by spillovers from subsequently treated groups. This assumption, which is likely to hold in many contexts, unifies the multiple definitions of the ATT, simplifying policy evaluation and aligning with the definition of ATT under SUTVA.

To estimate the ATT, we extend the TWFE model approach of Wooldridge [2022] to account for spillovers in linear and non-linear settings. In the case of a balanced panel, our approach can be used to easily calculate the ATT's standard error. We then revisit Gonzalez-Navarro [2013], who studied the effects of installing an auto theft prevention device known as Lojack. Our correction leads to a slightly larger effect of the policy relative to the original contribution's specification.

Finally, our Monte Carlo analysis brings attention to the inherent bias-variance trade-off involved in addressing staggered treatment and especially spillovers. We compare three different estimators: the traditional TWFE estimator, which overlooks both staggered adoption and spillovers; the estimator of Wooldridge [2022], which considers staggered adoption but not spillovers; and our proposed estimator, which addresses both factors. Our estimator proves to be competitive in various scenarios.

References

Tobias Berg, Markus Reisinger, and Daniel Streitz. Spillover effects in empirical corporate finance. *Journal of Financial Economics*, 2021.

Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, 2024.

Kyle Butts. Difference-in-Differences Estimation with Spatial Spillovers. Papers 2105.03737, arXiv.org, June 2023. URL https://ideas.repec.org/p/arx/papers/2105.03737.html.

Brantly Callaway and Pedro H. C. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 2020.

Damian Clarke. Estimating Difference-in-Differences in the Presence of Spillovers. MPRA Paper 81604, University Library of Munich, Germany, September 2017.

Clément de Chaisemartin and Xavier D'Haultfoeuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96, 2020.

Clément de Chaisemartin and Xavier D'Haultfœuille. Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey. *The Econometrics Journal*, 2021.

Marco Gonzalez-Navarro. Deterrence and geographical externalities in auto theft. *American Economic Journal: Applied Economics*, 5(4):92–110, 2013.

Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 2021.

Kevin Han, Guillaume Basse, and Iavor Bojinov. Population interference in panel experiments. *Journal of Econometrics*, 238(1), 2024.

Martin Huber and Andreas Steinmayr. A Framework for Separating Individual-Level Treatment Effects From Spillover Effects. *Journal of Business & Economic Statistics*, 39(2):422–436, March 2021.

Robert Minton and Casey B Mulligan. Difference-in-differences in the marketplace. Technical report, National Bureau of Economic Research, 2024.

Jonathan Roth and Pedro HC Sant'Anna. When is parallel trends sensitive to functional form? *Econometrica*, 91(2):737–747, 2023.

Jonathan Roth, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe. What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, 2023.

Fredrik Sävje, Peter Aronow, and Michael Hudgens. Average treatment effects in the presence of unknown interference. *Annals of statistics*, 49(2), 2021.

Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 2020.

Gonzalo Vazquez-Bare. Identification and estimation of spillover effects in randomized experiments. *Journal of Econometrics*, 237(1), 2023.

Jeffrey M Wooldridge. Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators, 2022.

Jeffrey M Wooldridge. Simple approaches to nonlinear difference-in-differences with panel data. *The Econometrics Journal*, 26(3):C31–C66, 2023.

A Proofs

A.1 Proof of Theorem 1

Under Assumptions 2 and 3, for each group g at time t, we can express Y_{it} as

$$\begin{aligned} Y_{it} &= Y_{it}(d_i, \mathbf{d}_{-i}) \\ &= Y_{it}(0, \mathbf{0}_{-i}) + [Y_{it}(d_i, \mathbf{0}_{-i}) - Y_{it}(0, \mathbf{0}_{-i})] + [Y_{it}(d_i, \mathbf{d}_{-i}) - Y_{it}(d_i, \mathbf{0}_{-i})] \\ &= \alpha_i + \delta_t + [Y_{it}(d_i, \mathbf{0}_{-i}) - Y_{it}(0, \mathbf{0}_{-i})] + [Y_{it}(d_i, \mathbf{d}_{-i}) - Y_{it}(d_i, \mathbf{0}_{-i})] + \varepsilon_{it}, \end{aligned}$$

