Beyond Parallel Trends: An Identification-Strategy-Robust Approach to Causal Inference with Panel Data

Brantly Callaway*   Derek Dyal†   Pedro H.C. Sant'Anna‡   Emmanuel S. Tsyawo§

November 27, 2025

Abstract

In this paper, we propose a new approach to causal inference with panel data. Instead of using panel data to adjust for differences in the distribution of unobserved heterogeneity between the treated and comparison groups, we instead use panel data to search for "close comparison groups"—groups that are similar to the treated group in terms of pre-treatment outcomes. Then, we compare the outcomes of the treated group to the outcomes of these close comparison groups in post-treatment periods. We show that this approach is often *identification-strategy-robust* in the sense that our approach recovers the ATT under many different non-nested panel data identification strategies, including difference-in-differences, change-in-changes, or lagged outcome unconfoundedness, among several others. We provide related, though non-nested, results under "time homogeneity", where outcomes do not systematically change over time for any comparison group. Our strategy asks more out of the research design—namely that there exists close comparison groups or time homogeneity (neither of which is required for most existing panel data approaches to causal inference)—but, when available, leads to more credible inferences.

1 Introduction

Panel data approaches to causal inference, such as difference-in-differences, are often categorized with instrumental variables (IV) and regression discontinuity (RD) as being the main quasi-experimental approaches to causal inference. Unlike natural experiment-based approaches to causal inference, such as IV and RD, where the full identification strategy essentially holds as a byproduct of the research design, panel data approaches to causal inference typically require auxiliary assumptions that are not implied by the research design. A leading example is the parallel trends assumption in the context of difference-in-differences. This distinction often results in panel data approaches being referred to as *model-based* in contrast to *design-based*, with the implication being that they are less credible, in the sense of relying on auxiliary assumptions

---

*John Munro Godfrey, Sr. Department of Economics, University of Georgia. `brantly.callaway@uga.edu`

†John Munro Godfrey, Sr. Department of Economics, University of Georgia. `ddyal@uga.edu`

‡Department of Economics, Emory University. `pedro.santanna@emory.edu`

§Department of Economics, Finance and Legal Studies, University of Alabama. `estsyawo@gmail.com`

not implied by the research design. For example, it is often *ex ante* unclear whether the parallel trends assumption is more plausible than the assumptions required for change-in-changes or lagged outcome unconfoundedness identification strategies in any given application.

In the current paper, we retain the conventional pre/post panel data causal inference research design that aims to exploit variation in treatment timing across units and access to untreated comparison groups. However, our aim in the current paper is to increase the credibility of causal inferences in pre/post designs by substantially relaxing the auxiliary assumptions required to go from research design to identification strategy. We mainly consider settings where there are multiple possible comparison groups and where the researcher observes microdata on units within each group. A leading example is the case where a researcher is interested in learning about the causal effect of a state-level policy, has access to data from multiple states, and observes underlying units such as counties, individuals, or firms. Instead of imposing particular auxiliary assumptions to move from research design to identification strategy, we instead look for a subset of the available comparison groups that are "very similar" to the treated group in terms of their pre-treatment outcomes—we refer to such groups as *close comparison groups*. We show that, in many cases, close comparison groups that satisfy certain relatively simple criteria in pre-treatment periods provide valid comparison groups under many different auxiliary assumptions coming from different identification strategies. To give a leading example, we show that, if we can find an untreated comparison group with the same distribution of untreated potential outcomes as the treated group in the period immediately before treatment, then comparing post-treatment outcomes for the treated group to the post-treatment outcomes for this group recovers the ATT under a difference-in-differences identification strategy, a change-in-changes identification strategy, or a lagged outcome unconfoundedness strategy. Similarly, we discuss conditions on comparison groups that rationalize using them under other identification strategies, such as unit-specific linear trends, interactive fixed effects models for untreated potential outcomes, and dynamic panel models for untreated potential outcomes.[1]

Our approach is conceptually much different from typical approaches to causal inference in pre/post designs, where a researcher chooses an identification strategy and uses pre-treatment periods to validate (typically via a placebo test) the identification strategy. Instead, we propose to choose a set of comparison groups that are valid for many different auxiliary assumptions required to move from design to identification strategy. A side-effect of choosing comparison groups that are robust to multiple identification strategies is that the number of applications where our approach can be implemented is smaller than would be the case if a researcher knew that the particular auxiliary assumptions rationalizing some identification strategy held. For example, one criteria that we use below is that the mean outcome for a candidate comparison

---

[1]Another important motivation for panel data approaches to causal inference is that they present the opportunity for the researcher to adjust for systematic latent differences between the treated group and comparison group. Many panel data approaches to causal inference can be thought of as trying to use panel data to adjust for these latent differences across groups. The exact way to make these between-group adjustments is the key distinguishing feature between different panel data identification strategies and is the main source of different estimates arising from different identification strategies. Oftentimes, there is no good way for researchers to rationalize one type of adjustment over another. Moreover, adjusting for latent differences across groups is obviously more challenging the "more different" the groups are. Instead of trying to use panel data to adjust for latent differences, our approach amounts to using panel data to find groups that have the same distribution of latent characteristics. If these groups exist, then our strategy fully obviates the need to make adjustments to make the groups more comparable.

