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

Mario Fiorini * Wooyong Lee † Gregor Pfeifer ‡

March 7, 2025

Abstract

We establish identifying assumptions and estimation procedures for the ATT in a Difference-in-Differences model 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. Moreover, we broaden our framework to nonlinear cases, offering estimation of the ATT by Poisson, Probit, and Logit regressions. We apply our method to revisit a corresponding application from the crime literature. Monte Carlo simulations suggest that our estimator performs competitively.

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

*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 [2023] 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. Our work contributes to this area and links two active DiD literature strands.

The first one focuses on estimation issues under staggered adoption and heterogeneous treatment effects across units and time, e.g., Borusyak et al. [2024], de Chaisemartin and D'Haultfoeuille [2020], Callaway and Sant'Anna [2021], Goodman-Bacon [2021], Sun and Abraham [2021], and Wooldridge [2022, 2023]. This literature highlights 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 nonlinear DiD settings.

The second DiD strand studies the identification of average treatment effects in the presence of spillovers, e.g., Berg et al. [2021] Butts [2023], Clarke [2017], Hettinger et al. [2023], Huber and Steinmayr [2021] and Xu [2023].[1] These studies highlight two main challenges for identification of the ATT if the treatment also impacts units that are not formally treated. First, untreated units might no longer be valid controls. Proposed solutions mostly centre around ruling out spillovers for a given group of units, often based on some spatial or network distance, allowing the researcher to use this latter group as a control.[2] Alternatively, if sufficient information exists, one can parametrize how units are exposed to spillovers. Second, multiple definitions of the ATT exist 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 that allow

[1]Violations of the SUTVA assumptions are also discussed in the context of cross-section observational studies (e.g., Forastiere et al. [2021], Zigler et al. [2020]) and in the context of randomized controlled trials, (e.g., Aronow and Samii [2017], Han et al. [2024], Hudgens and Halloran [2008], Sävje et al. [2021], and Vazquez-Bare [2023]).

[2]This structure is sometimes referred to as partial interference.

for the identification of the ATT without interference, despite the presence of spillovers. Moreover, we focus on the more complex staggered treatment adoption, offering consistent estimation of the ATT under staggered treatment adoption and the consequent evolution of spillover effects.

Specifically regarding contributions, first, we establish the identifying assumptions for the ATT without interference given a staggered DiD setting in the presence of spillovers. We show that, in addition to the canonical i) treatment irreversibility, ii) no-anticipation, and iii) parallel trends assumption, identification is established under the assumption that once a unit receives treatment, it is no longer influenced by spillover effects. This means that the unit forfeits any spillovers it may have previously received and remains unaffected by spillovers from subsequently treated groups. This assumption also makes the multiple definitions of the ATT equivalent, because they are the same with or without spillovers, simplifying policy evaluation and joining with the definition of ATT under SUTVA. In Section 3, we provide two examples in which such a scenario applies and discuss its testable implications. We further 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. Moreover, we show that the average spillover effects on untreated units (ASUT) are also identified under the same set of assumptions.

Our second contribution regards estimation. We extend the fully interacted TWFE regression approach of Wooldridge [2022] to account for spillovers.[3] Our approach applies to both linear and nonlinear parallel trends, offering estimation of the ATT and the ASUT by, for example, Probit, Logit and Poisson regressions, broadening the range of cases to which our approach can be applied.[4] For our empirical application, we revisit Gonzalez-Navarro [2013], who studied the effects of installing a stolen vehicle recovery device. We apply our nonlinear approach to estimate the device's effect on car theft, which is a count variable. Our proposed nonlinear estimator suggests a larger ATT relative to the original contribution's specification, which does not account for staggered treatment adoption. In addition, our nonlinear estimator finds that the ASUT is positive, indicating that units without the device experience higher theft counts when other units have the device installed. In contrast, the linear model yields an ASUT estimate that is not well-defined, in the sense that it implies negative counterfactual theft counts, highlighting the empirical relevance of our nonlinear approach. Lastly, diagnostic tests suggest that our assumptions are likely to hold.

[3] This approach is numerically equivalent to the imputation approach of Borusyak et al. [2024] in balanced panels.

[4] Roth and Sant'Anna [2023] point out that the parallel trends assumption is sensitive to the functional form, suggesting that the linear parallel trends assumption may not hold for outcomes such as binary and count data.

These contributions underscore the key distinctions between our work and Butts [2023], which appears to be most closely related. First, we provide assumptions for identifying the ATT without interference, whereas Butts [2023] identifies the sum of direct and spillover effects. Furthermore, while Butts [2023] focuses on linear models, we extend our framework to both linear and nonlinear settings.

As a third and final contribution, we perform a Monte Carlo analysis, highlighting the bias-variance trade-off implicit in the correction for staggered treatment adoption 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 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 lays out the formal DiD setup with staggered treatment adoption, after which Section 3 illustrates our setup with two motivating examples. 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 nonlinear case, and Section 7 discusses a corresponding application. Section 8 provides Monte Carlo simulations, and Section 9 concludes.

2 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:

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

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 $Y_{it}(g)$, which is the outcome value when the own group label G_i is set to g. In the presence of spillover effects, this outcome is affected by the group labels of all units. We define the potential outcome by

$$Y_{it}(g_i, \mathbf{g}_{-i}),$$

where g_i is the group label of unit i and $\mathbf{g}_{-i}$ is the group labels of all units other than i in the population. In the case of finite population, $\mathbf{g}_{-i}$ will be a $(N-1)$ vector, and in the case of infinite population, $\mathbf{g}_{-i}$ will be a mapping from an index to $\mathcal{G}$. This notation does not impose any restrictions on the structure of the spillover effects. An alternative notation used in the literature (e.g., Vazquez-Bare, 2023 and Butts, 2023) is

$$Y_{it}(g_i, h_i(\mathbf{g}_{-i})),$$

where $h_i(\cdot)$ is a deterministic mapping from $\mathbf{g}_{-i}$ to a non-negative real number (or vector) representing the spillover intensity. Hence, $Y_{it}(g_i, \mathbf{g}_{-i})$ can be interpreted as a generalization of $Y_{it}(g_i, h_i(\mathbf{g}_{-i}))$.

Similarly to the case of SUTVA, we define treatment effects by comparing the observed potential outcome against various counterfactual outcomes. Let G_i and $\mathbf{G}_{-i}$ be the observed group labels for unit i and the other units in the data, respectively, and define ∞ and ∞_{-i} as the group labels where unit i and the other units are all in group ∞, respectively. The following four types of potential outcomes will be relevant to our discussion.

- $Y_{it}(G_i, \mathbf{G}_{-i})$ represents the outcome where unit i is treated according to its observed group label, and other units are treated according to their observed group labels. This potential outcome is observed, i.e., $Y_{it} = Y_{it}(G_i, \mathbf{G}_{-i})$.

- $Y_{it}(G_i, \infty_{-i})$ represents the outcome where unit i is treated according to its observed group label, but all the other units are untreated.

- $Y_{it}(\infty, \mathbf{G}_{-i})$ represents the outcome where unit i is untreated, but other units are treated according to their observed group labels.
- $Y_{it}(\infty, \infty_{-i})$ represents the 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 $q \equiv \min\{t \mid t \in \mathcal{G}\}$ be the first period that any unit enters treatment.

Assumption 2 (no anticipation)**.** *For every i and for each $\mathbf{g}_{-i} \in \{\mathbf{G}_{-i}, \infty_{-i}\}$:*

$$\begin{aligned} Y_{it}(G_i, \mathbf{g}_{-i}) &= Y_{it}(\infty, \infty_{-i}) \quad \textit{for every } t < q, \textit{ and} \\ Y_{it}(G_i, \mathbf{g}_{-i}) &= Y_{it}(\infty, \mathbf{g}_{-i}) \quad \textit{for every } q \leq t < G_i. \end{aligned}$$

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 it to nonlinear parallel trends.

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

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

Remark 1. Assumption 3 is equivalent to (see the Appendix for the proof)

$$\mathbb{E}(Y_{it}(\infty, \infty_{-i})|G_i = g) = \mathbb{E}(\alpha_i|G_i = g) + \delta_t,$$

which is more commonly written as

$$Y_{it}(\infty, \infty_{-i}) = \alpha_i + \delta_t + \varepsilon_{it}, \tag{2}$$

where $\mathbb{E}(\varepsilon_{it}|G_i = g) = 0$ for all $g \in \mathcal{G}$. We refer to δ_t as the common time effect.

