UTS Business School – Department of Economics
Dr Chau Chak Wing Building
14 – 28 Ultimo Road, Ultimo NSW 2007

T: +61 2 9514 7713
economics@uts.edu.au

Department of Economics

Working Paper Series

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

Mario Fiorini [1]
Wooyong Lee [2]
Gregor Pfeifer [3]

[1 & 2] The University of Technology Sydney
[3] The University of Sydney, CESifo and IZA

No. 02 / 2024 **January 2024**

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

Mario Fiorini Wooyong Lee Gregor Pfeifer*

January 24, 2024

Abstract

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

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

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

1 Introduction

The Difference-in-Differences literature, particularly the one concerned with staggered treatment adoption, has experienced significant advances in the last few years, and papers by Roth et al. [2023] and de Chaisemartin and D'Haultfœuille [2021] have tried to summarize these developments. Within this array of advances, one area still understudied is the one linked to spillovers, which can be assumed to be present—and relevant—in many empirical settings (Roth et al., 2023). Spillovers imply that the Stable Unit Treatment Value Assumption (SUTVA) assumption does not hold. Our work contributes to this area and links to two active Difference-in-Differences literature strands.

The first one studies the identification of average treatment effects in the presence of spillover effects. Work by Berg et al. [2021], Butts [2023], Clarke [2017], and Huber and Steinmayr [2021] highlights the potential for biased estimates of the ATT if the treatment also impacts units that are not formally treated. Hereafter, we refer to this body of work as the Spillover literature. One issue is that untreated units are no longer valid controls. So far, proposed solutions mostly centre around ruling out spillovers for a given group of units, often based on some spatial distance, allowing the researcher to use this latter group as a control. Alternatively, if sufficient information exists, one can parametrize how units are exposed to spillovers. Another drawback is that multiple treatment effects can be defined in the presence of spillovers. The researcher might be interested in the treatment effect without interference (e.g., the one normally identified under SUTVA) or in a broader effect that accounts for spillovers. We contribute to this literature by providing an argument for identifying several treatments of interest. Our setting also departs from this literature since we focus on the more complex staggered treatment adoption, which has the potential for cumulative spillovers. Nevertheless, our results also apply to the more classical, simultaneous treatment case, which is the focus of the work mentioned above.

The second strand of the literature focuses on estimation issues under said staggered adoption and heterogeneous treatment effects across units and time. Borusyak et al. [2021], de Chaisemartin and D'Haultfoeuille [2020], Callaway and Sant'Anna [2020], Goodman-Bacon [2021], Sun and Abraham [2020] and Wooldridge [2022] highlight how OLS estimation of the Two Way Fixed Effect (TWFE) model might lead to point estimates that are different from the ATT, and most other estimands of interest, to the extreme of being uninterpretable. This constitutes an estimation issue rather than an identification problem, and the authors suggest alternative estimators that recover meaningful estimands by re-weighting appropriately. Hereafter, we refer to this body of work as the Weighting literature. We contribute to this literature by extending it to the case of spillovers in both linear and non-linear models.

Specifically regarding contributions, we first establish the identifying assumptions for the ATT given a staggered Difference-in-Differences set up in the presence of spillovers. We show that aside from the canonical i) treatment irreversibility, ii) no-anticipation, and iii) parallel trends assumptions, identification requires that once a unit is treated, it does not experience spillovers, past, present, and future. We name this assumption treated-immune. This assumption also unifies the ATTs, because they are the same with or without spillovers, simplifying policy evaluation and joining with the definition of ATT under SUTVA. Similar to the existing Spillover literature, we also assume that a set of never-treated units is not exposed to spillovers. The combination of these two assumptions allows for the identification of the ATT. Below, we argue that such a scenario applies to many contexts. Differently from Butts [2023], who is closest to our work, we directly focus on the staggered treatment scenario and, importantly, provide identification assumptions for all the ATTs, including the direct effect.[1]

Our second contribution regards estimation. We show that either the imputation approach of Borusyak et al. [2021] or the extended TWFE model approach of Wooldridge [2022] can be used to account for spillovers. Furthermore, we discuss identification and estimation in the non-linear case of count data, broadening the range of applications where our setup can be applied to.[2] For our empirical application, we revisit Gonzalez-Navarro [2013], who studied the effects of installing a device tracking cars in the event of theft. Gonzalez-Navarro [2013] does consider spillovers, but does not correct for the staggered treatment. Since car theft is a count data outcome, we implement the non-linear Poisson Difference-in-Differences adjusted for spillovers. Our correction leads to a larger effect of the policy relative to the original contribution's specification.

Finally, we perform a Monte Carlo analysis, highlighting the bias-variance trade-off implicit in the correction for staggered treatment and spillovers. Identification of time and group fixed effects can neither rely on the already treated units due to heterogeneous treatment effects, nor on the untreated units potentially exposed to spillovers. However, the benefit of excluding such units from estimation can be small if treatment effects are relatively homogeneous and if spillovers are small, while costing the researcher precision. We compare the traditional TWFE estimator, which ignores both staggered adoption and spillovers, the Borusyak et al. [2021] imputation 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,[3]

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

[2]While we extend identification and estimation to the non-linear case, Butts [2023] focuses his discussion on the linear setting.

[3]Borusyak et al. [2021]'s imputation estimator and Wooldridge [2022]'s extended TWFE estimator

showing that our estimator performs competitively in circumstances that reflect staggered treatment adoption and the presence of spillovers.

The remainder of the paper is organized as follows. Section 2 provides intuition alongside two motivating examples, after which Section 3 lays out the formal DiD setup with staggered treatment adoption. Section 4 establishes conditions for identifying the ATT, while Section 5 discusses estimation and inference considering the formerly established assumptions. Section 6 extends our model to the non-linear case, and Section 7 discusses a corresponding application, namely, Gonzalez-Navarro [2013]. Section 8 provides Monte Carlo simulations, and Section 9 concludes.

2 Motivating Examples and Intuition

To illustrate our setting, consider the case of three groups, A, B, and Z, observed over three periods, 1, 2, and 3. Groups are defined based on the timing of their treatment. Group A is treated in period 2, group B is treated in period 3, and group Z is not yet treated by the last period. Each group has two units, denoted a, a', b, b', z, z'. Figure 1 illustrates the treatment and spillover mechanisms using a DAG. For clarity, we omit nodes for untreated units and unit b'. Solid edges (i.e., lines) indicate effects under no interference (β_{it}), equivalent to a setting with SUTVA. We call them direct effects. Dotted edges indicate spillovers from unit j to other treated units (γ^j_{it}), and dashed ones represent spillovers from unit j to untreated units (η^j_{it}). This Figure highlights the key issues in the presence of spillovers: there are no valid controls and many treatment effects. Under SUTVA, only the solid edges would exist.

With some parametrization, we can also visualize a possible data pattern. Let the outcome of interest be deterministic and given by:

$$Y_{it} = 1 + \delta_t + \beta_{it} \cdot D_{it} + D_{it} \cdot \sum_{j \neq i} \gamma^j_{it} \cdot D_{jt} + (1 - D_{it}) \cdot \sum_{j \neq i} \eta^j_{it} \cdot D_{jt}, \quad (1)$$

where D_{it} is a binary variable equal to 1 when unit i is treated. Equation (1) is an example of a scenario where unit i's outcome is not only impacted by its own treatment effect, β_{it} but also by the treatment of other units via γ^j_{it} and η^j_{it}. Let the policy effect in the absence of interference be homogeneous across units and time: $\beta_{it} = \beta = -0.5 \ \forall i, t$. Therefore, $ATT \equiv \mathbb{E}(\beta \mid D_{it} = 1) = -0.5$. Let the time effect $\delta_t = 0.1 \cdot t$.

Without spillovers, the data would look like the left panel in Figure 2. The estimators proposed in the Weighting literature would exploit the never-treated group Z and the not-yet-treated observations in group-time $B2$ as the control group to identify the time

are numerically equivalent in our simulations.

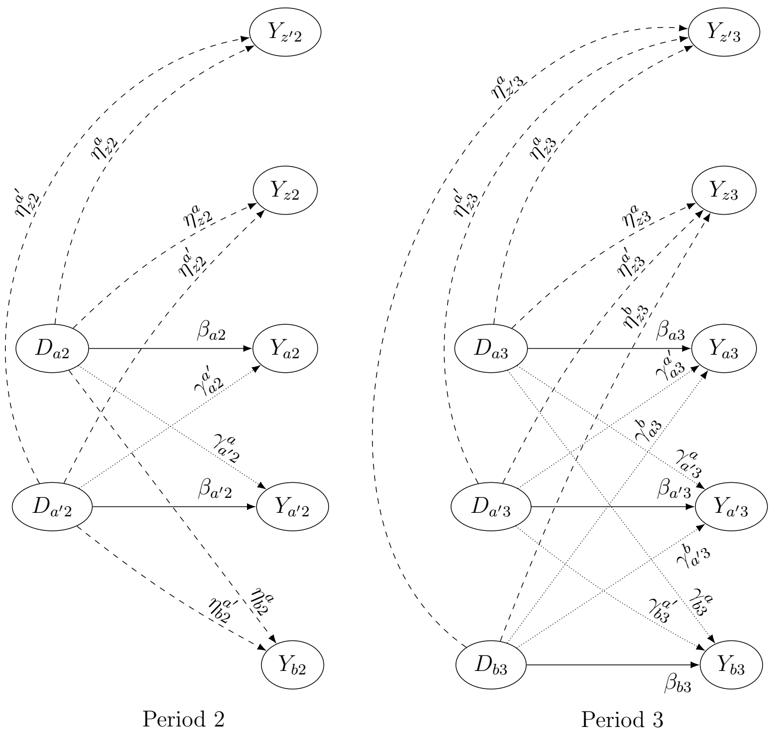


Figure 1: A DAG with Treatment and Spillover paths

effects and, in turn, the ATT.[4] Let $G_i = \min\{t \mid D_{it} = 1\}$ be the first period in which a unit is treated ($G_i = \infty$ if never treated). Wooldridge [2022] proposes a regression where D_{it} is interacted with indicators of G_i:

$$Y_{it} = \alpha_i + \delta_t + \sum_{g} \sum_{t'=g}^{T} \beta_{gt'} \cdot \mathbf{1}(G_i = g, t = t') \cdot D_{it} + \varepsilon_{it} \tag{2}$$

Under suitable conditions, it identifies the ATT for each group G_i at each time t by $\beta_{gt} = \mathbb{E}(\beta_{it} \mid G_i = g)$, all equal to -0.5 in our example.