where the last equality follows from Remark 1, in which $\mathbb{E}(\varepsilon_{it}|G_i = g) = 0$ for every group g at time t. Define

$$\begin{aligned} \beta_{it} &= Y_{it}(d_i, \mathbf{0}_{-i}) - Y_{it}(0, \mathbf{0}_{-i}), \\ \gamma_{it} &= Y_{it}(d_i, \mathbf{d}_{-i}) - Y_{it}(d_i, \mathbf{0}_{-i}). \end{aligned}$$

We can then simplify the expression for Y_{it} as

$$Y_{it} = \alpha_i + \delta_t + \beta_{it} + \gamma_{it} + \varepsilon_{it}.$$

In this expression, the parameter of interest $ATT_0(g,t)$ for $t \geq g$ is given by

$$ATT_0(g,t) = \mathbb{E}(\beta_{it}|G_i = g),$$

and $AST(g,t)$ for $t \geq g$ is given by

$$AST(g,t) = \mathbb{E}(\gamma_{it}|G_i = g).$$

Using these expressions, for every group $g \in \mathcal{G}$ such that $2 \leq g < \infty$ and time $t \geq g$, we can write the expectation of Y_{it} as

$$\mathbb{E}(Y_{it}|G_i = g) = \mathbb{E}(\alpha_i|G_i = g) + \delta_t + ATT_0(g,t) + AST(g,t), \tag{12}$$

where we used $\mathbb{E}(\varepsilon_{it}|G_i = g) = 0$.

Now we show that $ATT_0(g,t)$ is identified if and only if $\delta_t + AST(g,t)$ is identified, for every group $g \in \mathcal{G}$ such that $2 \leq g < \infty$ and time $t \geq g$. First, suppose that $\delta_t + AST(g,t)$

is identified. Let d_0 be the identified value. Then we can rewrite Equation (12) as

$$\mathbb{E}(Y_{it}|G_i = g) = \mathbb{E}(\alpha_i|G_i = g) + d_0 + ATT_0(g,t). \tag{13}$$

Now we show that $\mathbb{E}(\alpha_i|G_i = g)$ is identified from the data at $t = 1$. Note first that, under the assumptions of Theorem 1, all units are untreated at $t = 1$. This implies that

$$Y_{i1} = Y_{i1}(d_i, \mathbf{d}_{-i}) = Y_{i1}(0, \mathbf{0}_{-i}) = \alpha_i + \delta_1 + \varepsilon_{it},$$

where the last equality follows from Assumption 3. Then it follows that

$$\mathbb{E}(Y_{i1}|G_i = g) = \mathbb{E}(\alpha_i + \delta_1 + \varepsilon_{it}|G_i = g) = \mathbb{E}(\alpha_i|G_i = g), \tag{14}$$

where $\delta_1 = 0$ and $\mathbb{E}(\varepsilon_{it}|G_i = g) = 0$ by Assumption 3. We can then rewrite Equation (13) as

$$ATT_0(g,t) = \mathbb{E}(Y_{it}|G_i = g) - d_0 - \mathbb{E}(Y_{i1}|G_i = g),$$

which shows that $ATT_0(g,t)$ is identified because $\mathbb{E}(Y_{it})$ and $\mathbb{E}(Y_{i1})$ are identifiable whenever $g \in \mathcal{G}$, i.e., whenever the group appears in the data.