In the presence of spillover effects, multiple definitions of the ATT arise (Huber and Steinmayr [2021], Hudgens and Halloran [2008], Sävje et al. [2021]). We first introduce the ATT without interference (Huber and Steinmayr [2021],Vazquez-Bare [2023]):

$$ATT_0(g, t) \equiv \mathbb{E}(Y_{it}(G_i, \infty_{-i}) - Y_{it}(\infty, \infty_{-i})|G_i = g).$$

This definition of ATT 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. Alternatively, using the notation $Y_{it}(g_i, h_i(\mathbf{g}_{-i}))$, $ATT_0(g,t)$ is interpreted as the ATT when h_i is set to be zero for all units in group g. 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 [2021]).[5]

We can also consider an alternative definition of the ATT that incorporates the spillover effects from other treated units, namely the units with group labels $g \leq t$:

$$ATT_S(g,t) \equiv \mathbb{E}(Y_{it}(G_i, \mathbf{G}_{-i}) - Y_{it}(\infty, \infty_{-i})|G_i = g).$$

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

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

Lastly, we define the corresponding average spillover effect for untreated units, which we refer to as the average spillover effect on the untreated:

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

3 Intuition and Motivating Examples

To illustrate our setting, consider a panel data of six units ($i \in \{a, a', b, b', z, z'\}$) observed over three periods ($t = 1, 2, 3$), where units a, a' are treated in period 2, units b, b' are treated in period 3, and units z, z' are never-treated. The different effects of units' treatments on their outcomes are illustrated in Figure 1, focusing on periods 2 and 3. Solid arrows indicate the direct effect of a unit's own treatment, namely the integrand of ATT_0 (that is, $Y_{it}(G_i, \infty_{-i}) - Y_{it}(\infty, \infty_{-i})$). Dotted arrows indicate the spillovers on the treated, i.e., the influence of unit j's treatment on unit i's outcome while unit i is treated, which add up to the integrand of AST (that is, $Y_{it}(G_i, \mathbf{G}_{-i}) - Y_{it}(G_i, \infty_{-i})$). The dashed arrows represent the spillovers on the untreated, i.e., the influence of unit j's treatment on unit i's outcome while unit i is untreated, which add up to the integrand of $ASUT$ (that is, $Y_{it}(\infty, \mathbf{G}_{-i}) - Y_{it}(\infty, \infty_{-i})$). Under SUTVA, only the direct effects represented by solid lines would exist.[6]

We now introduce two alternative scenarios (so-called Examples 1 and 2) involving

[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.

[6] For brevity of illustration, Figure 1 omits dynamic treatment effects. For instance, there are no arrows linking the treatment in period 2 to the outcomes in period 3.

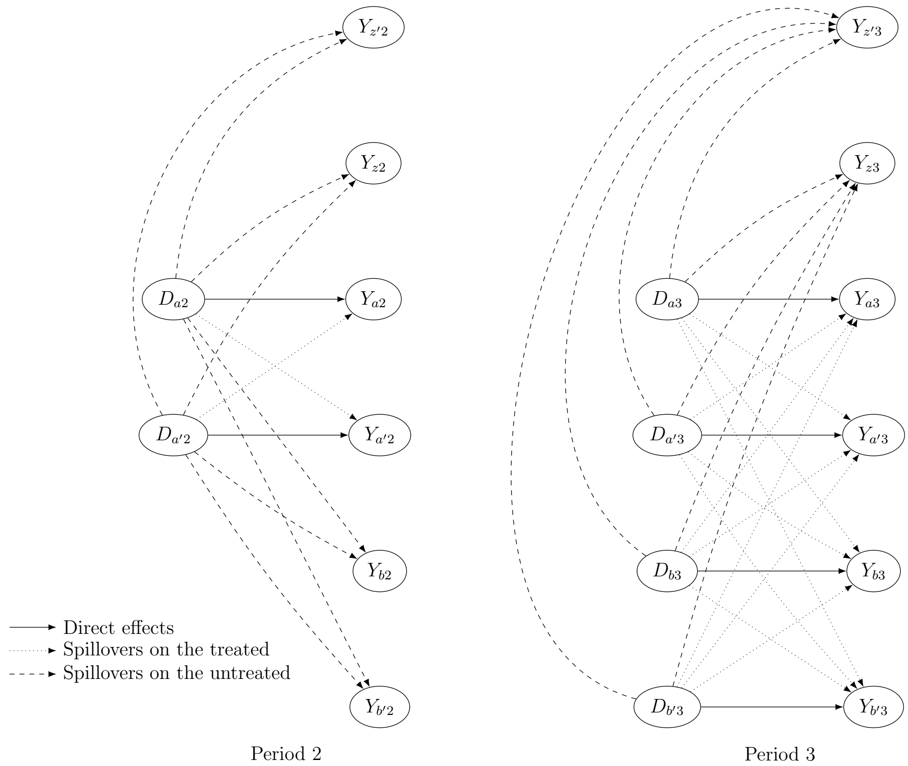


Figure 1: Illustration of the potential treatment and spillover paths.

spillover effects, which we will re-visit throughout the paper to motivate our assumptions and empirical application. In Example 1, the spillover is in the form of a diffusion effect, meaning that the direct effect and the spillover effects have the same sign. In Example 2, the spillover is in the form of displacement, where the direct effect and the spillover effects 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' 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.

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.

There are two key challenges arising from these settings. First, untreated units might be affected by spillovers. Changes in the outcomes of untreated units over time would then be a combination of time trends and spillover effects. Especially in the case of displacement with staggered treatment adoption, spillover effects may intensify over time as more and more treated units spill on an increasingly narrower pool of untreated units. For instance, in Example 2, car theft could concentrate on only a few untreated states and models, causing large bias if used to estimate the time effects. Second, even if time fixed effects could be identified, separating the direct and spillover effects remains infeasible. However, isolating the direct effect is often of interest. For instance, when a unit decides whether to participate in a policy, its decision likely hinges on the direct effect, as spillovers from other units are beyond its control. Understanding the direct effect also provides insight into the mechanisms driving overall outcomes, which is essential for refining policies and optimizing resource allocation. Moreover, the sum of direct and spillover effects might have limited external validity, as spillovers depend on the specific treatment histories of other units, which may differ in alternative contexts. Policymakers and researchers concerned with scalability or transferability can benefit from isolating the direct effect to evaluate the intervention's core efficacy across diverse scenarios.

Figure 2 visualizes our additional key assumptions, which will be formally introduced in Section 4, that resolve these issues and allow for identification of the direct effect. These are that units are not influenced by spillovers once treated, and that a subset of never-treated units remains unaffected by spillovers as well. Consequently, in this figure, dotted arrows and arrows to unit z' no longer appear. 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 the 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

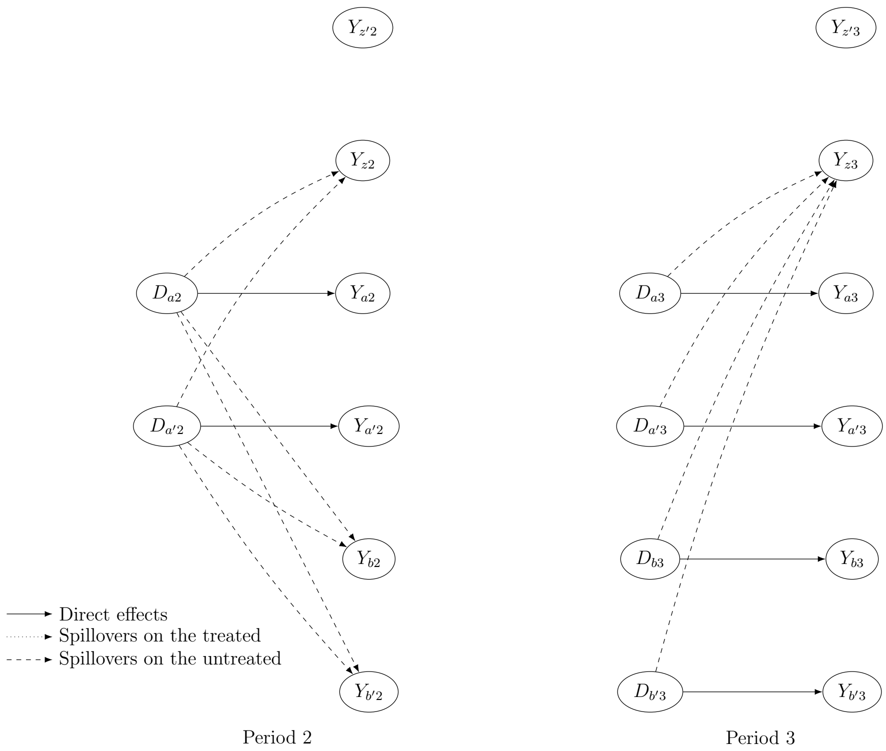


Figure 2: DAG under the key identification assumptions. There are no longer dotted arrows to treated outcomes and no arrows to $Y_{z'2}$ and $Y_{z'3}$.

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.[7]

[7]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. Even though this scenario might not be totally implausible, thieves would probably rather shift their focus to other, less protected assets or leave the criminal market entirely. For example, Cornish

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 $t \geq g$:

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

The intuition for Theorem 1 is that since $\mathbb{E}(\alpha_i|G_i = g)$ 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 $t \geq g$,*

$$\mathbb{E}(Y_{it}(G_i, \mathbf{G}_{-i})|G_i = g) = \mathbb{E}(Y_{it}(G_i, \infty_{-i})|G_i = g). \tag{3}$$

Remark 2. If the researcher is interested in $ATT_0(g,t)$ for a particular g, it is sufficient that Equation (3) holds for that particular g only. In this paper, we assume that Equation (3) holds for every g, since we will discuss aggregation of the $ATT_0(g,t)$ across g in

and Clarke [2017] point out that fitting steering column locks to all cars in West Germany during the 1960s resulted in a 60% reduction in car thefts without displacement, whereas introducing the locks solely on new cars in Great Britain merely displaced theft to older, unprotected vehicles. Moreover, 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.

later sections, which requires that Equation (3) holds for every g that is involved in the aggregation.