We introduce two alternative scenarios with spillovers that we will use throughout the paper to motivate our assumptions and the empirical application—called examples 1 and 2 below. In the first scenario, the spillover is in the form of a diffusion effect, meaning that β_{it} and $(\gamma^j_{it}, \eta^j_{it})$ have the same sign, hence, the direct effect and the spillover effect reinforce each other. In the second scenario, the spillover is in the form of displacement such that β_{it} and $(\gamma^j_{it}, \eta^j_{it})$ 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 nearby villages a and a' are the first to adopt

[4] In practice, given the homogeneous treatment effect, even the already treated $A3$ observations are valid controls, and the standard TWFE estimation would also recover the ATT.

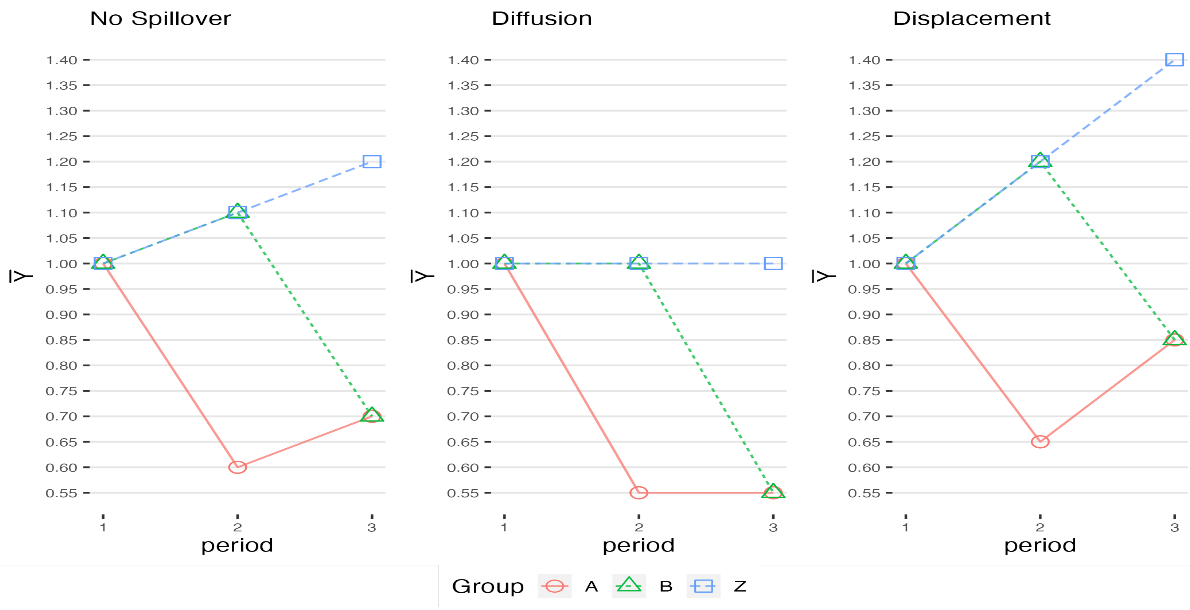


Figure 2: Data under equation (1)

the plant. Both of the group's A villages and the not-yet-treated downstream—villages might experience spillovers in the form of cleaner water. The data could then look like the middle panel in Figure 2, where we set the spillovers to also be homogeneous across units and time: $\gamma_{it}^{j} = \eta_{it}^{j} = \gamma = \eta = -0.1 \ \forall i, j, t$. First, there is no valid control group because the never-treated and the not-yet-treated experience spillovers leading to biased estimates of the time effects:

$$\mathbb{E}(Y_{it} - Y_{i,t-1} \mid D_{it} = D_{i,t-1} = 0) = \delta_t - \delta_{t-1} + \sum_{j \neq i} \left(\gamma_{it}^{j} \cdot D_{jt} - \gamma_{i,t-1}^{j} \cdot D_{j,t-1} \right)$$

For instance, in period 3, when only group Z is untreated, the bias would be $\gamma_{Z3}^{A} + \gamma_{Z3}^{B} - \gamma_{Z2}^{A} = -0.1$. Given the staggered treatment, even homogeneous spillovers would cause identification issues. Second, even if the never-treated group Z were not exposed to spillovers, allowing the researcher to identify the time effects, it would not be possible to identify the direct and spillover effects separately. The estimation methods proposed in the Weighting literature, such as (2), would at best identify the average sum of direct and spillover effects $\mathbb{E}(\beta + \sum_{j \neq i} \gamma_{it}^{j}) = \beta + \frac{7}{3}\gamma = -0.73$, equally weighting $\widehat{\beta}_{A2} = (\beta + \gamma) = -0.6$, and $\widehat{\beta}_{A3} = \widehat{\beta}_{B3} = (\beta + 3 \times \gamma) = -0.8$.

Example 2 (installation of stolen vehicle recovery devices)**.** Gonzalez-Navarro [2013] studies the effect of installing a stolen vehicle recovery device on car thefts. Treatment adoption is staggered across states (within a country) and only applies to specific car models. Hence, car theft might be displaced to unprotected models in treated states or protected models in untreated states. Gonzalez-Navarro [2013] finds a 52% increase in theft for protected models in untreated states. The data could follow the patterns of the right panel in Figure 2, where $\gamma_{it}^{j} = \eta_{it}^{j} = +0.1\ \forall i, j, t$. The same identification issues apply. Note that, especially in the case of displacement, spillovers could sequentially become very large as more and more treated units spill on an increasingly narrower pool of untreated units, exacerbating such identification issues. In this example, $\mathbb{E}(\beta + \sum_{j \neq i} \gamma_{it}^{j}) = -0.26$, equally weighting the A2 group-time, $(\beta + \gamma) = -0.4$, and the A3 and B3 ones, $(\beta + 3 \times \gamma) = -0.2$.[5]

In some scenarios, $\mathbb{E}(\beta + \sum_{j \neq i} \gamma_{it}^{j})$ might be all the researcher needs to know. However, it does not disentangle the different types of effects, limiting its usefulness for policy-making. For example, when a unit i decides whether to participate in a policy or treatment, it may only want to consider β_{it} because other units' decisions are out of its control. Even a policy-maker whose jurisdiction spans all units might want to determine the different channels. Figure 3 illustrates our treated-immune assumption that all treated units are not exposed to spillovers, combined with the assumption that a subset of the never-treated units are also not exposed to spillovers. There are no longer edges to treated groups, i.e., $\gamma_{it}^{j} = 0$, and there are no edges to unit z', allowing for identification of the time effects.[6] We also accomplish identification of the average spillovers to the not-yet-treated and a sub-set of the never-treated, i.e. those exposed to spillovers, allowing the researcher to shed light on the degree of interference across units.

Example 1 [continued]. We can use villages upstream from group A to identify time effects, because water only flows in the opposite direction. Alternatively, we could exploit time variation in another health outcome that does not depend on water quality but has similar trends pre-policy. We can then invoke the treated-immune assumption to identify the ATT. Since there is now a water treatment installation, future treatment of upstream villages should be inconsequential. Similarly, previous spillovers that had ameliorated water quality should be irrelevant, because the water is now fully treated. Health outcomes observed after the treatment would not incorporate any spillover effects, and only SUTVA-type direct effects would be present.

Example 2 [continued]. We can use faraway states or different car models to identify time effects. This implies that car thieves are only willing to travel a limited distance and that their network is geographically bounded, or that they focus on specific models. Gonzalez-Navarro [2013] shows that the data supports both the geographical bounds conjecture and the car-model targeting conjecture. Our treated-immune assumption implies

[5] For instance, under full displacement that is equally spread across untreated units, $\gamma_{it} = \eta_{it} = \frac{\kappa_t^T \cdot \beta}{\kappa_t^U}$, where κ_t^T and κ_t^U are the number of treated and untreated units at time t, respectively.

[6] In practice, we assume that multiple units in group Z are not exposed to spillovers to average out idiosyncratic shocks.

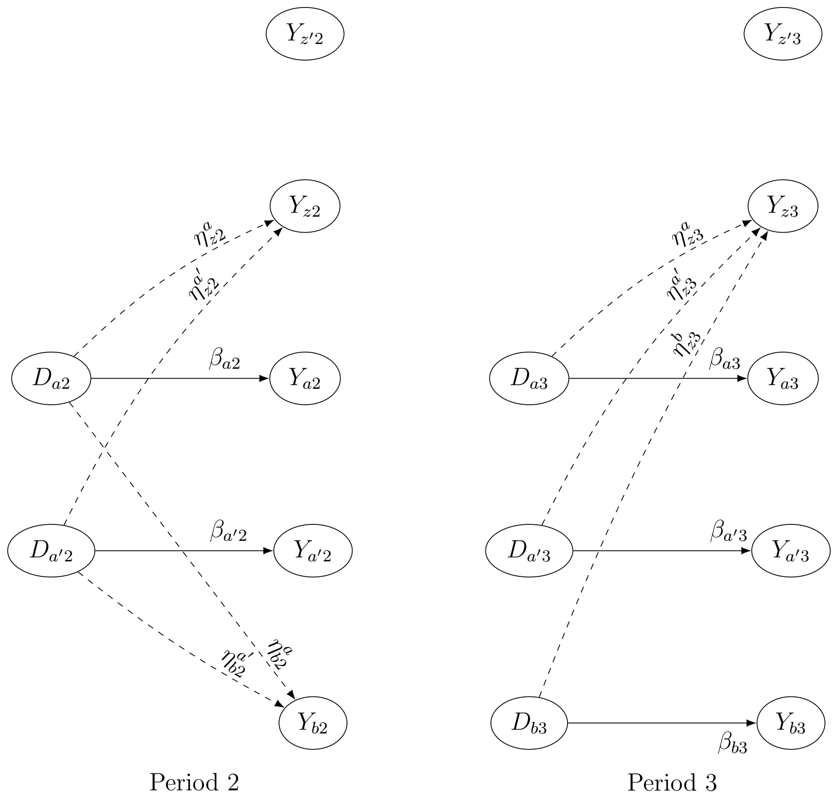


Figure 3: DAG under the key identification assumptions

that adopting stolen vehicle recovery devices should lead thieves to target unprotected states and models, leaving protected units unaffected by spillovers. Again, car theft observed after the treatment would include only SUTVA-type direct effects. One could argue that if the set of protected states and models were to expand to become almost universal, thieves might resort to stealing the protected cars again, violating the assumption. This is possible unless thieves switch their focus from cars to other, less protected targets, or exit the illegal market altogether. Nevertheless, almost universal treatment is an extreme case comprising other issues, such as a small set of control units to form a counterfactual.