Conversely, suppose that $ATT_0(g,t)$ is identified. Let b_0 be the identified value. Then we can rewrite Equation (12) as

$$\mathbb{E}(Y_{it}|G_i = g) = \mathbb{E}(\alpha_i|G_i = g) + \delta_t + b_0 + AST(g,t).$$

Using Equation (14), we can write

$$\delta_t + AST(g,t) = \mathbb{E}(Y_{it}|G_i = g) - b_0 - \mathbb{E}(Y_{i1}|G_i = g),$$

which shows that $\delta_t + AST(g,t)$ is identified. ■

A.2 Proof of Theorem 2

By Theorem 1, it suffices to show that $\delta_t + AST(g,t)$ is identified for every $t \geq 2$ under the assumptions of Theorem 2. Note first that Assumption 4 implies $AST(g,t) = 0$. In addition, Assumption 5 implies that the following quantity is identifiable for every $t \geq 2$:

$$\mathbb{E}(Y_{it} - Y_{i1}|G_i = \infty, H_i = 1) = \mathbb{E}(Y_{it}(0, \mathbf{d}_{-i}) - Y_{i1}(0, \mathbf{0}_{-i})|G_i = \infty, H_i = 1) = \delta_t,$$

where the last equality follows by Equation (5). This implies that δ_t is identified, which implies that $\delta_t + AST(g,t)$ is identified because $\delta_t + AST(g,t) = \delta_t + 0 = \delta_t$. In particular, it follows that

$$ATT_0(g,t) = \mathbb{E}(Y_{it}|G_i = g) - \mathbb{E}(Y_{i1}|G_i = g) - \mathbb{E}(Y_{it} - Y_{i1}|G_i = \infty, H_i = 1)$$

by the proof of Theorem 1. ■

A.3 Proof of Theorem 3

As in the proof of Theorem 1, under Assumptions 2 and 3, for each group g at time t, we can express Y_{it} as

$$\begin{aligned} Y_{it} &= Y_{it}(d_i, \mathbf{d}_{-i}) \\ &= Y_{it}(0, \mathbf{0}_{-i}) + [Y_{it}(d_i, \mathbf{0}_{-i}) - Y_{it}(0, \mathbf{0}_{-i})] + [Y_{it}(d_i, \mathbf{d}_{-i}) - Y_{it}(d_i, \mathbf{0}_{-i})] \\ &= \alpha_i + \delta_t + [Y_{it}(d_i, \mathbf{0}_{-i}) - Y_{it}(0, \mathbf{0}_{-i})] + [Y_{it}(d_i, \mathbf{d}_{-i}) - Y_{it}(d_i, \mathbf{0}_{-i})] + \varepsilon_{it}, \end{aligned}$$

where the last equality follows from Remark 1, in which $\mathbb{E}(\varepsilon_{it}|G_i = g) = 0$ for every g and t. Define

$$\begin{aligned} \beta_{it} &= Y_{it}(d_i, \mathbf{0}_{-i}) - Y_{it}(0, \mathbf{0}_{-i}), \\ \gamma_{it} &= Y_{it}(d_i, \mathbf{d}_{-i}) - Y_{it}(d_i, \mathbf{0}_{-i}). \end{aligned}$$

We can then simplify the expression for Y_{it} as

$$Y_{it} = \alpha_i + \delta_t + \beta_{it} + \gamma_{it} + \varepsilon_{it}.$$

For the group $G_i = \infty$, this expression simplifies to

$$Y_{it} = \alpha_i + \delta_t + \gamma_{it} + \varepsilon_{it}, \tag{15}$$

because $\beta_{it} = 0$. Now, by Assumption 5:

$$\begin{aligned} \mathbb{E}(Y_{it}|G_i = \infty, H_i = 1) &= \mathbb{E}(Y_{it}(0, \mathbf{d}_{-i})|G_i = \infty, H_i = 1) \\ &= \mathbb{E}(Y_{it}(0, \mathbf{0}_{-i})|G_i = \infty, H_i = 1) \\ &= \mathbb{E}(Y_{i1}(0, \mathbf{0}_{-i})|G_i = \infty, H_i = 1) + \delta_t. \end{aligned} \tag{16}$$

The first equality of Equation (16) and the definition of γ_{it} implies that

$$\mathbb{E}(\gamma_{it}|G_i = \infty, H_i = 1) = \mathbb{E}(Y_{it}(0, \mathbf{d}_{-i}) - Y_{it}(0, \mathbf{0}_{-i})|G_i = \infty, H_i = 1) = 0.$$