Remark 3. Assumption 4 is generally not testable. However, if the researcher has prior knowledge about the sign of $ATT_0(g,t)$ and the possible sign of $AST(g,t)$ (if it exists), and if these signs are opposite, they can check whether the estimated sign of $ATT_0(g,t)$ is as expected. For example, if prior knowledge suggests that $ATT_0(g,t) < 0$ and $AST(g,t) \geq 0$, one can check whether the estimated $ATT_0(g,t)$ is negative, as otherwise it could imply that $AST(g,t) > 0$.

Assumption 4 holds if $Y_{it}(G_i, \mathbf{G}_{-i}) = Y_{it}(G_i, \infty_{-i})$, 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 3, 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, which replaces Assumption 3.

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:*

$$\mathbb{E}(Y_{it}(\infty, \infty_{-i}) - Y_{i1}(\infty, \infty_{-i})|G_i = g) = \mathbb{E}(Y_{it}(\infty, \mathbf{G}_{-i}) - Y_{i1}(\infty, \infty_{-i})|G_i = \infty, H_i = 1) \tag{4}$$

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

Remark 4. Equation (4) can be interpreted as two separate equations:

$$\mathbb{E}(Y_{it}(\infty, \mathbf{G}_{-i})|G_i = \infty, H_i = 1) = \mathbb{E}(Y_{it}(\infty, \infty_{-i})|G_i = \infty, H_i = 1), \tag{5}$$

and

$$\begin{aligned} &\mathbb{E}(Y_{it}(\infty, \infty_{-i}) - Y_{i1}(\infty, \infty_{-i})|G_i = g) \\ &\qquad = \mathbb{E}(Y_{it}(\infty, \infty_{-i}) - Y_{i1}(\infty, \infty_{-i})|G_i = \infty, H_i = 1), \end{aligned} \tag{6}$$

where Equation (5) states that there exists a subgroup of never-treated units that are not affected by spillovers, and Equation (6) states that parallel trends apply to said subgroup. These two equalities have testable implications, which means that Assumption 5 can be tested with data. We will discuss the details of these tests in the next section.

Remark 5. Assumption 5 can be written equivalently as, for every $g \in \mathcal{G}$:

$$\mathbb{E}(Y_{it}(\infty, \infty_{-i})|G_i = g) = \mathbb{E}(\alpha_i|G_i = g) + \delta_t,$$

where $\alpha_i = Y_{i1}(\infty, \infty_{-i})$, $\delta_t = \mathbb{E}(Y_{it}(\infty, \mathbf{G}_{-i}) - Y_{i1}(\infty, \infty_{-i})|G_i = \infty, H_i = 1)$.

Remark 6. In practice, the researcher may take a conservative approach by setting $H_i = 1$ only for units strongly believed to satisfy Equation (4), and $H_i = 0$ otherwise.

Remark 7. 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:

$$\mathbb{E}(Y_{it}(\infty, \infty_{-i}) - Y_{i1}(\infty, \infty_{-i})|G_i = g) = \mathbb{E}(Y_{it}(\infty, \mathbf{G}_{-i}) - Y_{i1}(\infty, \infty_{-i})|G_i = g', H_i = 1). \tag{7}$$

for some g' in a subset of $\mathcal{G}$.

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 6). 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 7). In this case, the control group is not fixed but shrinks over t as more states adopt treatment.

Extending beyond Remark 7, there may be cases where different groups g' should be used for the parallel trend between period 1 and t, i.e., there exists $(g_2, \ldots, g_t) \in \mathcal{G}$ such that:

$$\begin{aligned}&\mathbb{E}(Y_{it}(\infty, \infty_{-i}) - Y_{i,t-1}(\infty, \infty_{-i})|G_i = g)\\ &\qquad = \mathbb{E}(Y_{it}(\infty, \mathbf{G}_{-i}) - Y_{i,t-1}(\infty, \infty_{-i})|G_i = g_t, H_i = 1)\end{aligned}$$

for each t and for all g. While this case also allows for identification and consistent estimation of ATT_0, the estimation procedure that will be introduced later becomes more complicated than under Assumption 5 or Remark 7, especially if the pattern of g_t is correlated with the potential outcomes. Since one of our focuses is to provide an estimation procedure that is straightforward to implement, we do not extend our setting to this case.

We conclude this section by showing that $ATT_0(g, t)$ is identified under these two assumptions, and that $ASUT(g, t)$ is also identified under the same set of assumptions.

Theorem 2. *Suppose that Assumptions 1, 2, 4 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,*

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

In addition, under the same set of assumptions, $ASUT(g,t)$ is identified for every $g \in \mathcal{G}$ and $q \leq t < g$. In particular,

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

Remark 8. When there are multiple periods where all units are untreated, i.e., when $q > 2$, Assumption 5 can be relaxed to:

$$\begin{aligned}&\mathbb{E}(Y_{it}(\infty, \infty_{-i}) - Y_{i,q-1}(\infty, \infty_{-i})|G_i = g)\\ &\qquad = \mathbb{E}(Y_{it}(\infty, \infty_{-i}) - Y_{i,q-1}(\infty, \infty_{-i})|G_i = \infty, H_i = 1),\end{aligned}$$

where $q-1$ is the last period in which all units are untreated. Note that, in the absence of spillovers, it is sufficient to assume

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

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

5 Estimation and Inference

In this section, we discuss estimation and inference of $ATT_0(g,t)$ and $ASUT(g,t)$. 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 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:

$$\begin{aligned} Y_{it} = \alpha_{G_i H_i} + \delta_t &+ \sum_{(g',h') \in \tilde{\mathcal{G}}} \sum_{t'} \beta_{g'h't'} \cdot \mathbf{1}((G_i, H_i, t) = (g', h', t')) \cdot D_{it} \\ &+ \sum_{(g',h') \in \tilde{\mathcal{G}}} \sum_{t'} \gamma_{g'h't'} \cdot \mathbf{1}((G_i, H_i, t) = (g', h', t')) \cdot S_{it} + \varepsilon_{it}. \end{aligned} \tag{8}$$

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

$$\begin{aligned} Y_{it} = \alpha_{G_i H_i} + \delta_t + \sum_{g' \in \mathcal{G} \setminus \{\infty\}} \sum_{t'=g'}^{T} \beta_{g't'} \cdot \mathbf{1}((G_i, t) = (g', t')) \cdot D_{it} \\ + \sum_{(g',h') \in \tilde{\mathcal{G}} \setminus \{(\infty,1)\}} \sum_{t'=q}^{\min\{g'-1,T\}} \gamma_{g'h't'} \cdot \mathbf{1}((G_i, H_i, t) = (g', h', t')) \cdot S_{it} + \varepsilon_{it}. \end{aligned} \tag{9}$$

Note that Equation (9) involves the *group*-level coefficients $(\alpha_{gh}, \beta_{gt})$, as opposed to the *unit*-level coefficients (α_i, β_{it}).[8] The following result shows that, despite this simplification, the population regression of Equation (9) correctly identifies $ATT_0(g, t)$ and $ASUT(g, t)$.

Theorem 3. *Suppose that the assumptions of Theorem 2 hold. Consider the population regression of Equation (9). Then*

$$\begin{aligned} \beta_{gt} &= \mathbb{E}(Y_{it}(G_i, \infty_{-i}) - Y_{it}(\infty, \infty_{-i}) | G_i = g), \\ \gamma_{g,0,t} &= \mathbb{E}(Y_{it}(\infty, \mathbf{G}_{-i}) - Y_{it}(\infty, \infty_{-i}) | G_i = g, H_i = 0). \end{aligned}$$

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

Remark 9. Theorem 3 implies that $\beta_{gt} = ATT_0(g,t)$ and $\gamma_{g,0,t} = ASUT(g,t)$ for $2 \leq g < \infty$. In addition, $\gamma_{\infty,0,t}$ equals to $ASUT$ for the subgroup $(G_i = \infty, H_i = 0)$. Then, $ASUT(\infty, t)$ can be calculated as

$$ASUT(\infty, t) = w_{\infty,0,t} \cdot \gamma_{\infty,0,t} + w_{\infty,1,t} \cdot 0 = w_{\infty,0,t}\gamma_{\infty,0,t},$$

where $w_{\infty,0,t}$ and $w_{\infty,1,t}$ are the weights assigned to the subgroups $(G_i = \infty, H_i = 0)$ and $(G_i = \infty, H_i = 1)$, respectively. Note that the spillover effect on the group $(G_i = \infty, H_i = 1)$ is zero.

Theorem 3 yields a simple procedure for estimation and inference of $ATT_0(g,t)$ (or $ASUT(g,t)$), since $\widehat{\beta}_{gt}$ (or $\widehat{\gamma}_{g,0,t}$) and its standard error can be easily obtained by implementing Equation (9) using standard statistical software. Estimation and inference of an aggregate ATT_0 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, e.g., via e(V) in Stata.

It is worth noting that Equation (9) is numerically equivalent to the following extension of the imputation 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
$$\widehat{Y}_{it}(\infty, \infty_{-i}) = \widehat{\alpha}_i + \widehat{\delta}_t.$$

3. For each unit i treated at time t, compute
$$\widehat{\beta}_{it}^{imp} \equiv Y_{it} - \widehat{Y}_{it}(\infty, \infty_{-i}),$$
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 (9), 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. In comparison, the use of group-level coefficients in Equation (9) eases the calculation of standard errors, which can be done conveniently using standard statistical software.