3 Setup

We consider a DID model with a staggered adoption design, observed over the time periods $t \in \{1, \ldots, T\}$. For each unit at each time t, we consider a 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 subsequent periods.

Assumption 1 (irreversibility)**.** *For any two time periods* (s,t) *such that* $s < t$*, if a unit has a treatment status of 1 at time* s*, then it also has a treatment status of 1 at time* t*.*

Under Assumption 1, we define the group label $\mathcal{G}$ to be the subset of $\{1,\ldots,T,\infty\}$, representing the collection of periods when units enter treatment. We then assign a unit to a group $g \in \mathcal{G}$ if it enters treatment at period g, except for the group labelled ∞, which remains untreated until time T. For each group $g \in \mathcal{G}$, we consider the population of units indexed by i. We denote unit i in group g by a (i,g) pair, and we let Λ_g represent the set of all (i,g) indices within group g in the population, with $\Lambda \equiv \bigcup_g \Lambda_g$ being the set of all indices across all groups. We define $D_{igt} \in \{0,1\}$ as the binary treatment indicator and let $D_{ig} \equiv (D_{ig1},\ldots,D_{igT})$ represent the treatment history. Furthermore, we define the vector d_g as follows:

$$d_g \equiv (\underbrace{0,\ldots,0}_{t<g},\underbrace{1,\ldots,1}_{t\geq g}), \tag{3}$$

which represents the treatment statuses for group g across all periods and let d_g^t denote the treatment history up to time t. In cases with no ambiguity, we will use 0 to represent d_∞, since d_∞ corresponds to the vector of zeros.

Let $Y_{igt}(\{d_{jh}\}_{(j,h)\in\Lambda})$ be the potential outcome for unit i in group g at time t when $\{D_{jh}\}_{(j,h)\in\Lambda}$ is set to $\{d_{jh}\}_{(j,h)\in\Lambda}$. It is important to note that the potential outcome depends on the treatment statuses of all units $(j,h)\in\Lambda$, whereas under SUTVA it would be a function of the unit's own treatment status only, i.e., $Y_{igt}(\{d_{jh}\}_{(j,h)\in\Lambda}) = Y_{igt}(d_{ig})$. To facilitate future discussions, we rewrite the potential outcome by partitioning the population's treatment into the unit's own treatment status and those of the other units:

$$Y_{igt}(d_{ig},\{d_{jh}\}_{(j,h)\in\Lambda\setminus\{(i,g)\}}).$$

This notation emphasizes the possibility of unit (i,g) being affected by spillover effects from units not included in the sample, as Λ represents the index set of the entire population. Note that Assumption 1 and the definition of the group labels $\mathcal{G}$ imply that we observe $d_{jh} = d_h$ for every $(j,h)\in\Lambda$ in the data.

We define $\mathbf{d}_{(i,g)}$ to be a particular value of $\{d_{jh}\}_{(j,h)\in\Lambda\setminus\{(i,g)\}}$ where $d_{jh} = d_h$ for every $(j,h)\in\Lambda\setminus\{(i,g)\}$, representing the treatment status for units other than (i,g) according to their group labels. In addition, we define $\mathbf{0}_{(i,g)}$ to be another value of $\{d_{jh}\}_{(j,h)\in\Lambda\setminus\{(i,g)\}}$ where $d_{jh} = 0$ for every $(j,h)\in\Lambda\setminus\{(i,g)\}$, representing the treatment status of no treatment for units other than (i,g).

These definitions lead to the following four types of potential outcomes that are relevant to our discussion:

- $Y_{igt}(d_g,\mathbf{d}_{(i,g)})$ corresponds to the *observed* treatment status where (i,g) is treated according to d_g, and other units (j,h) are treated according to d_h.

- $Y_{igt}(d_g, \mathbf{0}_{(i,g)})$ corresponds to a counterfactual treatment status where (i,g) is treated according to d_g, but all the other units (j,h) are untreated.
- $Y_{igt}(0, \mathbf{d}_{(i,g)})$ corresponds to a counterfactual treatment status where (i,g) is untreated, but other units (j,h) are treated according to d_h.
- $Y_{igt}(0, \mathbf{0}_{(i,g)})$ corresponds to a counterfactual treatment status where both (i,g) and all the other units (j,h) are untreated.

We assume that there is no anticipatory effect for these four types of potential outcomes, a standard assumption in DID models.

Assumption 2 (no anticipation)**.** *$Y_{igt}(d_{ig}, \mathbb{d}_{(i,g)}) = Y_{igt}(d_{ig}^t, \mathbb{d}_{(i,g)}^t)$ where $d_{ig} \in \{0, d_g\}$ and $\mathbb{d}_{(i,g)} \in \{\mathbf{0}_{(i,g)}, \mathbf{d}_{(i,g)}\}$.*

Under this assumption, the not-yet-treated group has the same potential outcomes as the "never-treated" group. Hence, we will refer to the group labelled ∞ as the never-treated group. Assumption 2 still allows the treatment duration to affect the potential outcomes.

Next, we introduce the parallel trend assumption, specifically for a linear DID model. We discuss nonlinear DID models in later sections.

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

$$\mathbb{E}(Y_{igt}(0^t, \mathbf{0}_{(i,g)}^t)|\alpha_{ig}) = \alpha_{ig} + \delta_t,$$

where $\alpha_{ig} \in \mathbb{R}$ is the unit fixed effect and $\delta_t \in \mathbb{R}$ is a common time effect.

Assumption 3 can also be expressed in a standard form commonly found in the literature on DID models [Borusyak et al., 2021]:

$$Y_{igt}(0^t, \mathbf{0}_{(i,g)}^t) = \alpha_{ig} + \delta_t + \varepsilon_{igt},$$

where $\mathbb{E}(\varepsilon_{igt}|\alpha_{ig}) = 0$ for every group g at time t.

We now introduce the estimand of interest, which is the average treatment effect on the treated (ATT) for group g at time t, denoted as $ATT(g,t)$. Without SUTVA, multiple definitions of $ATT(g,t)$ can be used. We first introduce $ATT(g,t)$ without accounting for any spillovers:

$$ATT_0(g,t) \equiv \mathbb{E}(Y_{igt}(d_g^t, \mathbf{0}_{(i,g)}^t) - Y_{igt}(0^t, \mathbf{0}_{(i,g)}^t)).$$

$ATT_0(g,t)$ captures the expected treatment effect when unit i is the only treated unit in the population, thereby excluding any spillover effects from the other units. In other words, $ATT_0(g,t)$ captures only the own direct effect from the treatment, illustrated by the solid edges in Figure 1, and corresponds to the standard definition of ATT when SUTVA holds. We can then define an aggregate ATT_0 by $ATT_0 = \sum_{g,t} w_{gt} ATT_0(g,t)$ where w_{gt} is a weight chosen by the econometrician (see Callaway and Sant'Anna [2020]).

We can also adopt an alternative definition of the ATT inclusive of the spillovers:

$$ATT_S(g,t) \equiv \mathbb{E}(Y_{igt}(d_g^t, \mathbf{d}_{(i,g)}^t) - Y_{igt}(0^t, \mathbf{0}_{(i,g)}^t)).$$

Relative to the $ATT_0(g,t)$, $ATT_S(g,t)$ incorporates the spillover effect(s) from other treated units. Note that, at time t, all units with group labels $g \leq t$ are treated. Therefore, $ATT_S(g,t)$ includes the spillover effect(s) from all units with group labels $g \leq t$.

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

$$AST(g,t) \equiv \mathbb{E}(Y_{igt}(d_g^t, \mathbf{d}_{(i,g)}^t) - Y_{igt}(d_g^t, \mathbf{0}_{(i,g)}^t)).$$

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

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

4 Identification

This section establishes the conditions for identifying the $ATT_0(g,t)$. The discussion 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 discuss 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 $t \geq g$, the parameter $ATT_0(g,t)$ is identified if and only if $\delta_t + AST(g,t)$ is identified.*

Proof. See Appendix. □

The proof of Theorem 1 shows that, for $t \geq g$:

$$\mathbb{E}(Y_{igt}) = \mathbb{E}(\alpha_{ig}) + \delta_t + ATT_0(g,t) + AST(g,t).$$

The intuition for Theorem 1 is that since $\mathbb{E}(\alpha_{ig})$ 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 (time effect) and $AST(g,t)$ (the average spillover effect on the treated). In general, Assumptions 1 to 3 are not sufficient to deliver identification of these two parameters. Note that $AST(g,t) = 0$ under SUTVA, in which case identification of the ATT only requires knowledge of the time-effects.

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 group g at time t such that $t \geq g$,*

$$Y_{igt}(d_g^t, \mathbf{d}_{(i,g)}^t) = Y_{igt}(d_g^t, \mathbf{0}_{(i,g)}^t).$$

This assumption requires that once a unit receives treatment, it no longer experiences spillover effects. This means the unit forfeits any spillovers it may have previously received and remains unaffected by spillovers from subsequently treated groups. As previously discussed, it will likely apply to many contexts (see Examples 1 and 2). Under Assumption 4, $AST(g,t) = 0$, and therefore $ATT_0(g,t) = ATT_S(g,t)$, unifying the definition of $ATT(g,t)$.[7]

Now, we state the second assumption.

For every group $g \in \mathcal{G}$, let $\Lambda_g^0 \subseteq \Lambda_g$ be a collection of units such that, for every untreated period $t < g$:

$$\mathbb{E}(Y_{igt}(0^t, \mathbf{d}_{(i,g)}^t)|\alpha_{ig}, (i,g) \in \Lambda_g^0) = \mathbb{E}(Y_{igt}(0^t, \mathbf{0}^t)|\alpha_{ig}, (i,g) \in \Lambda_g^0) = \alpha_{ig} + \delta_t. \quad (4)$$

Hence, the Λ_g^0 set consists of units within group g that are not affected by spillover effects while they are untreated.

Assumption 5 (Existence of never-treated units without spillover effects)**.** *Λ_∞^0 has a positive measure.*

This assumption states that there exists a nontrivial proportion of never-treated units that are not affected by spillovers, allowing for the identification of the time effects δ_t. In

[7] This not to say that $ATT_S(g,t)$ is never interesting when Assumption 4 does not hold. However, the researcher should be mindful that we can only identify $ATT_S(g,t)$ for the observed treatment status $\mathbf{d}_{(i,g)}^t$, weakening external validity. For example, we cannot make any statements about $ATT_S(g,t)$ when other groups are treated earlier than what is observed in the data.

practice, the econometrician may not have complete knowledge of Λ_g^0 and take a conservative approach by selecting the smaller subset of units strongly believed to be unaffected by spillovers, denoted by $\tilde{\Lambda}_g^0 \subseteq \Lambda_g^0$. We abuse notation and use Λ_g^0 interchangeably with $\tilde{\Lambda}_g^0$. All the results we discuss below apply to both Λ_g^0 and $\tilde{\Lambda}_g^0$.