Building on this finding, Equation (15) and the second equality of Equation (16) implies that

$$
\begin{aligned}
&\mathbb{E}(\varepsilon_{it}|G_i = \infty, H_i = 1) \\
&= \mathbb{E}(Y_{it} - \alpha_i - \delta_t|G_i = \infty, H_i = 1) \\
&= \mathbb{E}(Y_{it} - \alpha_i|G_i = \infty, H_i = 1) - \delta_t \\
&= \mathbb{E}(Y_{it}(0, \mathbf{d}_{-i}) - Y_{i1}(0, \mathbf{0}_{-i})|G_i = \infty, H_i = 1) - \delta_t = 0.
\end{aligned}
$$

Therefore, it follows that $\mathbb{E}(\gamma_{it}|G_i = \infty, H_i = 1) = \mathbb{E}(\varepsilon_{it}|G_i = \infty, H_i = 1) = 0$. In addition, since $\mathbb{E}(\varepsilon_{it}|G_i = \infty) = 0$, it follows that

$$\mathbb{E}(\varepsilon_{it}|G_i = \infty, H_i = 0) = 0,$$

which then implies that $\mathbb{E}(\varepsilon_{it}|G_i = g, H_i = h) = 0$ for every (g, h) in the extended group label set $\tilde{\mathcal{G}}$. Consequently, we can express Y_{it} for any extended group $(G_i, H_i) = (g, h)$ at time t as

$$Y_{it} = \alpha_i + \delta_t + \beta_{it} + \gamma_{it} + \varepsilon_{it},$$

where $\mathbb{E}(\varepsilon_{it}|G_i = g, H_i = h) = 0$ for every $(g, h) \in \tilde{\mathcal{G}}$ and $\mathbb{E}(\gamma_{it}) = 0$ if $(g, h) = (\infty, 0)$.

Now, for every unit satisfying $(G_i, H_i) \in \tilde{\mathcal{G}}$, we can write

$$
\begin{aligned}
Y_{it} &= \alpha_i + \delta_t + \varepsilon_{it} && \text{for} \quad 1 \le t < q, \\
Y_{it} &= \alpha_i + \delta_t + \gamma_{it} + \varepsilon_{it} && \text{for} \quad q \le t < g, \\
Y_{it} &= \alpha_i + \delta_t + \beta_{it} + \varepsilon_{it} && \text{for} \quad g \le t \le T.
\end{aligned}
$$

These expressions are obtained by the following arguments:

- The first expression is obtained by recognizing that, when all units are untreated, neither the treatment effect nor the spillover effects are present, represented by $\beta_{it} = 0$ and $\gamma_{it} = 0$.
- The second expression is obtained by recognizing that, in the periods where some units are treated but units in group g are not yet treated, there is no treatment effect ($\beta_{it} = 0$), while spillover effects may occur, represented by γ_{it}.
- The third expression is obtained by recognizing that, in the periods where units in group g have been treated, the treatment effect may present, represented by β_{it}, but units are not subject to spillover effects by Assumption 4, represented by $\gamma_{it} = 0$.

We can combine these three expressions into one unified expression, encompassing every group $(G_i, H_i) \in \tilde{\mathcal{G}}$ at every period $1 \leq t \leq T$, as follows:

$$Y_{it} = \alpha_i + \delta_t + \sum_{t'=G_i}^{T} \beta_{it'} \mathbf{1}(t = t') + \sum_{t'=q}^{\min\{G_i-1,T\}} \gamma_{it'} \mathbf{1}(t = t') + \varepsilon_{it},$$

where $\sum_{t'=G_i}^{T}$ is considered a null summation if $G_i = \infty$. We can write this further as

$$Y_{it} = \alpha_i + \delta_t + \sum_{t'=G_i}^{T} \beta_{it'} \mathbf{1}(t = t') D_{it} + \sum_{t'=q}^{\min\{G_i-1,T\}} \gamma_{it'} \mathbf{1}(t = t') S_{it} + \varepsilon_{it},$$