Our procedure also offers a diagnostic test for validity of Assumption 5. As discussed in Remark 4, Assumption 5 can be interpreted as two separate equations, given by Equations (5) and (6), which have testable implications. First, Equation (5) leads to the following placebo test:

1. Divide the group $(G_i = \infty, H_i = 1)$ into two subgroups based on how Assumption 5 is likely to apply, and label one of them as $(G_i = \infty, H_i = 2)$. For instance, in Example 2, we could choose states in $G_i = \infty$ that are closest to the treated states and label them as $(G_i = \infty, H_i = 2)$.
2. For this group $(G_i = \infty, H_i = 2)$, define $S_{it} = 1$ for all $t \geq q$ and 0 otherwise.
3. Run the regression specified in Equation (8).
4. If Equation (5) holds, then $\gamma_{\infty,2,t}$ should be zero for all $t \geq q$.

This placebo test splits the group $(G_i = \infty, H_i = 1)$ into two subgroups and considers one group as if it were affected by spillover effects, which is false if Equation (5) is satisfied. Therefore, under Equation (5), the $\gamma_{\infty,2,t}$ coefficients must be zero.

Second, Equation (6) can be tested using the standard pre-trend test:

1. Select an integer k such that $1 \leq k \leq q-1$, and redefine S_{it} as:

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

2. Run the regression specified in Equation (8).
3. If Equation (6) holds, then $\gamma_{g,0,t}$ should be zero for $q-k \leq t < q$.

This test is similar to the standard pre-trend test, but as discussed in Remark 8, the pre-treatment periods in the presence of spillovers are characterized by $t < q$, periods where all groups are untreated. Under SUTVA, the pre-treatment periods are defined according to each group's own treatment state instead. In empirical applications, it is common to report the average of the $\gamma_{g,0,t}$s instead of the individual ones, for instance, in the event-study plot. Following this practice, we report the test results about this average in our application in Section 7.

Next, we consider the case where there are also units with $H_i = 1$ under $G_i \neq \infty$. 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 and $ASUT$ for these subgroups as follows:

$$\begin{aligned} ATT_0(g,h,t) &= \mathbb{E}(Y_{it}(G_i, \infty_{-i}) - Y_{it}(\infty, \infty_{-i}) | G_i = g, H_i = h), \\ ASUT(g,h,t) &= \mathbb{E}(Y_{it}(\infty, \mathbf{G}_{-i}) - Y_{it}(\infty, \infty_{-i}) | G_i = g, H_i = h). \end{aligned}$$

We introduce this new definition because the ATT_0 (or $ASUT$) for the subgroups $(G_i = g, H_i = 0)$ and $(G_i = g, H_i = 1)$ may differ.[9] The aggregate ATT_0 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 (8), and the coefficient β_{ght} equals to $\beta_{ght} = ATT_0(g,h,t)$ and γ_{ght} equals to $ASUT(g,h,t)$.

Lastly, if the data is an unbalanced panel, the regression in Equation (9) 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 (9) is asymptotically exact.

6 Extension to Nonlinear DiD Models

In this section, we extend our approach to the case where Y_{it} is a limited dependent variable, such as binary or count variables, for which the linear parallel trends assumption may 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 nonlinear models, following Wooldridge [2023]. Let $F : \mathbb{R} \mapsto \mathbb{R}$ be a monotonically increasing function.

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

$$\begin{aligned} &F\left(\mathbb{E}(Y_{it}(\infty, \infty_{-i}) | G_i = g)\right) - F\left(\mathbb{E}(Y_{i1}(\infty, \infty_{-i}) | G_i = g)\right) \\ &= F\left(\mathbb{E}(Y_{it}(\infty, \infty_{-i}) | G_i = \infty)\right) - F\left(\mathbb{E}(Y_{i1}(\infty, \infty_{-i}) | G_i = \infty)\right) \end{aligned}. \tag{10}$$

[9] 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 $ATT(g,t) = ATT(g,0,t)$.

Remark 10. Common choices of F are canonical link functions, corresponding to nonlinear regressions that are frequently used in empirical applications. For instance, if Y_{it} is binary, F can be set as the inverse of the Gaussian CDF, corresponding to probit regression, or the inverse of the Logistic function, corresponding to logit regression. If Y_{it} is a count variable, F can be set as the log function, which corresponds to Poisson regression.

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 nonlinear models, replacing Assumption 3'.

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:*

$$\begin{aligned} &F\left(\mathbb{E}(Y_{it}(\infty,\infty_{-i})|G_i = g)\right) - F\left(\mathbb{E}(Y_{i1}(\infty,\infty_{-i})|G_i = g)\right) \\ &= F\left(\mathbb{E}(Y_{it}(\infty,\mathbf{G}_{-i})|G_i = \infty, H_i = 1)\right) - F\left(\mathbb{E}(Y_{i1}(\infty,\infty_{-i})|G_i = \infty, H_i = 1)\right) \end{aligned} \tag{11}$$

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

Remark 11. Assumptions 3 and 5 are special cases of Assumptions 3' and 5' where F is an identity function, i.e., $F(x) = x$.

Remark 12. Assumption 5' can be written equivalently as, for every $g \in \mathcal{G}$:

$$\mathbb{E}(Y_{it}(\infty,\infty_{-i})|G_i = g) = F^{-1}(\alpha_g + \delta_t), \tag{12}$$

where

$$\begin{aligned} \alpha_g &= F(\mathbb{E}(Y_{i1}(\infty,\infty_{-i})|G_i = g)), \\ \delta_t &= F(\mathbb{E}(Y_{it}(\infty,\mathbf{G}_{-i})|G_i = \infty, H_i = 1)) - F(\mathbb{E}(Y_{i1}(\infty,\infty_{-i})|G_i = \infty, H_i = 1)). \end{aligned}$$

Corollary 1. *Suppose that Assumptions 1, 2, and 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 $F^{-1}(\alpha_g + \delta_t) + AST(g,t)$ is identified.*

Corollary 2. *Suppose that Assumptions 1, 2, 4, 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,*

$$\begin{aligned} ATT_0(g,t) = \mathbb{E}(Y_{it}|G_i = g) - F^{-1}\Big[&F(\mathbb{E}(Y_{i1}|G_i = g)) + \\ &F(\mathbb{E}(Y_{it}|G_i = \infty, H_i = 1)) - F(\mathbb{E}(Y_{i1}|G_i = \infty, H_i = 1))\Big]. \end{aligned}$$

In addition, under the same set of assumptions, $ASUT(g,t)$ is identified for every $g \in \mathcal{G}$ and $q \leq t < g$. In particular,

$$ASUT(g,t) = \mathbb{E}(Y_{it}|G_i = g) - F^{-1}\Big[F(\mathbb{E}(Y_{i1}|G_i = g)) + F(\mathbb{E}(Y_{it}|G_i = \infty, H_i = 1)) - F(\mathbb{E}(Y_{i1}|G_i = \infty, H_i = 1))\Big].$$

Remark 13. For example, when F is the log function, $ATT_0(g,t)$ is identified by

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

for $t \geq g$.

Estimation and inference are also similar to Section 5. To be concrete, we discuss details on estimation and inference in the case where F is a log function, corresponding to Poisson regression for count data, which we use in our application in the next section. The estimation and inference procedures can be similarly applied for any canonical link function F.

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:

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

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

Consider 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:

$$\begin{aligned}\ln \mathbb{E}(Y_{it}|\mathbf{X}_{it}) =&\alpha_{G_iH_i} + \delta_t + \sum_{g'\in\mathcal{G}\backslash\{\infty\}} \sum_{t'=g'}^{T} \beta_{g't'} \cdot \mathbf{1}((G_i,t)=(g',t')) \cdot D_{it} \\ &+ \sum_{(g',h')\in\tilde{\mathcal{G}}\backslash\{(\infty,0)\}} \sum_{t'=q}^{\min\{g'-1,T\}} \gamma_{g'h't'} \cdot \mathbf{1}((G_i,H_i,t)=(g',h',t')) \cdot S_{it},\end{aligned} \tag{13}$$

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. For $2 \leq g < \infty$ and $t \geq g$, estimate $ATT_0(g,t)$ by:

$$\widehat{ATT}_0(g,t) = \exp\{\widehat{\alpha}_{g,0} + \widehat{\delta}_t + \widehat{\beta}_{gt}\} - \exp\{\widehat{\alpha}_{g,0} + \widehat{\delta}_t\},$$

or estimate $ATTP_0(g,t)$ by $\widehat{ATTP}_0(g,t) = \exp\{\widehat{\beta}_{gt}\} - 1$. Similarly, for $2 \leq g < \infty$ and $q \leq t < g$, estimate $ASUT(g,t)$ by

$$\widehat{ASUT}(g,t) = \exp\{\widehat{\alpha}_{g,0} + \widehat{\delta}_t + \widehat{\gamma}_{g,0,t}\} - \exp\{\widehat{\alpha}_{g,0} + \widehat{\delta}_t\}.$$

For a generic canonical link function F, $\widehat{ATT}_0(g,t) = F^{-1}(\widehat{\alpha}_{g,0} + \widehat{\delta}_t + \widehat{\beta}_{gt}) - F^{-1}(\widehat{\alpha}_{g,0} + \widehat{\delta}_t)$, where $(\widehat{\alpha}_{g,0}, \widehat{\beta}_{gt}, \widehat{\delta}_t)$ are estimates from the generalized linear model corresponding to F, such as probit or logit. A similar argument applies to $\widehat{ASUT}(g,t)$.