Note that Assumption 5 does not impose any requirement about the size of Λ_g^0 for $g \neq \infty$. For instance, in the staggered adoption setting of Gonzalez-Navarro [2013] described in Example 2, the author used the never-treated Mexican states that are farthest from the treated states as controls. Alternatively, assuming that spillover effects occur only among adjacent states, Λ_g^0 can be set to be a nonempty set consisting of all units in group g that are not adjacent to any treated states until they are treated. In this case, the set of "controls" for time t, defined by $\bigcup_{g>t} \Lambda_g^0$, decreases in t as more states adopt the treatment over time, resulting in fewer untreated states that are not adjacent to any treated ones.[8]

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

Theorem 2. *Suppose that Assumptions 1 to 5 hold, and all units are untreated at $t = 1$. Then $ATT_0(g,t)$ is identified for all $t \geq g$.*

Proof. See Appendix. □

5 Estimation and inference

In this section, we discuss estimation and inference regarding $ATT_0(g,t)$ under Assumptions 1 to 5. Consider a balanced panel of T periods, where all units are untreated at $t = 1$. The units are indexed as $i = 1, \ldots, N_g$ for each group label $g \in \mathcal{G}$. Given the sets $\{\Lambda_g^0\}_{g \in \mathcal{G}}$, we define S_{igt} as a binary indicator denoting units affected by spillover effects. Let $S_{igt} = 0$ for all pre-treatment periods ($t < \min\{t | t \in \mathcal{G}\}$). For post-treatment periods ($t \geq \min\{t | t \in \mathcal{G}\}$), let

$$S_{igt} = \begin{cases} 0 & \text{if } (i,g) \in \Lambda_g^0 \text{ or } D_{igt} = 1 \\ 1 & \text{otherwise} \end{cases}$$

where S_{igt} equals 0 if unit (i,g) is either treated (Assumption 4) or belongs to Λ_g^0.

We first consider the case where Λ_∞^0 is the only nonempty set. We propose the following extension of Wooldridge [2022] as the estimation procedure. Define an extended

[8]The time effects could also be identified from variation in a different type of unit or from a different outcome. In Gonzalez-Navarro [2013], time effects could be estimated from non-protected car models (e.g. non-Ford) as long as they were not exposed to spillovers or from time variation in a different crime outcome (e.g. robbery) as long as it shows similar trends.

group label $\mathbf{g}$, created by partitioning the never-treated group ∞ into $(\infty, 0)$ and $(\infty, 1)$. For example, if $\mathcal{G} = \{2, \ldots, T, \infty\}$, the extended group label is given by:

$$\mathbf{g} \in \tilde{\mathcal{G}} \equiv \{2, \ldots, T, (\infty, 0), (\infty, 1)\},$$

where these new group labels $(\infty, 0)$ and $(\infty, 1)$ represent units not affected by spillovers (Λ^0_∞) and those affected $(\Lambda_\infty - \Lambda^0_\infty)$, respectively. The other group labels $\{2, \ldots, T\}$ are unchanged. With this, we also apply the definition of $Y_{i\mathbf{g}t}$, $D_{i\mathbf{g}t}$ and $S_{i\mathbf{g}t}$ to this extended group label.

We estimate the linear regression model where $Y_{i\mathbf{g}t}$ is the outcome variable, and the regressors are:

- indicators of $\mathbf{g}$ (the "extended group fixed effects"),
- indicators of t (the "time fixed effects"),
- interactions between indicators of $(\mathbf{g}, t)$ and $D_{i\mathbf{g}t}$, and
- interactions between indicators of $(\mathbf{g}, t)$ and $S_{i\mathbf{g}t}$.

$$\begin{aligned} Y_{i\mathbf{g}t} = \alpha_{\mathbf{g}} + \delta_t + \sum_{g' \in \mathcal{G} \setminus \{\infty\}} \sum_{t'=g'}^{T} \beta_{g't'} \cdot \mathbf{1}((\mathbf{g}, t) = (g', t')) \cdot D_{i\mathbf{g}t} \\ + \sum_{\mathbf{g}' \in \tilde{\mathcal{G}}} \sum_{t'=2}^{g'-1} \gamma_{\mathbf{g}'t'} \cdot \mathbf{1}((\mathbf{g}, t) = (\mathbf{g}', t')) \cdot S_{i\mathbf{g}t} + \varepsilon_{i\mathbf{g}t}. \end{aligned} \tag{5}$$

Then, the estimate of β_{gt} in this linear regression model, denoted by $\widehat{\beta}_{gt}$, is the estimate of $ATT_0(g, t)$.

Note that the regressors involve indicators of $\mathbf{g}$, the *group* fixed effect, as opposed to indicators of $(i, \mathbf{g})$, the *unit* fixed effect. This applies similarly to the treatment effects, where the regression model involves group-level treatment effects (β_{gt}) instead of unit-level treatment effects. This leads to simple steps for estimation and inference of $ATT_0(g, t)$, because its estimate $\widehat{\beta}_{gt}$ and standard error can be easily obtained using any software package running linear regressions. Furthermore, estimation and inference for an aggregate ATT is also simple because its estimate is given by $\sum_{g,t} w_{gt} \widehat{\beta}_{gt}$ and its standard error is computed straightforwardly 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 variances and covariances of $\widehat{\beta}_{gt}$'s are available in any software package.

Alternatively, the following extension of the imputation procedure proposed by [Borusyak et al., 2021] is a numerically equivalent method of obtaining β_{gt} in (5) :

1. Estimate the linear model
$$Y_{i\mathbf{g}t} = \alpha_{i\mathbf{g}} + \delta_t + \varepsilon_{i\mathbf{g}t},$$
using observations (i, g, t) such that $D_{igt} = 0$ and $S_{igt} = 0$. These observations consist of all observations in the pre-treatment periods $(t < \min\{t'|t' \in \mathcal{G}\})$ and observations for units (i, g) belonging to Λ^0_∞ across all post-treatment periods.

2. Let $\widehat{\alpha}_{i\mathbf{g}}$ and $\widehat{\delta}_t$ be the estimates of $\alpha_{i\mathbf{g}}$ and δ_t from the previous linear model. Impute the baseline outcome for unit $(i, \mathbf{g})$ at time t as
$$\widehat{Y}_{igt}(0^t, \mathbf{0}^t_{(i,\mathbf{g})}) = \widehat{\alpha}_{i\mathbf{g}} + \widehat{\delta}_t.$$

3. Estimate $ATT_0(g, t)$ for each $g \in \mathcal{G}$ by
$$\frac{1}{N_g}\sum_{i=1}^{N_g}\left[Y_{igt} - \widehat{Y}_{igt}(0^t, \mathbf{0}^t_{(i,g)})\right].$$

In this procedure, $ATT_0(g, t)$ is estimated by the average difference between the observed (treated) outcome and the counterfactual (baseline) outcome. The baseline outcome is estimated using observations in the control group that are unaffected by spillovers.

Note that the estimate of $ATT_0(g, t)$ in the imputation procedure equals to
$$\frac{1}{N_g}\sum_{i=1}^{N_g} Y_{igt} - \frac{1}{N_g}\sum_{i=1}^{N_g}\widehat{\alpha}_{i\mathbf{g}} - \widehat{\delta}_t.$$

The regression in (5) directly computes $(1/N_g)\sum_{i=1}^{N_g}\widehat{\alpha}_{i\mathbf{g}}$, and not individual $\widehat{\alpha}_{i\mathbf{g}}$, through the group-level fixed effect $\alpha_{\mathbf{g}}$ to estimate $ATT_0(g, t)$. The following proposition shows that, despite this simplification in estimation, the population regression of (5) correctly identifies $ATT_0(g, t)$ in the presence of unit-and-time level treatment effect heterogeneity. The consistency and asymptotic normality of $\widehat{\beta}_{gt}$ follows directly from the validity of the population regression.

Proposition 1. *Suppose that the assumptions of Theorem 2 hold. Consider the population regression of (5), and let β_{gt} be the population regression coefficient for the interaction between the indicator of (g, t) and D_{igt}. Then $\beta_{gt} = ATT_0(g, t)$.*

Proof. See Appendix. □

Next, we consider the case where Λ_g^0 might be non-empty for some $g \neq \infty$. We define an extended group label $\mathbf{g}$ by

$$\mathbf{g} \equiv (g, \lambda) \in \mathcal{G} \times \{0, 1\},$$

which partitions each group g further into $(g, 0)$ and $(g, 1)$, representing units not affected by spillovers (units within Λ_g^0) and those affected (units within $\Lambda_g - \Lambda_g^0$), respectively. We extend the definition of $Y_{i\mathbf{g}t}$, $D_{i\mathbf{g}t}$ and $S_{i\mathbf{g}t}$ to this new group label and define ATT_0 accordingly:

$$ATT_0(\mathbf{g}, t) = \mathbb{E}(Y_{i\mathbf{g}t}(d_g^t, \mathbf{0}_{(i,g)}^t) - Y_{i\mathbf{g}t}(0^t, \mathbf{0}_{(i,g)}^t)).$$

The aggregate ATT can then be defined as $ATT_0 = \sum_{\mathbf{g},t} w_{\mathbf{g}t} ATT_0(\mathbf{g}, t)$ where $w_{\mathbf{g}t}$ is a weight chosen by the econometrician. The previously described regression procedure in (5) can then be straightforwardly extended with this group label, except that the coefficients β_{gt} on the interactions between indicators of $(\mathbf{g}, t)$ and $D_{i\mathbf{g}t}$ are replaced by $\beta_{\mathbf{g}t}$ where $\beta_{\mathbf{g}t} = ATT_0(\mathbf{g}, t)$.

Lastly, if the data is an unbalanced panel, the population regression of (5) is no longer valid. It is still possible to implement the imputation-based estimation procedure of Borusyak et al. [2021], but the standard error will be conservative in general (see Borusyak et al., 2021, Section 4.3). In contrast, in the case of a balanced panel, the standard error computed from (5) maintains the size.

6 Extension to nonlinear DID models

This section focuses on situations where Y_{igt} is a count variable, such that the linear parallel trend condition (Assumption 3) does not hold. This extension contributes to the literature on nonlinear DID models [Wooldridge, 2023], expanding the applicability of our results to a wider array of empirical applications.