since $D_{it} = 1$ for $t \geq G_i$ and $S_{it} = 1$ for $q \leq t < G_i$ according to their definitions. We can equivalently write this last expression as

$$\begin{aligned} Y_{it} &= \sum_{(g,h)\in\tilde{\mathcal{G}}} \mathbf{1}(G_i = g, H_i = h) \left(\alpha_i + \delta_t + \sum_{t'=g}^{T} \beta_{it'} \mathbf{1}(t = t') D_{it} + \sum_{t'=q}^{\min\{g-1,T\}} \gamma_{it'} \mathbf{1}(t = t') S_{it} + \varepsilon_{it} \right) \\ &= \alpha_i + \delta_t + \sum_{(g,h)\in\tilde{\mathcal{G}}} \sum_{t'=g}^{T} \beta_{it'} \mathbf{1}(G_i = g, H_i = h) \mathbf{1}(t = t') D_{it} \\ &\quad + \sum_{(g,h)\in\tilde{\mathcal{G}}} \sum_{t'=q}^{\min\{g-1,T\}} \gamma_{it'} \mathbf{1}(G_i = g, H_i = h) \mathbf{1}(t = t') S_{it} + \varepsilon_{it}. \end{aligned} \tag{17}$$

Now we proceed to prove the theorem. Note that Equation (7) is a pooled regression of the variables that encompass all groups $(g, h) \in \tilde{\mathcal{G}}$. Let $\mathbf{X}_{it}$ be the vector of regressors in Equation (7), namely indicators of (g, h), indicators of t, interactions between D_{it} and indicators of (g, h, t), and interactions between S_{it} and indicators of (g, h, t). Note that the regressors $(\mathbf{X}_{i1}, \dots, \mathbf{X}_{iT})$ identify the extended group label (g, h) and vice versa, because the interactions of D_{it} identifies the original group label $g \in \mathcal{G}$ and the interactions of S_{it} distinguishes $(\infty, 0)$ and $(\infty, 1)$. This implies that

$$\mathbb{E}(Y_{it} | \mathbf{X}_{i1}, \dots, \mathbf{X}_{iT}) = \mathbb{E}(Y_{it} | G_i = g(\mathbf{X}_{i1}, \dots, \mathbf{X}_{iT}), H_i = h(\mathbf{X}_{i1}, \dots, \mathbf{X}_{iT}))$$

where $g(\mathbf{X}_{i1}, \dots, \mathbf{X}_{iT})$ and $h(\mathbf{X}_{i1}, \dots, \mathbf{X}_{iT})$ are the extended group label identified by

$(\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iT})$. Then, by Equation (17):

$$\begin{aligned}\mathbb{E}(Y_{it}|\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iT}) &= \mathbb{E}(Y_{it}|G_i = g(\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iT}), H_i = h(\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iT}))\\ &= \mathbb{E}(\alpha_i|G_i = g, H_i = h) + \delta_t + \sum_{t'=g}^{T} \mathbb{E}(\beta_{it'}|G_i = g, H_i = h)\mathbf{1}(t = t')D_{it}\\ &\quad + \sum_{t'=q}^{\min\{g-1,T\}} \mathbb{E}(\gamma_{it'}|G_i = g, H_i = h)\mathbf{1}(t = t')S_{it}.\end{aligned}$$

In the case where units with $H_i = 1$ exist only within group $G_i = \infty$, H_i is degenerate with $H_i = 0$ for units with $G_i \neq \infty$. Therefore, for units with $G_i \neq \infty$, the above equation simplifies to:

$$\begin{aligned}\mathbb{E}(Y_{it}|\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iT}) = \mathbb{E}(\alpha_i|G_i = g) + \delta_t + \sum_{t'=g}^{T} \mathbb{E}(\beta_{it'}|G_i = g)\mathbf{1}(t = t')D_{it}\\ + \sum_{t'=q}^{\min\{g-1,T\}} \mathbb{E}(\gamma_{it'}|G_i = g)\mathbf{1}(t = t')S_{it}.\end{aligned}$$