The validity of the population regression of Equation (13) 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 (13) using any standard statistical software that runs Poisson regressions.

Note that most statistical software 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 $Y_{it}(\infty, \infty_{-i})$ follows a Poisson distribution, as opposed to only specifying its mean as in Assumption 5', 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.[10] 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

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

In practice, $\widehat{\mathrm{Var}}(\hat{\theta})$ is available in most software packages, e.g., via vce(robust) in Stata. This variance can then be used to compute the standard errors of the ATT_0 and $ATTP_0$ estimates via the delta method. For a generic canonical link function F, the procedure applies similarly, where the score and Hessian functions are computed according to the nonlinear regression that corresponds to F. The resulting variance matrix will be available in most statistical software, e.g., via vce(robust) in Stata.

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

[10] 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 $t \geq 2001$.

To apply our method, we need to check the credibility of Assumptions 4 and 5'. Assumption 4 states 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' states 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.[11] Later in this section, we also show results for the diagnostic tests mentioned in Remarks 3 and 4, which support the credibility of Assumptions 4 and 5'.

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:

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

This aligns with Assumption 5, where the combination (m, s, a) plays the role of i. Second, we consider Poisson parallel trends:

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

which aligns with Assumption 5'. 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

[11] 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	-6.084	2.637	-63.5%	-5.525	1.706	-62.7%
ATT_0^0	-3.978	1.988	-45.9%	-3.764	1.083	-44.9%
ATT_0^1	-6.786	2.803	-77.5%	-6.162	1.934	-75.8%
ATT_0^2	-16.259	8.201	-79.3%	-13.356	5.241	-85.3%

Table 1: Estimates of the aggregate ATT_0. The standard errors are clustered at the state level. The "Reduction" column stands for the reduction rate, which is calculated using the formula for computing $ATTP_0$.

estimate the following aggregate ATT_0:

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

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

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

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

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 state level. The analysis reveals a notable average reduction in thefts of 63.5% for the linear model and 62.7% 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. This highlights the increasing effectiveness of Lojack in preventing auto thefts over time.

For comparison, we also report the estimated ATT_0 from two misspecified models. First, we consider the standard TWFE specification that incorporates spillover effects but overlooks the staggered nature of the treatment. Second, we consider the specification of

	TWFE (linear)		BW (linear)	
	Estimate	Reduction	Estimate	Reduction
ATT_0	-6.373	-64.1%	-7.834	-70.1%
ATT_0^0	N/A		-5.637	-55.4%
ATT_0^1	N/A		-8.453	-81.0%
ATT_0^2	N/A		-19.138	-87.8%
	TWFE (Poisson)		**BW (Poisson)**	
	Estimate	Reduction	Estimate	Reduction
ATT_0	-5.158	-59.1%	-5.851	-63.9%
ATT_0^0	N/A		-3.957	-46.3%
ATT_0^1	N/A		-6.489	-76.7%
ATT_0^2	N/A		-14.574	-86.4%

Table 2: Estimates of the aggregate ATT_0 using the TWFE specification that does not account for staggered adoption (the "TWFE" columns), and the specification of Borusyak et al. [2024] and Wooldridge [2022] that does not account for spillover effects (the "BW" columns), for each of linear and Poisson specifications. The "Reduction" columns stand for the reduction rate, which is calculated using the formula for computing $ATTP_0$.

Borusyak et al. [2024] and Wooldridge [2022] that accounts for staggered adoption but does not include spillover effects. Results from these models are presented in Table 2. We find that the TWFE regression estimates are biased when compared to the estimates presented in Table 1, where our preferred Poisson model specification indicates a slightly stronger reduction effect. Further, the estimates that neglect spillover effects exhibit a downward 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.

Next, we report the estimates of the spillover effects and the diagnostic test results. We estimate the following aggregate $ASUT$:

$$\begin{aligned}
ASUT_{2001} &= \sum_{g>2001} \frac{N_g}{N_{2002}+N_{2003}+N_{2004}+N_{\infty,0}} ASUT(g,0,2001),\\
ASUT_{2002} &= \sum_{g>2002} \frac{N_g}{N_{2003}+N_{2004}+N_{\infty,0}} ASUT(g,0,2002),\\
ASUT_{2003} &= \sum_{g>2003} \frac{N_g}{N_{2004}+N_{\infty,0}} ASUT(g,0,2003),\\
ASUT_{2004} &= ASUT(\infty,0,2004),
\end{aligned}$$

where $N_{\infty,0}$ is the number of units in the group $(G_i=\infty, H_i=0)$. Table 3 presents the estimated $ASUT$ values obtained from both linear and Poisson model specifications,

	Linear		Poisson	
	Estimate	Std Error	Estimate	Std Error
$ASUT_{2001}$	-1.268	0.887	-0.027	0.119
$ASUT_{2002}$	0.635	1.679	-0.009	0.272
$ASUT_{2003}$	N/A	N/A	0.452	0.416
$ASUT_{2004}$	N/A	N/A	0.476	0.502

Table 3: Estimates of the aggregate $ASUT$. The standard errors are clustered at the state level. We do not report $ASUT_{2003}$ and $ASUT_{2004}$ for the linear model since the estimates imply negative counterfactual theft counts for units affected by spillover effects.

with standard errors clustered at the state level. Although the spillover effects are not statistically significant, they exhibit a pattern where the estimated $ASUT$ increases over time. Note that statistical insignificance does not mean that there are no spillover effects, as the average spillover effect may be close to zero even if individual spillover effects are not. The result in Gonzalez-Navarro [2013] suggests that there may be spillover effects from a Lojack-installed model to the same model in other states without Lojack, but that spillover effects to other models in the same state are less likely.

For the diagnostic test of Assumption 4 described in Remark 3, we expect that $ATT_0(g,t) < 0$ and $AST(g,t) \geq 0$. This is because the installation of Lojack would reduce auto thefts for the treated models, while potentially causing car thieves to redirect their efforts toward other car models, leading to a possible increase in theft attempts. We find that all estimated $ATT_0(g,t)$ values are negative (not reported here), not detecting any evidence that $AST(g,t) > 0$.

For the diagnostic tests of Assumption 5 described in Remark 4, we note that our dataset has two pre-treatment periods, namely 1999 and 2000. We perform a pre-trend test for the year 2000, and we find that we do not reject the null hypothesis of parallel trends, with p-values of 0.164 for the linear model and 0.053 for the Poisson model. For the placebo tests, we split the $(G_i = \infty, H_i = 1)$ group into two subgroups by dividing the units based on two sets of states. We find that we reject the null hypothesis of the placebo test for the linear model, but we do not reject it for the Poisson model, with a p-value of 0.647.

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. 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.

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 extended group $g \in \tilde{\mathcal{G}} \equiv \{2, \dots, T, (\infty, 0), (\infty, 1)\}$, meaning that we have a total of $N = (T + 1)M$ units in the dataset. We consider the following data generating process (DGP) that embeds Assumptions 1, 2, 4 and 5. Depending on the specification of the outcome model—linear or Poisson—we adapt the relevant assumption, replacing Assumption 5 with 5' as necessary. The DGP for the linear model is given by:

$$Y_{it} = \alpha_{G_i} + \delta_t + \beta_{it} D_{it} + (1 - D_{it}) \cdot \sum_{j=1}^{M} \gamma_{it}^{j} \cdot D_{jt} + \varepsilon_{it},$$

where γ_{it}^{j} represents the spillover effect from unit j to unit i. The DGP for the Poisson model is specified similarly with the link function $F(x) = \exp(x)$. 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 $\beta_{it} = \beta_{G_i t}$ and $\gamma_{it}^{j} = \gamma_{G_i t}^{G_j}$.

- 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 group $(\infty, 0)$. 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 $(\infty, 1)$. As a result, each untreated unit receives a spillover effect of $-\rho \cdot \beta_{G_i t} / U_t$ from the treated unit i.

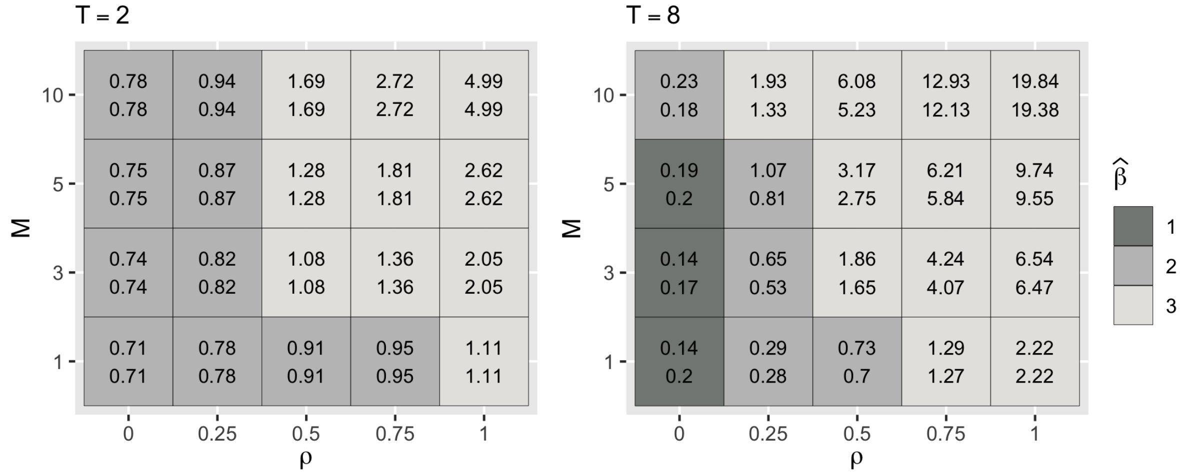
With this parametrization, Y_{it} is generated with an independent additive error term $\epsilon_{it} \sim N(0, \max(\alpha_{G_i})/10)$ for the linear model, and according to Poisson distribution for the Poisson model. We then estimate the aggregate ATT_0 defined as in Section 7, namely $ATT_0 = (1/\overline{G}) \sum_{g=2}^{T} \sum_{t=g}^{T} ATT_0(g,t)$, 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 estimator from standard TWFE regression, which neither accounts for staggered treatment adoption nor for spillovers.