We introduce the following assumption regarding parallel trends in the context of count data.

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

$$\ln \mathbb{E}(Y_{igt}(0^t, \mathbf{0}_{(i,g)}^t) | \alpha_{ig}) = \alpha_{ig} + \delta_t.$$

By replicating the arguments in Theorems 1 and 2, we can show that $ATT_0(g, t)$ is identified under assumptions similar to those in Theorem 2.

In doing so, we abuse notation and define $\Lambda_g^0 \subseteq \Lambda_g$ for every group $g \in \mathcal{G}$ as a collection of units such that, for every untreated period $t < g$:

$$\ln \mathbb{E}(Y_{igt}(0^t, \mathbf{d}^t_{(i,g)})|\alpha_{ig}, (i,g) \in \Lambda^0_g) = \ln \mathbb{E}(Y_{igt}(0^t, \mathbf{0}^t)|\alpha_{ig}, (i,g) \in \Lambda^0_g) = \alpha_{ig} + \delta_t. \quad (6)$$

Theorem 3. *Suppose that Assumptions 1 and 2 and assumption 3' hold, and that all units are untreated at $t = 1$. Then, for $t \geq g$, the parameter $ATT_0(g,t)$ is identified if and only if $\mathbb{E}(\exp\{\alpha_{ig}\}|G_i = g) \cdot \exp\{\delta_t\} + AST(g,t)$ is identified.*

Proof. See Appendix. □

Theorem 4. *Suppose that Assumptions 1, 2, 4 and 5 and assumption 3' hold, and all units are untreated at $t = 1$. Then $ATT_0(g,t)$ is identified for all $t \geq g$.*

Proof. See Appendix. □

Note that, despite a nonlinear setting, the identification holds in a short panel setting, implying consistent estimation of $ATT_0(g,t)$ under the asymptotics where T remains fixed.

Let S_{igt} be defined as in previous sections, and consider a balanced panel of T periods where units are indexed as $i = 1, \ldots, N_g$ for each group label g, and all units are untreated at $t = 1$. Our parameter of interest is still $ATT_0(g,t)$. In the case of count data, the average treatment effect in terms of percentage changes is also often reported:

$$ATTP_0(g,t) = \frac{ATT_0(g,t)}{\mathbb{E}(Y_{igt}(0^t, \mathbf{0}^t_{(i,g)})|G_i = g)}.$$

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

The estimation and inference procedure discussed in Section 5 can be straightforwardly extended to the count data. For example, when Λ^0_g is empty for all $g \neq \infty$, we define the extended group label $\mathbf{g}$ to be as defined in Section 5 and we use the following simple estimation procedure that involves a parsimonious generalized linear model.

1. Estimate the Poisson regression model

$$\begin{aligned}\ln \mathbb{E}(Y_{i\mathbf{g}t}|\alpha_{\mathbf{g}}, D_{i\mathbf{g}t}, S_{i\mathbf{g}t}) = \alpha_{\mathbf{g}} + \delta_t + \sum_{g' \in \mathcal{G}\setminus\{\infty\}} \sum_{t'=g'}^{T} \beta_{g't'} \cdot \mathbf{1}((\mathbf{g}, t) = (g', t')) \cdot D_{i\mathbf{g}t} \\ + \sum_{\mathbf{g}' \in \tilde{\mathcal{G}}} \sum_{t'=2}^{g'-1} \gamma_{\mathbf{g}'t'} \cdot \mathbf{1}((\mathbf{g}, t) = (\mathbf{g}', t')) \cdot S_{i\mathbf{g}t},\end{aligned} \quad (7)$$

   that is, the Poisson regression of Y_{igt} on:

   - indicators of $\mathbf{g}$ (the "extended group fixed effects"),

- indicators of t (the "time fixed effects"),
- interactions between indicators of $(\mathbf{g}, t)$ and $D_{i\mathbf{g}t}$, and
- interactions between indicators of $(\mathbf{g}, t)$ and $S_{i\mathbf{g}t}$.

Let $\widehat{\alpha}_{\mathbf{g}}$, $\widehat{\delta}_t$, and $\widehat{\beta}_{gt}$ be the estimates of $\alpha_{\mathbf{g}}$, δ_t, and β_{gt} from this model, respectively.

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

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

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

The validity of the population regression of (7) can be shown by an immediate extension of Proposition 1, and we omit the proof here. Then the consistency and asymptotic normality of $\widehat{ATT}(g,t)$ and $ATTP(g,t)$ follow directly from the validity of the population regression.

Note that most software packages that run Poisson regressions produce standard errors of $(\widehat{\alpha}_{\mathbf{g}}, \widehat{\delta}_t, \widehat{\beta}_{gt})$ based on the maximum likelihood. This assumes that the distribution of $Y_{i\mathbf{g}t}(0^t, \mathbf{0}^t_{(i,g)})$ conditional on $\alpha_{i\mathbf{g}}$ follows Poisson distribution (as opposed to only specifying its mean as in Assumption 3'), ruling out heteroskedasticity. The estimates' standard errors that rely on Assumption 3', therefore allowing for heteroskedasticity, can be obtained by interpreting them as the quasi maximum likelihood estimator (QMLE). Specifically, let θ be the vector of all coefficients in the Poisson regression (i.e., all of α_g, δ_t, β_{gt} and κ_{gt}), $\hat{\theta}$ be their maximum likelihood estimates (i.e., all of $\widehat{\alpha}_{\mathbf{g}}$, $\widehat{\delta}_t$, $\widehat{\beta}_{gt}$ and $\widehat{\gamma}_{\mathbf{g}t}$), and $X_{i\mathbf{g}t}$ be the vector of all regressors (i.e, all of the indicators of $\mathbf{g}$, indicators of t, interactions between indicators of $(\mathbf{g}, t)$ and $D_{i\mathbf{g}t}$, and interactions between indicators of $(\mathbf{g}, t)$ and $S_{i\mathbf{g}t}$). Define

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

be the clustered outer product of the score function, where $\widehat{Y}_{i\mathbf{g}t} = \exp\{X'_{i\mathbf{g}t}\hat{\theta}\}$ is the "fitted value" of $Y_{i\mathbf{g}t}$ in the Poisson regression[9] and $\{\Lambda^c\}_{c=1}^{C}$ refers to the partition of units according to which the units are clustered. In addition, define

$$\mathcal{H} = \sum_{(i,\mathbf{g})\in\cup_{c=1}^{C}\Lambda^c} \sum_{t} X_{i\mathbf{g}t}X'_{i\mathbf{g}t}\widehat{Y}_{i\mathbf{g}t}$$

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

be the negative Hessian function. The variance-covariance matrix of $\hat{\theta}$, when we interpret $\hat{\theta}$ as the QMLE, is given by

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

This variance-covariance matrix can then be used to compute the standard error of aggregate ATT and $ATTP$ estimates via the delta method.

7 Application to auto theft prevention policy

In this section, we apply our method to the dataset of Gonzalez-Navarro [2013], who studied the effects of installing Lojack, a device tracking cars in the event of theft.

Lojack 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. By 2004, it was installed in 32 *model* $\times$ *state* combinations. The dataset of Gonzalez-Navarro [2013] provides comprehensive information on car theft for each *model* $\times$ *state* $\times$ *vintage* (referring to the year of the car model) combination, for each calendar year. In our analysis, we use the labels 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 potential sources of spillover. Firstly, there is the possibility of spillover to non-Lojack car models within the same state. Since it was known that Lojack was installed in a particular car model and state, criminals involved in auto theft within that state might shift their focus from Lojack-installed car models to non-Lojack car models. Secondly, there could also be geographical spillover to the same car model in other states where Lojack is not installed.

Gonzalez-Navarro [2013] estimates the treatment effect of Lojack installation, considering these spillover effects. However, the estimation does not account for heterogeneity in the treatment effect. In this section, we apply our method to the original dataset and provide estimates of the treatment effect across various combinations of g and t, which also documents heterogeneity in the treatment effect.

Once Lojack is installed in a specific combination of car model and state for a particular vintage, it is subsequently installed in all subsequent vintages. This setup allows us to treat the situation as a staggered adoption design, where the unit of analysis is *model* (m) $\times$ *state* (s) $\times$ *age* (a), with *age* defined as the difference between the calendar year (t) and the vintage year (v), meaning that $a = t - v$.

We define the binary treatment indicator as D_{msat}. For example, consider newly released ($age = 0$) Ford Windstar models in Jalisco. For this unit, Lojack is installed in new models in Jalisco starting from 2001. In this case, we have $D_{Windstar,Jalisco,0,t} = 1$ for $t \geq 2001$.

Our method requires Assumptions 4 and 5. Assumption 4 implies that once a *model* $\times$ *state* $\times$ *age* combination is Lojack-installed, it does not experience spillover effects. Generally, we can reasonably expect thieves to shift their target only to cars that are not protected, which are either non-Lojack models within the same state or models in states without Lojack installation, or to other goods. Thus, it is reasonable to assume that Assumption 4 holds.

Assumption Assumption 5 requires that some *model* (m) $\times$ *state* (s) are not affected. Gonzalez-Navarro [2013] finds evidence that this is the case. That is, states far from the treated ones and different car models do not seem exposed to spillovers.

Let Y_{msat} represent the number of auto thefts that occurred in a given calendar year t for a specific combination of *model* (m), *state* (s), and *age* (a). We consider the two models for the untreated outcomes. First, we consider a linear parallel trend:

$$\mathbb{E}(Y_{msat}(0^t, \mathbf{0}^t_{(msa,t)})|\alpha_{msa}) = \alpha_{msa} + \delta_t.$$

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

$$\ln \mathbb{E}(Y_{msat}(0^t, \mathbf{0}^t_{(msa,t)})|\alpha_{msa}) = \alpha_{msa} + \delta_t.$$

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

We define S_{msat} to be a binary indicator, which is equal to 1 if $D_{msat} = 0$ and the state s is not adjacent to any state s' that contains at least one combination (m', s', a') where $D_{m's'a't'} = 1$ (that is, s is not adjacent to any of the treated states). We estimate the following aggregate ATTs:

$$\begin{aligned}
ATT &= \sum_{(g,t):t\geq g\geq 2} \frac{N_g}{\sum_{(g,t):t\geq g\geq 2} N_g} ATT_0(g,t), \\
ATT^0 &= \sum_{g:g\leq T} \frac{N_g}{\sum_{g:g\leq T} N_g} ATT_0(g,g), \\
ATT^1 &= \sum_{g:g\leq T-1} \frac{N_g}{\sum_{g:g\leq T-1} N_g} ATT_0(g,g+1), \\
ATT^2 &= \sum_{g:g\leq T-2} \frac{N_g}{\sum_{g:g\leq T-2} N_g} ATT_0(g,g+2).
\end{aligned}$$

Here, ATT represents the sample-size-weighted average of all $ATT(g,t)$ values across g and t. ATT^k represents the sample-size-weighted average of the ATTs for the k-th year after installation of Lojack. For example, ATT^0 represents the effect in the same year when Lojack is installed, ATT_1 represents the effect one year after the installation of Lojack, and so forth.