This shows that, for units with $G_i \neq \infty$, the coefficient associated with $\mathbf{1}(G_i = g)\mathbf{1}(t = t')D_{it}$ equals to $\mathbb{E}(\beta_{it}|G_i = g)$. Then, by the definition of β_{it}:

$$\mathbb{E}(\beta_{it}|G_i = g) = \mathbb{E}(Y_{it}(d_i, \mathbf{0}_{-i}) - Y_{it}(0, \mathbf{0}_{-i})|G_i = g),$$

where the right-hand side is the definition of $ATT_0(g, t)$. ■

A.4 Proof of Corollary 1

Similarly to the proof of Theorem 1, for each group g at time t, we can express Y_{it} as

$$\begin{aligned}Y_{it} &= Y_{it}(d_i, \mathbf{d}_{-i})\\ &= Y_{it}(0, \mathbf{0}_{-i}) + [Y_{it}(d_i, \mathbf{0}_{-i}) - Y_{it}(0, \mathbf{0}_{-i})] + [Y_{it}(d_i, \mathbf{d}_{-i}) - Y_{it}(d_i, \mathbf{0}_{-i})].\end{aligned}$$

Then, under Equation (9), we can write the expectation of Y_{it} as

$$\mathbb{E}(Y_{it}|G_i = g) = \exp\{\alpha_g + \delta_t\} + ATT_0(g, t) + AST(g, t),$$

where

$$ATT_0(g, t) = \mathbb{E}(Y_{it}(d_i, \mathbf{0}_{-i}) - Y_{it}(0, \mathbf{0}_{-i})|G_i = g),$$

and

$$AST(g,t) = \mathbb{E}(Y_{it}(d_i, \mathbf{d}_{-i}) - Y_{it}(d_i, \mathbf{0}_{-i})|G_i = g).$$

Then, by replicating the arguments in Theorem 1 that starts from Equation (12), it is straightforward to show that $ATT_0(g,t)$ is identified if and only if $\exp\{\alpha_g+\delta_t\}+AST(g,t)$ is identified. ■

A.5 Proof of Corollary 2

By Corollary 1, it suffices to show that $\exp\{\alpha_g+\delta_t\}+AST(g,t)$ is identified. We proceed by separately identifying the three objects: α_g, δ_t, and $AST(g,t)$. First, Assumption 4 implies that $AST(g,t)=0$, identifying $AST(g,t)$. Second, for units such that $G_i=\infty$ and $H_i=1$, Equations (9) and (10) imply that

$$\begin{aligned}
&\ln \mathbb{E}(Y_{it}|G_i=\infty, H_i=1) - \ln \mathbb{E}(Y_{i1}|G_i=\infty, H_i=1) \\
&= \ln \mathbb{E}(Y_{it}(0,\mathbf{0}_{-i})|G_i=\infty, H_i=1) - \ln \mathbb{E}(Y_{i1}(0,\mathbf{0}_{-i})|G_i=\infty, H_i=1) \\
&= \delta_t,
\end{aligned}$$

which identifies $\exp\{\delta_t\}$. In addition, because all units are untreated at $t=1$ by the assumption, it follows that

$$\mathbb{E}(Y_{i1}|G_i=g) = \mathbb{E}(Y_{i1}(0,\mathbf{0}_{-i})|G_i=g) = \exp\{\alpha_g\}.$$

This implies that α_g is identified for every $g\in\mathcal{G}$, because Y_{i1} is identifiable whenever $g\in\mathcal{G}$, i.e., whenever the group is present in the data. Therefore, $ATT_0(g,t)$ is identified by Corollary 1. In particular,

$$ATT_0(g,t) = \mathbb{E}(Y_{it}|G_i=g) - \frac{\mathbb{E}(Y_{it}|G_i=\infty, H_i=1)}{\mathbb{E}(Y_{i1}|G_i=\infty, H_i=1)}\mathbb{E}(Y_{i1}|G_i=g).$$

■