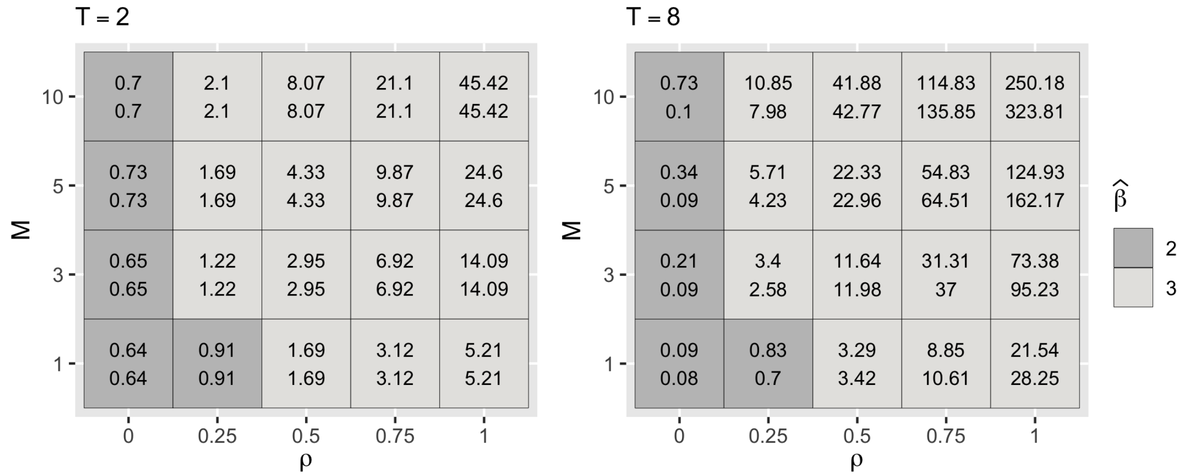
($\widehat{\beta}_2$) The estimator from fully interacted TWFE regression by Wooldridge [2022], which accounts for staggered treatment adoption but does not account for spillovers. This estimator is numerically equivalent to the imputation estimator by Borusyak et al. [2024].

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

Figure 3a and Table 4 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 3b and Table 5 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 3: 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 4: 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 5: 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 is established under the assumption 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 without interference, we extend the TWFE regession approach of Wooldridge [2022] to account for spillovers in linear and nonlinear 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

Peter M. Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. *Annals of Applied Statistics*, 11(4), 2017.

Tobias Berg, Markus Reisinger, and Daniel Streitz. Spillover effects in empirical corporate finance. *Journal of Financial Economics*, 142(3), 2021.

Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, 91(6), 2024.

Kyle Butts. Difference-in-Differences Estimation with Spatial Spillovers. *arXiv preprint arXiv:2105.03737*, 2023.

Brantly Callaway and Pedro H. C. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 2021.

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

Derek B Cornish and Ronald V Clarke. Understanding crime displacement: An application of rational choice theory. In *Crime Opportunity Theories*. Routledge, 2017.

Clément de Chaisemartin and Xavier D'Haultfoeuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9), 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*, 26(3), 2023.

Laura Forastiere, Edoardo M Airoldi, and Fabrizia Mealli. Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association*, 116(534), 2021.

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

Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 2021.

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

Gary Hettinger, Christina Roberto, Youjin Lee, and Nandita Mitra. Estimation of policy-relevant causal effects in the presence of interference with an application to the philadelphia beverage tax. *arXiv preprint arXiv:2301.06697*, 2023.

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

Michael G Hudgens and M Elizabeth Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482), 2008.

Jonathan Roth and Pedro HC Sant'Anna. When is parallel trends sensitive to functional form? *Econometrica*, 91(2), 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*, 235(2), 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*, 225(2), 2021.

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. *Available at SSRN 3906345*, 2022.

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

Ruonan Xu. Difference-in-differences with interference: A finite population perspective. *arXiv preprint arXiv:2306.12003*, 2023.

Corwin Zigler, Vera Liu, Fabrizia Mealli, and Laura Forastiere. Bipartite interference and air pollution transport: Estimating health effects of power plant interventions. *arXiv preprint arXiv:2012.04831*, 2020.

A Proofs

A.1 Proof of Remark 1

We show that Assumption 3 is equivalent to Equation (2). First, it is straightforward to show that Equation (2) implies Assumption 3, so we omit the proof of this direction. In what follows, we show that Assumption 3 implies Equation (2). Define

$$\begin{aligned}\alpha_i &\equiv Y_{i1}(\infty, \infty_{-i}),\\ \delta_t &\equiv \mathbb{E}(Y_{it}(\infty, \infty_{-i}) - Y_{i1}(\infty, \infty_{-i})|G_i = \infty).\end{aligned}$$

Then Assumption 3 can be written as

$$\mathbb{E}(Y_{it}(\infty, \infty_{-i}) - \alpha_i|G_i = g) = \delta_t,$$

which we rearrange as

$$\mathbb{E}(Y_{it}(\infty, \infty_{-i})|G_i = g) = \mathbb{E}(\alpha_i|G_i = g) + \delta_t. \tag{14}$$

Next, define

$$\varepsilon_{it} \equiv Y_{it}(\infty, \infty_{-i}) - Y_{i1}(\infty, \infty_{-i}) - \mathbb{E}(Y_{it}(\infty, \infty_{-i}) - Y_{i1}(\infty, \infty_{-i})|G_i = \infty),$$

from which, by rearranging, we obtain

$$Y_{it}(\infty, \infty_{-i}) = Y_{i1}(\infty, \infty_{-i}) + \mathbb{E}(Y_{it}(\infty, \infty_{-i}) - Y_{i1}(\infty, \infty_{-i})|G_i = \infty) + \varepsilon_{it},$$

which we can rewrite as, using the definition of α_i and δ_t,

$$Y_{it}(\infty, \infty_{-i}) = \alpha_i + \delta_t + \varepsilon_{it},$$

which is Equation (2).

It remains to show that $\mathbb{E}(\varepsilon_{it}|G_i = g) = 0$. This follows from the definition of ε_{it}:

$$\begin{aligned}
&\mathbb{E}(\varepsilon_{it}|G_i = g) \\
&= \mathbb{E}(Y_{it}(\infty, \infty_{-i}) - Y_{i1}(\infty, \infty_{-i})|G_i = g) - \mathbb{E}(Y_{it}(\infty, \infty_{-i}) - Y_{i1}(\infty, \infty_{-i})|G_i = \infty) \\
&= 0,
\end{aligned}$$

where the last equality follows from Assumption 3.

A.2 Proof of Theorem 1

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

$$\begin{aligned}
Y_{it} &= Y_{it}(G_i, \mathbf{G}_{-i}) \\
&= Y_{it}(\infty, \infty_{-i}) + [Y_{it}(G_i, \infty_{-i}) - Y_{it}(\infty, \infty_{-i})] + [Y_{it}(G_i, \mathbf{G}_{-i}) - Y_{it}(G_i, \infty_{-i})] \\
&= \alpha_i + \delta_t + [Y_{it}(G_i, \infty_{-i}) - Y_{it}(\infty, \infty_{-i})] + [Y_{it}(G_i, \mathbf{G}_{-i}) - Y_{it}(G_i, \infty_{-i})] + \varepsilon_{it},
\end{aligned}$$

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

$$\begin{aligned}
\beta_{it} &= Y_{it}(G_i, \infty_{-i}) - Y_{it}(\infty, \infty_{-i}), \\
\gamma_{it} &= Y_{it}(G_i, \mathbf{G}_{-i}) - Y_{it}(G_i, \infty_{-i}).
\end{aligned}$$

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

$$\mathbb{E}(Y_{it}|G_i = g) = \mathbb{E}(\alpha_i|G_i = g) + \delta_t + ATT_0(g,t) + AST(g,t), \tag{15}$$

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 (15) as

$$\mathbb{E}(Y_{it}|G_i = g) = \mathbb{E}(\alpha_i|G_i = g) + d_0 + ATT_0(g,t). \tag{16}$$