Table 1 displays estimates of the ATT values using the linear and Poisson specifications. Under the linear model, the impact of Lojack installation on theft reduction becomes increasingly pronounced as more time elapses following installation. Compared to the overall average reduction rate of 64% for the linear model and 61% for the Poisson model, it becomes evident that the effect intensifies starting one year after the installation of Lojack. These findings remain robust across linear and Poisson specifications.

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

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

For comparison, we also report the estimated ATTs for the misspecified linear models. Specifically, we consider the TWFE specification that does not account for staggered adoption and the specification of Borusyak et al. [2021], Wooldridge [2022] that do not account for spillover effects. The estimates are presented in Table 2. We observe that the TWFE regression estimate is similar to the estimates in the first line of Table 1, whereas the policy estimates that do not account for spillover effects are biased upward compared to estimates in Table 1. This is what we would expect in the presence of displacement, where installing Lojack in a treated unit increases theft for units without Lojack.

8 Monte Carlo

We study the finite sample properties of our estimator in a simulated dataset with either a contemporaneous or staggered adoption design. No group is treated in period 1, and the set of untreated units is always non-empty. Within a group, there are κ units. Like the estimators suggested in the Weighting literature, our approach trades precision for bias. In the absence of actual spillovers, excluding group-time data points potentially exposed to spillovers is inefficient. However, when spillovers are present, those cells introduce bias.

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

Table 2: Estimates of the aggregate ATTs using the TWFE specification that does not account for staggered adoption (the "TWFE" columns), and the specification of Borusyak et al. [2021], Wooldridge [2022] that do not account for spillover effects (the "NoSpillover" columns) for linear and Poisson specifications. The "Reduction" columns stands for the reduction rate, which is calculated using the formula for computing $ATTP$.

When κ is small, the lack of precision might be too costly if the bias from spillover is not large. To highlight this, we use a tuning parameter ($0 \leq \rho \leq 1$): $\rho = 0$ corresponds to no spillovers, whereas $\rho = 1$ corresponds to full displacement. We do not consider diffusion since the source of the bias is the same, only of the opposite sign.

8.1 DGP

Our Data Generating Process embeds the identification assumptions Assumptions 1 to 5, with Assumption 3 being replaced by 3' if the data follows a Poisson distribution. The Data Generating Process is given by:

$$\mathbb{E}[Y_{it}] = F\left(\alpha_i + \delta_t + \beta_{it} D_{it} + (1 - D_{it}) \cdot \sum_{j \neq i} D_{jt} \cdot \eta_{it}^j \cdot S_{it}^j\right)$$

where $F(.)$ is the identity function or Poisson. We parametrize the DGP such that

- Groups are named alphabetically, A, B, ..., Z. Group A is treated first in period 2, group B is treated second in period 3, and so on. Groups W and Z are never treated, but units in group W are immune to spillovers, while units in group Z are not. Each group has the same number of units, and within a group, units are identical.

- Starting in period 2, a set of units (group) is treated in each period until $t = T$.

Hence, the number of groups is a function of the number of time periods, i.e. $G = T + 1$. E.g. if $T = 3$, groups A, B, W and Z exist.

- Unit fixed effects: $\alpha_i = 26 - g + 1$, where g is the group index starting from $A = 1$. The first group to be treated has the largest fixed effect, leading to selection into treatment. The number 26 corresponds to the number of letters in the alphabet. Unit fixed effects are homogeneous within a group: $\alpha_{ig} = \alpha_{i'g}\ \forall i, i' \in g$.[10]

- Time effects: $\delta_t = \bar{\alpha} \times 0.1 \times ((t - 1) + \sin(t))$. The time effect is divided into two parts, both proportional to the average unit fixed effect: a linear upward trend $(t - 1)$ and a period-specific effect that follows a $\sin(.)$ pattern

- Direct policy effect: $\beta_{it} = \frac{0.5\alpha_i}{t}$. The policy effect is heterogeneous across groups and time but homogeneous within a group. It is largest for the first group with the highest μ_i, decreasing over time. This parametrization implies sorting on gain since α_i also correlates with treatment timing.

- Spillovers (Displacement): $\eta_{it}^{j} = \frac{\beta_{jt} \cdot \rho}{\sum_{i \neq j} S_{it}^{j}}$, where the denominator gives the number of units exposed to unit j spillovers. Hence, the spillover from unit j generated by unit j's direct policy effect spills evenly among the untreated groups at time t. The magnitude of the spillover is regulated by the tuning parameter ρ.

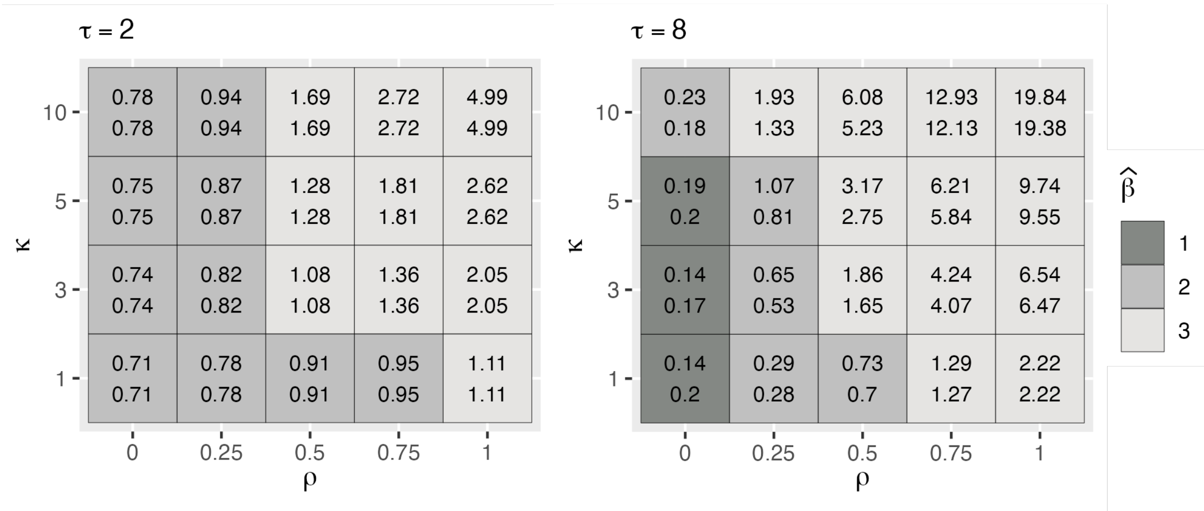
This parametrization is fixed over the simulations, so $\mathbb{E}[Y_{it}]$ is constant. If $F(\lambda)$ is the identity function, randomness is given by different draws of $\epsilon_{it} \sim N(0, \max(\alpha_i)/10)$ added to $\mathbb{E}[Y_{it}]$. If $F(.)$ is Poisson, then randomness is given by random generation from the Poisson distribution with mean $\mathbb{E}[Y_{it}]$. The target estimand is $\mathbb{E}[\beta_{it} | D_{it} = 1]$. We compare the mean Absolute Bias and the Mean Squared Error across the:

($\widehat{\beta}_1$) TWFE estimator, which does not account either for the staggered treatment or for spillovers.

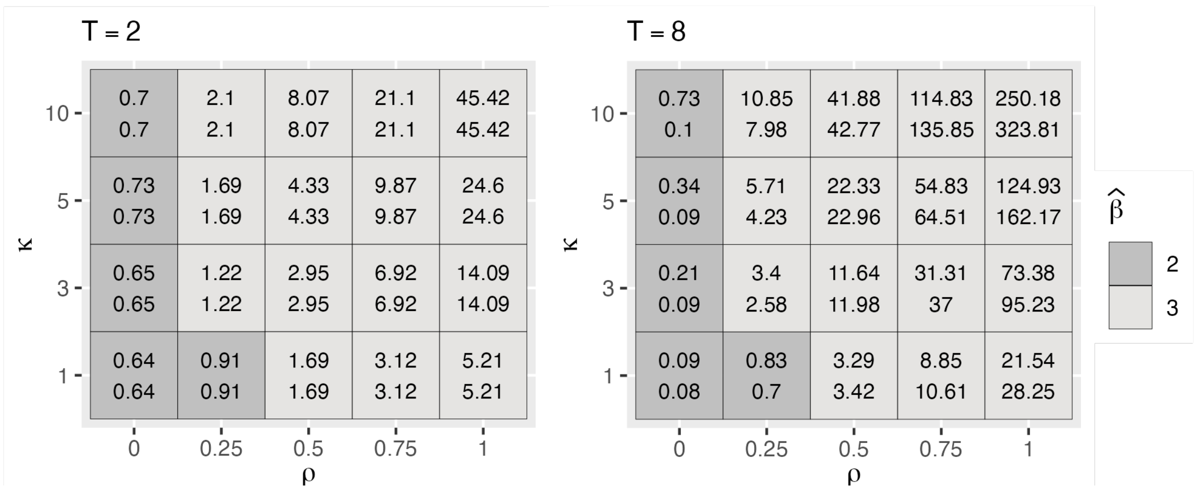
($\widehat{\beta}_2$) Imputation estimator suggested in Borusyak et al. [2021] which accounts for treatment effects heterogeneity by excluding group-time observation after treatment, but does not account for spillovers. This estimator is numerically equivalent to Wooldridge [2022] extended TWFE.

($\widehat{\beta}_3$) Our estimator, which uses imputation but also excludes group-time observation potentially exposed to spillovers.

[10] In the case of Poisson data we replace α_i with $\tilde{\alpha}_i = \log(\alpha_i)$.



(a) Identity



(b) Poisson

Figure 4: Monte Carlo Simulation - comparing MSE.
Cell background color is based on the best-performing estimator. Numbers in cells correspond to MSE ratios $\frac{MSE_1}{MSE_3}$ and $\frac{MSE_2}{MSE_3}$ respectively. Subscript refers to: (1) TWFE, (2) BJS estimator and (3) our approach accounting for spillovers.