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

$$Y_{i1} = Y_{i1}(G_i, \mathbf{G}_{-i}) = Y_{i1}(\infty, \infty_{-i}) = \alpha_i + \delta_1 + \varepsilon_{it}$$

by Assumptions 2 and 3. Then it follows that

$$\mathbb{E}(Y_{i1}|G_i = g) = \mathbb{E}(\alpha_i + \delta_1 + \varepsilon_{it}|G_i = g) = \mathbb{E}(\alpha_i|G_i = g), \tag{17}$$

where $\delta_1 = 0$ and $\mathbb{E}(\varepsilon_{it}|G_i = g) = 0$ by Remark 1. We can then rewrite Equation (16) 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 (15) as

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

Using Equation (17), 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.3 Proof of Theorem 2

We first prove identification of $ATT_0(g,t)$. Recall that, for $t \geq g$:

$$ATT_0(g,t) = \mathbb{E}(Y_{it}(G_i, \infty_{-i}) - Y_{it}(\infty_i, \infty_{-i})|G_i = g),$$

which we can rewrite as

$$\begin{aligned} ATT_0(g,t) = \mathbb{E}(Y_{it}(G_i, \mathbf{G}_{-i}) - Y_{i1}(\infty_i, \infty_{-i})|G_i = g) \\ - \mathbb{E}(Y_{it}(G_i, \mathbf{G}_{-i}) - Y_{it}(G_i, \infty_{-i})|G_i = g) \\ - \mathbb{E}(Y_{it}(\infty_i, \infty_{-i}) - Y_{i1}(\infty_i, \infty_{-i})|G_i = g). \end{aligned}$$

We now simplify each term in the right-hand side. The first term simplifies to:

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

In addition, the second term is zero for $t \geq g$ by Assumption 4:

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

Lastly, using Assumption 5, the third term simplifies to:

$$\begin{aligned} &\mathbb{E}(Y_{it}(\infty_i, \infty_{-i}) - Y_{i1}(\infty_i, \infty_{-i})|G_i = g) \\ &= \mathbb{E}(Y_{it}(\infty, \mathbf{G}_{-i}) - Y_{i1}(\infty, \infty_{-i})|G_i = \infty, H_i = 1) \\ &= \mathbb{E}(Y_{it} - Y_{i1}|G_i = \infty, H_i = 1). \end{aligned}$$

Therefore, it follows that

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

for $t \geq g$.

Next, we prove identification of $ASUT(g,t)$. Recall that, for $q \leq t < g$:

$$ASUT(g,t) = \mathbb{E}(Y_{it}(\infty_i, \mathbf{G}_{-i}) - Y_{it}(\infty_i, \infty_{-i})|G_i = g),$$

which we can rewrite as

$$
\begin{aligned}
ASUT(g,t) = \mathbb{E}(Y_{it}(\infty_i, \mathbf{G}_{-i}) - Y_{i1}(\infty_i, \infty_{-i})|G_i = g) \\
- \mathbb{E}(Y_{it}(\infty_i, \infty_{-i}) - Y_{i1}(\infty_i, \infty_{-i})|G_i = g).
\end{aligned}
$$

We now simplify each term in the right-hand side. Using Assumption 2, the first term simplifies to:

$$
\begin{aligned}
\mathbb{E}(Y_{it}(\infty_i, \mathbf{G}_{-i}) - Y_{i1}(\infty_i, \infty_{-i})|G_i = g) &= \mathbb{E}(Y_{it}(G_i, \mathbf{G}_{-i}) - Y_{i1}(G_i, \mathbf{G}_{-i})|G_i = g) \\
&= \mathbb{E}(Y_{it} - Y_{i1}|G_i = g).
\end{aligned}
$$

In addition, using Assumptions 2 and 5, the second term simplifies to:

$$
\begin{aligned}
&\mathbb{E}(Y_{it}(\infty_i, \infty_{-i}) - Y_{i1}(\infty_i, \infty_{-i})|G_i = g) \\
&= \mathbb{E}(Y_{it}(\infty, \mathbf{G}_{-i}) - Y_{i1}(\infty, \infty_{-i})|G_i = \infty, H_i = 1) \\
&= \mathbb{E}(Y_{it} - Y_{i1}|G_i = \infty, H_i = 1).
\end{aligned}
$$

Therefore, it follows that

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

for $q \leq t < g$. ■

A.4 Proof of Theorem 3

Using Remark 5, we can express Y_{it} for every $2 \leq g < \infty$ as:

$$
\begin{aligned}
&\mathbb{E}(Y_{it}|G_i = g) \\
&= \mathbb{E}(Y_{it}(G_i, \mathbf{G}_{-i})|G_i = g) \\
&= \mathbb{E}(Y_{it}(\infty, \infty_{-i}) + [Y_{it}(G_i, \infty_{-i}) - Y_{it}(\infty, \infty_{-i})] + [Y_{it}(G_i, \mathbf{G}_{-i}) - Y_{it}(G_i, \infty_{-i})]|G_i = g) \\
&= \mathbb{E}(Y_{i1}(\infty, \infty_{-i})|G_i = g) + \delta_t + \mathbb{E}(Y_{it}(G_i, \infty_{-i}) - Y_{it}(\infty, \infty_{-i})|G_i = g) \\
&\qquad + \mathbb{E}(Y_{it}(G_i, \mathbf{G}_{-i}) - Y_{it}(G_i, \infty_{-i})|G_i = g).
\end{aligned}
$$

Since $G_i \neq \infty$ implies $H_i = 0$, we can rewrite the above as

$$
\begin{aligned}
\mathbb{E}(Y_{it}|G_i = g, H_i = 0) = \mathbb{E}(Y_{i1}(\infty, \infty_{-i})|G_i = g, H_i = 0) + \delta_t \\
+ \mathbb{E}(Y_{it}(G_i, \infty_{-i}) - Y_{it}(\infty, \infty_{-i})|G_i = g, H_i = 0) \\
+ \mathbb{E}(Y_{it}(G_i, \mathbf{G}_{-i}) - Y_{it}(G_i, \infty_{-i})|G_i = g, H_i = 0).
\end{aligned} \tag{18}
$$

In addition, using Remark 5, we can express Y_{it} for $G_i = \infty$ and $H_i = 1$ as:

$$
\begin{aligned}
&\mathbb{E}(Y_{it}|G_i = \infty, H_i = 1) \\
&= \mathbb{E}(Y_{it}(\infty_i, \mathbf{G}_{-i})|G_i = \infty, H_i = 1) \\
&= \mathbb{E}(Y_{it}(\infty, \infty_{-i}) + [Y_{it}(\infty_i, \mathbf{G}_{-i}) - Y_{it}(\infty_i, \infty_{-i})]|G_i = \infty, H_i = 1) \\
&= \mathbb{E}(Y_{i1}(\infty, \infty_{-i})|G_i = \infty, H_i = 1) + \delta_t.
\end{aligned} \tag{19}
$$

Lastly, Remark 5 and the definition of δ_t respectively implies that:

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

which implies that

$$
\mathbb{E}(Y_{it}(\infty, \infty_{-i})|G_i = \infty, H_i = 0) = \mathbb{E}(Y_{i1}(\infty, \infty_{-i})|G_i = \infty, H_i = 0) + \delta_t.
$$

Therefore, we can express Y_{it} for $G_i = \infty$ and $H_i = 0$ as:

$$
\begin{aligned}
&\mathbb{E}(Y_{it}|G_i = \infty, H_i = 0) \\
&= \mathbb{E}(Y_{it}(\infty_i, \mathbf{G}_{-i})|G_i = \infty, H_i = 0) \\
&= \mathbb{E}(Y_{it}(\infty, \infty_{-i}) + [Y_{it}(\infty_i, \mathbf{G}_{-i}) - Y_{it}(\infty_i, \infty_{-i})]|G_i = \infty, H_i = 0) \\
&= \mathbb{E}(Y_{i1}(\infty, \infty_{-i})|G_i = \infty, H_i = 0) + \delta_t \\
&\qquad\qquad + \mathbb{E}(Y_{it}(\infty_i, \mathbf{G}_{-i}) - Y_{it}(\infty_i, \infty_{-i})|G_i = \infty, H_i = 0).
\end{aligned} \tag{20}
$$

Now, let

$$
\begin{aligned}
A_{gh} &\equiv \mathbb{E}(Y_{i1}(\infty, \infty_{-i})|G_i = g, H_i = h), \\
B_{ght} &\equiv \mathbb{E}(Y_{it}(G_i, \infty_{-i}) - Y_{it}(\infty, \infty_{-i})|G_i = g, H_i = h), \\
C_{ght} &\equiv \mathbb{E}(Y_{it}(G_i, \mathbf{G}_{-i}) - Y_{it}(G_i, \infty_{-i})|G_i = g, H_i = h).
\end{aligned}
$$

We can rewrite Equations (18), (19) and (20) as:

$$
\begin{aligned}
\mathbb{E}(Y_{it}|G_i = g, H_i = 0) \quad &= A_{gh} + \delta_t + B_{ght} + C_{ght} \quad \text{if } 2 \le g < \infty, \\
\mathbb{E}(Y_{it}|G_i = \infty, H_i = 0) \quad &= A_{\infty,h} + \delta_t + C_{\infty,h,t}, \\
\mathbb{E}(Y_{it}|G_i = \infty, H_i = 1) \quad &= A_{\infty,h} + \delta_t.
\end{aligned}
$$

In addition, by the definitions of B_{ght} and C_{ght} and Assumptions 2 and 4, for $2 \le g < \infty$, it follows that

$$
\begin{aligned}
&B_{ght} = 0 \text{ if } t < g, \\
&C_{ght} = 0 \text{ if } t < q \quad \text{or} \quad g \le t,
\end{aligned}
$$

and $C_{\infty,h,t} = 0$ if $t < q$. Therefore, we can further rewrite Equations (18), (19) and (20) as:

$$\mathbb{E}(Y_{it}|G_i = g, H_i = h) = \begin{cases} A_{gh} + \delta_t & \text{if } t < q \\ A_{gh} + \delta_t & \text{if } g = \infty, h = 1 \text{ and } q \leq t \\ A_{gh} + \delta_t + C_{\infty,h,t} & \text{if } g = \infty, h = 0 \text{ and } q \leq t \\ A_{gh} + \delta_t + C_{ght} & \text{if } 2 \leq g < \infty \text{ and } q \leq t < g \\ A_{gh} + \delta_t + B_{ght} & \text{if } 2 \leq g < \infty \text{ and } t \geq g \end{cases}.$$

Note that, according to the definition of D_{it} in the main text, D_{it} is equal to 1 if $2 \leq G_i < \infty$ and $t \geq G_i$. Similarly, according to the definition in the main text, S_{it} is equal to 1 if $(G_i = \infty, H_i = 0)$ and $q \leq t$ holds or if $2 \leq G_i < \infty$ and $q \leq t < g$. Then we can combine the above expressions into one unified expression as follows:

$$\begin{aligned} \mathbb{E}(Y_{it}|G_i = g, H_i = h) = A_{gh} + \delta_t &+ \sum_{g',h'} \sum_{t'=g}^{T} B_{g'h't'} \mathbf{1}(g' = g, h' = h, t' = t) D_{it} \\ &+ \sum_{g',h'} \sum_{t'=q}^{\min\{g'-1,T\}} C_{g'h't'} \mathbf{1}(g' = g, h' = h, t' = t) S_{it}. \end{aligned} \tag{21}$$

Now we proceed to prove the theorem. Let $\mathbf{X}_{it}$ be the vector of regressors in Equation (9), 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). Then the regressors $\mathbf{X}_i \equiv (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iT})$ 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

$$\mathbb{E}(Y_{it}|\mathbf{X}_i) = \mathbb{E}(Y_{it}|G_i = g(\mathbf{X}_i), H_i = h(\mathbf{X}_i))$$

where $g(\mathbf{X}_i)$ and $h(\mathbf{X}_i)$ are the extended group label identified by $\mathbf{X}_i$. Then we can write Equation (21) as:

$$\begin{aligned} \mathbb{E}(Y_{it}|\mathbf{X}_i) = A_{g(\mathbf{X}_i)h(\mathbf{X}_i)} &+ \delta_t \\ &+ \sum_{g',h'} \sum_{t'=g(\mathbf{X}_i)}^{T} B_{g'h't'} \mathbf{1}(g' = g(\mathbf{X}_i), h' = h(\mathbf{X}_i), t' = t) D_{it} \\ &+ \sum_{g',h'} \sum_{t'=q}^{\min\{g'-1,T\}} C_{g'h't'} \mathbf{1}(g' = g(\mathbf{X}_i), h' = h(\mathbf{X}_i), t' = t) S_{it}, \end{aligned}$$

which coincides with the population regression of Equation (9). Then it follows that

$$\begin{aligned} \beta_{gt} &= B_{g,0,t} \quad \text{for } 2 \leq g < \infty \text{ and } t \geq g, \\ \gamma_{g,0,t} &= C_{g,0,t} \quad \text{for } q \leq t < g. \end{aligned}$$

which means that

$$\begin{aligned} \beta_{gt} &= \mathbb{E}(Y_{it}(G_i, \infty_{-i}) - Y_{it}(\infty, \infty_{-i}) | G_i = g) \quad \text{for } 2 \leq g < \infty \text{ and } t \geq g, \\ \gamma_{g,0,t} &= \mathbb{E}(Y_{it}(\infty_i, \mathbf{G}_{-i}) - Y_{it}(\infty_i, \infty_{-i}) | G_i = g, H_i = 0) \quad \text{for } q \leq t < g, \end{aligned}$$

by the fact that $G_i \neq \infty$ implies $H_i = 0$ and by Assumption 2. ■

A.5 Proof of Corollary 1

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

$$\begin{aligned} Y_{it} &= Y_{it}(G_i, \mathbf{G}_{-i}) \\ &= Y_{it}(\infty, \infty_{-i}) + [Y_{it}(G_i, \infty_{-i}) - Y_{it}(\infty, \infty_{-i})] + [Y_{it}(G_i, \mathbf{G}_{-i}) - Y_{it}(G_i, \infty_{-i})] \,. \end{aligned}$$

Then, using Equation (12), we can write the expectation of Y_{it} as

$$\mathbb{E}(Y_{it} | G_i = g) = F^{-1}(\alpha_g + \delta_t) + ATT_0(g, t) + AST(g, t),$$

where

$$ATT_0(g, t) = \mathbb{E}(Y_{it}(G_i, \infty_{-i}) - Y_{it}(\infty, \infty_{-i}) | G_i = g),$$

and

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

Then, by replicating the arguments in Theorem 1 that starts from Equation (15), it is straightforward to show that $ATT_0(g, t)$ is identified if and only if $F^{-1}(\alpha_g + \delta_t) + AST(g, t)$ is identified. ■

A.6 Proof of Corollary 2

Similarly to Theorem 2, we first prove identification of $ATT_0(g, t)$. Recall that, for $t \geq g$:

$$ATT_0(g, t) = \mathbb{E}(Y_{it}(G_i, \infty_{-i}) - Y_{it}(\infty_i, \infty_{-i}) | G_i = g),$$

which we can rewrite as

$$ATT_0(g,t) = \mathbb{E}(Y_{it}(G_i, \mathbf{G}_{-i}) - Y_{it}(\infty_i, \infty_{-i})|G_i = g)$$
$$- \mathbb{E}(Y_{it}(G_i, \mathbf{G}_{-i}) - Y_{it}(G_i, \infty_{-i})|G_i = g).$$

We now simplify each term in the right-hand side. Using Remark 12 and Assumption 2, the first term simplifies to:

$$\begin{aligned}
&\mathbb{E}(Y_{it}(G_i, \mathbf{G}_{-i}) - Y_{it}(\infty_i, \infty_{-i})|G_i = g) \\
&= \mathbb{E}(Y_{it}|G_i = g) - \mathbb{E}(Y_{it}(\infty_i, \infty_{-i})|G_i = g) \\
&= \mathbb{E}(Y_{it}|G_i = g) - F^{-1}\Big[F(\mathbb{E}(Y_{i1}|G_i = g)) + \\
&\qquad F(\mathbb{E}(Y_{it}|G_i = \infty, H_i = 1)) - F(\mathbb{E}(Y_{i1}|G_i = \infty, H_i = 1))\Big].
\end{aligned}$$

In addition, the second term is zero for $t \geq g$ by Assumption 4:

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

Therefore, it follows that

$$\begin{aligned}
ATT_0(g,t) = \mathbb{E}(Y_{it}|G_i = g) - F^{-1}\Big[&F(\mathbb{E}(Y_{i1}|G_i = g)) + \\
&F(\mathbb{E}(Y_{it}|G_i = \infty, H_i = 1)) - F(\mathbb{E}(Y_{i1}|G_i = \infty, H_i = 1))\Big]
\end{aligned}$$

for $t \geq g$.

Next, we prove identification of $ASUT(g,t)$. Recall that, for $q \leq t < g$:

$$ASUT(g,t) = \mathbb{E}(Y_{it}(\infty_i, \mathbf{G}_{-i}) - Y_{it}(\infty_i, \infty_{-i})|G_i = g).$$

which we can rewrite as

$$ASUT(g,t) = \mathbb{E}(Y_{it}(\infty_i, \mathbf{G}_{-i})|G_i = g) - \mathbb{E}(Y_{it}(\infty_i, \infty_{-i})|G_i = g).$$

We now simplify each term in the right-hand side. Using Assumption 2, the first term simplifies to:

$$\begin{aligned}
\mathbb{E}(Y_{it}(\infty_i, \mathbf{G}_{-i})|G_i = g) &= \mathbb{E}(Y_{it}(G_i, \mathbf{G}_{-i})|G_i = g) \\
&= \mathbb{E}(Y_{it}|G_i = g).
\end{aligned}$$

In addition, using Remark 12 and Assumption 2, the second term simplifies to:

$$\mathbb{E}(Y_{it}(\infty_i, \infty_{-i})|G_i = g) = F^{-1}\Big[F(\mathbb{E}(Y_{i1}|G_i = g)) + F(\mathbb{E}(Y_{it}|G_i = \infty, H_i = 1)) - F(\mathbb{E}(Y_{i1}|G_i = \infty, H_i = 1))\Big].$$

Therefore, it follows that

$$ASUT(g,t) = \mathbb{E}(Y_{it}|G_i = g) - F^{-1}\Big[F(\mathbb{E}(Y_{i1}|G_i = g)) + F(\mathbb{E}(Y_{it}|G_i = \infty, H_i = 1)) - F(\mathbb{E}(Y_{i1}|G_i = \infty, H_i = 1))\Big].$$

for $q \leq t < g$. ■