Table 3: Monte Carlo Simulation - identity

ρ	T	κ	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, (2) BJS estimator and (2) our approach accounting for spillovers. Lowest value across estimators in bold.

Figure 4a and Table 3 show the linear DGP results. The Figure compares the MSE across the three estimators to illustrate their relative performance under different scenarios, while the Table provides MSE and Absolute Bias values. Note that When $T = 2$, the TWFE and Imputation estimators are identical since treatment is not staggered. Overall, the best-performing estimator depends on the degree of spillovers, staggered treatment,

Table 4: Monte Carlo Simulation - Poisson

ρ	T	κ	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, (2) BJS estimator and (2) our approach accounting for spillovers. Lowest value across estimators in bold.

and how many units are 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 imputation estimator becomes the best-performing one. It corrects for staggered treatment and is not too biased. Once spillovers are not negligible and observations increase, our estimator has the lowest MSE, often by a large margin. Our estimator also performs better as treatment becomes more staggered ($T = 8$) since cumulative spillovers to untreated units will strongly affect the counterfactual. Figure 4b and Table 4 show the results for the Poisson DGP. Our estimator performs even better relative to the TWFE and the Imputation ones.

9 Conclusion

We establish identifying assumptions, and estimation procedures, for the ATT in a Difference-in-Differences setting with staggered treatment adoption and spillovers. Aside from the canonical Difference-in-Differences assumptions, identification requires that once a unit is treated, it does not experience spillovers, past, present, or future. This assumption, which is likely to hold in many contexts, unifies the ATTs simplifying policy evaluation and joining with the definition of ATT under SUTVA.

To estimate the ATT we extend the TWFE model approach of Wooldridge [2022] and the imputation approach of Borusyak et al. [2021] to account for spillovers in linear and non-linear settings. In the case of a balanced panel, Wooldridge [2022] 's approach can also be used to directly calculate the ATT's standard error. We then revisit Gonzalez-Navarro [2013], who studied the effects of installing a device tracking cars in the event of theft. 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. Point identification requires excluding units affected by spillovers from estimation. However, such exclusions may have minimal benefits if spillovers are negligible, or if the number of units is small, all the while sacrificing precision. We compare different estimators: the traditional TWFE estimator, which overlooks both staggered adoption and spillovers; the Wooldridge [2022] and Borusyak et al. [2021] estimators, which consider staggered adoption but not spillovers; and our proposed estimator, which addresses both factors. Our estimator proves to be highly competitive in a majority of scenarios.

References

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

Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting event study designs: Robust and efficient estimation. Technical report, arXiv.org, 2021.

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

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

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

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

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

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

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

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

László Mátyás and Patrick Sevestre. *The econometrics of panel data: fundamentals and recent developments in theory and practice*. Springer, 2008.

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

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

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

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

A Proofs

A.1 Proof of Theorem 1

Under Assumption 2, we can write, for each group g at time t,

$$\begin{aligned}
Y_{igt} &= Y_{igt}(d_g^t, \mathbf{d}_{(i,g)}^t) \\
&= Y_{igt}(0^t, \mathbf{0}_{(i,g)}^t) + \left[Y_{igt}(d_g^t, \mathbf{0}_{(i,g)}^t) - Y_{igt}(0^t, \mathbf{0}_{(i,g)}^t)\right] + \left[Y_{igt}(d_g^t, \mathbf{d}_{(i,g)}^t) - Y_{igt}(d_g^t, \mathbf{0}_{(i,g)}^t)\right] \\
&= \alpha_{ig} + \delta_t + \left[Y_{igt}(d_g^t, \mathbf{0}_{(i,g)}^t) - Y_{igt}(0^t, \mathbf{0}_{(i,g)}^t)\right] + \left[Y_{igt}(d_g^t, \mathbf{d}_{(i,g)}^t) - Y_{igt}(d_g^t, \mathbf{0}_{(i,g)}^t)\right] + \varepsilon_{igt}.
\end{aligned}$$

Define

$$\begin{aligned}
\beta_{igt} &= Y_{igt}(d_g^t, \mathbf{0}_{(i,g)}^t) - Y_{igt}(0^t, \mathbf{0}_{(i,g)}^t), \\
\gamma_{igt} &= Y_{igt}(d_g^t, \mathbf{d}_{(i,g)}^t) - Y_{igt}(d_g^t, \mathbf{0}_{(i,g)}^t).
\end{aligned}$$

We can then write

$$Y_{igt} = \alpha_{ig} + \delta_t + \beta_{igt} + \gamma_{igt} + \varepsilon_{igt}.$$

The parameter of interest $ATT_0(g,t)$ is then given by

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

and $AST(g,t)$ is given by

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

Then, under Assumption 3, we can write, for every group g at time t,

$$\mathbb{E}(Y_{igt}) = \mathbb{E}(\alpha_{ig}) + \delta_t + ATT_0(g,t) + AST(g,t).$$

Now we show that $ATT_0(g,t)$ is identified if and only if $\delta_t + AST(g,t)$ is identified. First, suppose that $\delta_t + AST(g,t)$ is identified. Let d_0 be the identified value. Then we can write

$$\mathbb{E}(Y_{igt}) = \mathbb{E}(\alpha_{ig}) + d_0 + ATT_0(g,t).$$

Note that $\mathbb{E}(\alpha_{ig})$ is identified from the data for group g at $t = 1$:

$$\mathbb{E}(Y_{ig1}) = \mathbb{E}(\alpha_{ig}). \tag{8}$$

We can then write

$$ATT_0(g,t) = \mathbb{E}(Y_{igt}) - d_0 - \mathbb{E}(Y_{ig1}).$$

This shows that $ATT_0(g,t)$ is identified.

Next, suppose that $ATT_0(g,t)$ is identified. Let b_0 be the identified value. Then:

$$\mathbb{E}(Y_{igt}) = \mathbb{E}(\alpha_{ig}) + \delta_t + b_0 + AST(g,t).$$

Using (8), we can write

$$\delta_t + AST(g,t) = \mathbb{E}(Y_{igt}) - b_0 - \mathbb{E}(Y_{ig1}).$$

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

A.2 Proof of Theorem 2

By Theorem 1, it suffices to show that $\delta_t + AST(g,t)$ is identified. First, Assumption 4 implies that $AST(g,t) = 0$. In addition, for every group g at time $t \geq 2$, Assumptions 3 and 5 jointly implies:

$$\mathbb{E}(Y_{igt}|(i,g) \in \Lambda_t^0) - \mathbb{E}(Y_{ig1}) = \mathbb{E}(Y_{igt}(0^t, \mathbf{d}_{(i,g)}^t)|(i,g) \in \Lambda_t^0) - \mathbb{E}(\alpha_{ig}) = \delta_t,$$

which implies that δ_t is identified for every $t \geq 2$. Consequently, it follows that

$$\delta_t + AST(g,t) = \mathbb{E}(Y_{igt}|(i,g) \in \Lambda_t^0) - \mathbb{E}(Y_{ig1}) + 0,$$

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

A.3 Proof of Proposition 1

As in the proof of Theorem 1, under Assumption 2, we can write, for each group g at time t,

$$\begin{aligned}
Y_{igt} &= Y_{igt}(d_g^t, \mathbf{d}_{(i,g)}^t) \\
&= Y_{igt}(0^t, \mathbf{0}_{(i,g)}^t) + \left[Y_{igt}(d_g^t, \mathbf{0}_{(i,g)}^t) - Y_{igt}(0^t, \mathbf{0}_{(i,g)}^t)\right] + \left[Y_{igt}(d_g^t, \mathbf{d}_{(i,g)}^t) - Y_{igt}(d_g^t, \mathbf{0}_{(i,g)}^t)\right] \\
&= \alpha_{ig} + \delta_t + \left[Y_{igt}(d_g^t, \mathbf{0}_{(i,g)}^t) - Y_{igt}(0^t, \mathbf{0}_{(i,g)}^t)\right] + \left[Y_{igt}(d_g^t, \mathbf{d}_{(i,g)}^t) - Y_{igt}(d_g^t, \mathbf{0}_{(i,g)}^t)\right] + \varepsilon_{igt}.
\end{aligned}$$

Define

$$\begin{aligned}
\beta_{igt} &= Y_{igt}(d_g^t, \mathbf{0}_{(i,g)}^t) - Y_{igt}(0^t, \mathbf{0}_{(i,g)}^t), \\
\gamma_{igt} &= Y_{igt}(d_g^t, \mathbf{d}_{(i,g)}^t) - Y_{igt}(d_g^t, \mathbf{0}_{(i,g)}^t).
\end{aligned}$$

We can then write

$$Y_{igt} = \alpha_{ig} + \delta_t + \beta_{igt} + \gamma_{igt} + \varepsilon_{igt}.$$

By the definition of Λ_∞^0, it follows that

$$\mathbb{E}(\varepsilon_{i,\infty,t}|(i,\infty) \in \Lambda_\infty^0) = 0.$$

In addition, since $\mathbb{E}(\varepsilon_{i,\infty,t}) = 0$ by Assumption 3, it follows that

$$\mathbb{E}(\varepsilon_{i,\infty,t}|(i,\infty) \in \Lambda_\infty - \Lambda_\infty^0) = 0.$$

Therefore, with the extended group label $\mathbf{g} \in \{2,\ldots,T,(\infty,0),(\infty,1)\}$, we can write

$$Y_{i\mathbf{g}t} = \alpha_{i\mathbf{g}} + \delta_t + \beta_{i\mathbf{g}t} + \gamma_{i\mathbf{g}t} + \varepsilon_{i\mathbf{g}t},$$

where $\mathbb{E}(\varepsilon_{i\mathbf{g}t}) = 0$ for every $\mathbf{g}$.

Now, for each $\mathbf{g} \in \{2,\ldots,T\}$, we have

$$\begin{aligned} Y_{i\mathbf{g}1} &= \alpha_{i\mathbf{g}} + \varepsilon_{i\mathbf{g}1}, \\ Y_{i\mathbf{g}t} &= \alpha_{i\mathbf{g}} + \delta_t + \gamma_{i\mathbf{g}t} + \varepsilon_{i\mathbf{g}t} \quad \text{for} \quad 2 \le t < g, \\ Y_{i\mathbf{g}t} &= \alpha_{i\mathbf{g}} + \delta_t + \beta_{i\mathbf{g}t} + \varepsilon_{i\mathbf{g}t} \quad \text{for} \quad g \le t \le T, \end{aligned}$$

where we used the following to derive the above equations:

- For the first equality, we used that all units are untreated at $t = 1$ and that δ_1 is set to 0 for normalization.
- For the second equality, we used that the units are untreated at $t < g$, in which case $\beta_{igt} = 0$.
- For the third equality, we used that the units are treated at $t \ge g$, in which case $\gamma_{i\mathbf{g}t} = 0$ because the unit is not exposed to spillover effects by Assumption 4.

We can then combine the three equations and write them as one equation where we write, for each $\mathbf{g} \in \{2,\ldots,T\}$:

$$Y_{i\mathbf{g}t} = \alpha_{i\mathbf{g}} + \delta_t + \sum_{t'=g}^{T} \mathbf{1}(t=t')\beta_{i\mathbf{g}t} + \sum_{t'=2}^{g-1} \mathbf{1}(t=t')\gamma_{i\mathbf{g}t} + \varepsilon_{i\mathbf{g}t},$$

which we can further write

$$Y_{i\mathbf{g}t} = \alpha_{i\mathbf{g}} + \delta_t + \sum_{t'=g}^{T} \mathbf{1}(t=t')D_{i\mathbf{g}t}\beta_{i\mathbf{g}t} + \sum_{t'=2}^{g-1} \mathbf{1}(t=t')S_{i\mathbf{g}t}\gamma_{i\mathbf{g}t} + \varepsilon_{i\mathbf{g}t}, \tag{9}$$

since $D_{i\mathbf{g}t} = 1$ for $t \ge g$ and $S_{i\mathbf{g}t} = 1$ for $2 \le t < g$ for each $\mathbf{g} \in \{2,\ldots,T\}$ by definition.

Now, for the group label $\mathbf{g} \in \{(\infty, 0), (\infty, 1)\}$, we can write

$$\begin{aligned}
Y_{i,(\infty,0),1} &= \alpha_{i\mathbf{g}} + \varepsilon_{i\mathbf{g}1}, \\
Y_{i,(\infty,0),t} &= \alpha_{i\mathbf{g}} + \delta_t + \varepsilon_{i\mathbf{g}t} \quad \text{for} \quad 2 \le t, \\
Y_{i,(\infty,1),1} &= \alpha_{i\mathbf{g}} + \varepsilon_{i\mathbf{g}1}, \\
Y_{i,(\infty,1),t} &= \alpha_{i\mathbf{g}} + \delta_t + \gamma_{i\mathbf{g}t} + \varepsilon_{i\mathbf{g}t} \quad \text{for} \quad 2 \le t,
\end{aligned}$$

which can be summarized as

$$\begin{aligned}
Y_{i,(\infty,0),t} &= \alpha_{i,(\infty,0)} + \delta_t + \varepsilon_{i\mathbf{g}t}, \\
Y_{i,(\infty,1),t} &= \alpha_{i,(\infty,1)} + \delta_t + \sum_{t'=2}^{T} \mathbf{1}(t = t')\gamma_{i\mathbf{g}t} + \varepsilon_{i\mathbf{g}t},
\end{aligned}$$

Then, these can be further written as, using that $D_{i\mathbf{g}t} = 0$ for $\mathbf{g} \in \{(\infty, 0), (\infty, 1)\}$ and $S_{i\mathbf{g}t} = 0$ for $(i, \mathbf{g}) \in \Lambda^0_\infty$,

$$Y_{i\mathbf{g}t} = \alpha_{i\mathbf{g}} + \delta_t + \sum_{t'=g}^{T} \mathbf{1}(t = t')D_{i\mathbf{g}t}\beta_{i\mathbf{g}t} + \sum_{t'=2}^{g-1} \mathbf{1}(t = t')S_{i\mathbf{g}t}\gamma_{i\mathbf{g}t} + \varepsilon_{i\mathbf{g}t}, \tag{10}$$

where we define $\sum_{t'=2}^{g-1}$ as $\sum_{t'=2}^{T}$ when $g = \infty$.

We can then combine (9) and (10) and write them as

$$\begin{aligned}
Y_{i\mathbf{g}t} &= \sum_{\mathbf{g}'} \mathbf{1}(\mathbf{g} = \mathbf{g}') \left(\alpha_{i\mathbf{g}} + \delta_t + \sum_{t'=g}^{T} \mathbf{1}(t = t')D_{i\mathbf{g}t}\beta_{i\mathbf{g}t} + \sum_{t'=2}^{g-1} \mathbf{1}(t = t')S_{i\mathbf{g}t}\gamma_{i\mathbf{g}t} + \varepsilon_{i\mathbf{g}t} \right) \\
&= \alpha_{i\mathbf{g}'} + \delta_t + \sum_{\mathbf{g}'} \sum_{t'=g}^{T} \mathbf{1}(\mathbf{g} = \mathbf{g}')\mathbf{1}(t = t')D_{i\mathbf{g}t}\beta_{i\mathbf{g}t} + \sum_{\mathbf{g}'} \sum_{t'=2}^{g-1} \mathbf{1}(\mathbf{g} = \mathbf{g}')\mathbf{1}(t = t')S_{i\mathbf{g}t}\gamma_{i\mathbf{g}t} + \varepsilon_{i\mathbf{g}t}.
\end{aligned} \tag{11}$$

Let $\mathbf{X}$ be the vector of regressors for the regression in (5). Note first that there exists a one-to-one mapping between the sequence $(D_{i\mathbf{g}t}, S_{i\mathbf{g}t})_{t=1}^{T}$ and the extended group label $\mathbf{g}$, because $(D_{i\mathbf{g}t})_{t=1}^{T}$ identifies the group label $g \in \{2, \dots, T\}$ and $(S_{i\mathbf{g}t})_{t=1}^{T}$ separates $(\infty, 0)$ and $(\infty, 1)$. This implies that

$$\overline{\mathbb{E}}(Y_{i\mathbf{g}t}|\mathbf{X}) = \overline{\mathbb{E}}(Y_{i\mathbf{g}t}|\mathbf{X}, \mathbf{g}) = \mathbb{E}(Y_{i\mathbf{g}t}|\mathbf{X}).$$

where $\overline{\mathbb{E}}$ refers to the expectation across all groups, while $\mathbb{E}$ refers to the expectation across i for a given $\mathbf{g}$. Then, by (11):

$$\mathbb{E}(Y_{\mathbf{igt}}|\mathbf{X}) = \mathbb{E}(\alpha_{\mathbf{ig'}})+\delta_t+\sum_{\mathbf{g'}}\sum_{t'=g}^{T}\mathbf{1}(\mathbf{g}=\mathbf{g'})\mathbf{1}(t=t')D_{\mathbf{igt}}\mathbb{E}(\beta_{\mathbf{igt}})+\sum_{\mathbf{g'}}\sum_{t'=2}^{g-1}\mathbf{1}(\mathbf{g}=\mathbf{g'})\mathbf{1}(t=t')S_{\mathbf{igt}}\mathbb{E}(\gamma_{\mathbf{igt}}),$$

which shows that the coefficient on $\mathbf{1}(\mathbf{g}=\mathbf{g'})\mathbf{1}(t=t')D_{\mathbf{igt}}$ for $g \in \{2,\ldots,T\}$ is

$$\mathbb{E}(\beta_{igt}) = \mathbb{E}(Y_{igt}(d_g^t, \mathbf{0}_{(i,g)}^t) - Y_{igt}(0^t, \mathbf{0}_{(i,g)}^t)),$$

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

A.4 Proof of Theorem 3

Similarly to the proof of Theorem 1, we can write, for each $(i,g) \in \Lambda$ and $t \leq T$,

$$\begin{aligned} Y_{igt} &= Y_{igt}(d_g^t, \mathbf{d}_{(i,g)}^t) \\ &= Y_{igt}(0^t, \mathbf{0}_{(i,g)}^t) + \left[Y_{igt}(d_g^t, \mathbf{0}_{(i,g)}^t) - Y_{igt}(0^t, \mathbf{0}_{(i,g)}^t)\right] + \left[Y_{igt}(d_g^t, \mathbf{d}_{(i,g)}^t) - Y_{igt}(d_g^t, \mathbf{0}_{(i,g)}^t)\right]. \end{aligned}$$

Then, under Assumption 3', we can write

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

where

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

and

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

Then, by replicating the arguments in Theorem 1, it is straightforward to show that $ATT_0(g,t)$ is identified if and only if $\mathbb{E}(\exp\{\alpha_{ig}\}|G_i = g)\exp\{\delta_t\} + AST(g,t)$ is identified.

A.5 Proof of Theorem 4

By Theorem 3, it suffices to show that $\mathbb{E}(\exp\{\alpha_{ig}\}|G_i = g) \cdot \exp\{\delta_t\} + AST(g,t)$ is identified. In what follows, we show that the three objects $\mathbb{E}(\exp\{\alpha_{ig}\}|G_i = g)$, $\exp\{\delta_t\}$, and $AST(g,t)$ are identified separately. First, Assumption 4 assumes that $AST(g,t) = 0$. Second, to show that $\exp\{\delta_t\}$ is identified, consider the units (i,g) such that $(i,g) \in \Lambda_\infty^0$. These are untreated units that are not affected by spillover effects. For these units, the following moment equality holds for any t under Assumption 3' [Mátyás and Sevestre, 2008, Chapter 18.3.1]:

$$\mathbb{E}\left(Y_{igt} - \exp\{\delta_t\} \frac{(1/T)\sum_{t'=1}^{T} Y_{igt'}}{(1/T)\sum_{t'=1}^{T} \exp\{\delta_{t'}\}} \middle| (i,g) \in \Lambda_{0t} \right) = 0 \quad \text{for all } t = 1, \ldots, T,$$

which can be verified using the law of iterated expectations conditional on α_i. Then, evaluating this moment equality at $t = 1$ yields (with normalizing $\delta_1 = 0$)

$$\mathbb{E}\left(Y_{ig1} - \frac{(1/T)\sum_{t'=1}^{T} Y_{igt'}}{(1/T)\sum_{t'=1}^{T} \exp\{\delta_{t'}\}} \middle| (i,g) \in \Lambda_{\infty}^{0} \right) = 0,$$

which identifies the term $(1/T)\sum_{t'=1}^{T} \exp\{\delta_{t'}\}$. Based on this, evaluating the moment equality at $t \geq 2$ identifies $\exp\{\delta_t\}$ for each $t \geq 2$.

Lastly, because all units are untreated at $t = 1$ by assumption, it follows that

$$\mathbb{E}(Y_{ig1}) = \mathbb{E}(Y_{ig1}(0, \mathbf{0}_{(i,g)}^{1})) = \mathbb{E}(\exp\{\alpha_{ig}\}),$$

meaning that $\mathbb{E}(\exp\{\alpha_{ig}\})$ is identified. ■