group is the same as the mean outcome for the treated group in the pre-treatment period. We show that, if parallel trends holds, then a level-comparison of mean outcomes between this comparison group and the treated group in the post-treatment period is equal to the ATT. However, it is possible that there are no comparison groups that satisfy this criteria—a DiD identification strategy would still work in this case, but our approach would not deliver any estimate of the ATT. We view this as a feature, rather than a bug, of our approach—if such a comparison group exists, results using our approach will be more robust to different identification strategies (for example, our approach would be robust to certain violations of parallel trends), and effectively less dependent on functional form assumptions. This suggests that inferences using our approach, when it is feasible, are likely to be substantially more credible in a pre/post research design than inferences that rely on auxiliary assumptions such as parallel trends.

We provide several extensions of these results. First, we provide related results for settings with time homogeneity —where the outcomes do not systematically change over time for any of the comparison groups. We show that time homogeneity often provides robustness to multiple identification strategies as well, though the exact kind of robustness differs from using close comparison groups. Thus, we also propose a way to combine both close comparison groups and time homogeneity to further increase robustness. Second, we discuss how to extend our results to settings where close comparison groups are formed conditional on covariates. Third, we discuss how to extend our results to settings with more than one treated group and variation in treatment timing across units.

Related Work

Our approach is related to synthetic control-type methods (Abadie, Diamond, and Hainmueller (2010) and Arkhangelsky et al. (2021)), which construct weights on untreated units in order to match the treated group's pre-treatment characteristics. A key difference between our approach and the synthetic control method is that we consider a setting with multiple possible comparison groups where all units within the same comparison group "count equally". This explains why our results apply in settings with few time periods, while synthetic control methods typically require many time periods to estimate unit-specific weights. The other key difference is that we focus on finding a set of close comparison groups, any of which could be used as a valid comparison group. This is a more stringent requirement on the research design than finding a synthetic control—i.e., there may be applications where the set of close comparison groups is empty, but one could construct a synthetic control. However, when close comparison groups exist, our approach leads to more robust inferences as we show that these groups can be used under many different identification strategies. Kellogg, Mogstad, Pouliot, and Torgovitsky (2021) discuss the trade-off between interpolation bias from synthetic control and extrapolation bias from matching. Gunsilius (2023, 2025) use micro-data and construct group-level comparison groups based on applying weights to either the quantiles or cdfs; this is related to our idea of forming close comparison groups, but we employ a tighter definition of comparison groups.

Our paper is more closely related to recent work that has explored using multiple comparison groups in panel data settings. Part of this literature has used multiple comparison groups to increase efficiency (Marcus and Sant'Anna (2021) and Chen, Sant'Anna, and Xie (2025)). An-

other part of this literature has used multiple comparison groups to relax the parallel trends assumption (Callaway and Tsyawo (2023), Sun, Xie, and Zhang (2025), Rincón and Song (2025), and Liu (2025)), often with an interactive fixed effects model for untreated potential outcomes as motivation.

Our paper is related to a growing literature on robust causal inference with panel data (Manski and Pepper (2018), Rambachan and Roth (2023), Arkhangelsky and Imbens (2022), Arkhangelsky, Imbens, Lei, and Luo (2024), Leavitt and Hatfield (2025), and Athey, Imbens, Qu, and Viviano (2025)). Finally, our results on time homogeneity are related to the idea of using "time as an instrument" (Chernozhukov, Fernandez-Val, Hahn, and Newey (2013)) and regression discontinuity in time (Hausman and Rapson (2018) and Cattaneo, Diaz, and Titiunik (2024)).

2 Setup

2.1 Notation and Assumptions

We consider a setting where a researcher observes T periods of panel data. Let D_{it} denote unit i's treatment status in time period t, and let $D_i = (D_{i1}, D_{i2}, \ldots, D_{iT})'$ denote unit i's vector of treatments across all periods. Let G_i denote a unit's group, and let $\mathcal{G}$ denote the support of G_i. We consider the case where all units in the same group experience the same treatment regime. This means that, for $g \in \mathcal{G}$, we can write $\mathbf{d}(g)$ to indicate the treatment regime experienced by group g. Next, let Y_{it} denote unit i's observed outcome in period t. Next, we define potential outcomes. Let $Y_{it}(\mathbf{d})$ denote the outcome unit i would experience in time period t if it experienced treatment regime $\mathbf{d} = (d_1, d_2, \ldots, d_T)'$. We use the notation $\mathbf{0} = (0, 0, \ldots, 0)'$ to denote the treatment regime of being untreated in all periods, and refer to $Y_{it}(\mathbf{0})$ as unit i's untreated potential outcome in time period t. And, slightly abusing notation, we often use the shorthand notation $Y_{it}(g)$ for $Y_{it}(\mathbf{d}(g))$, which is the potential outcome for unit i in time period t if it experienced the treatment regime implied by being in group g. We maintain the following assumptions throughout the paper:

Assumption 1 (Sampling). The observed data consists of $\left\{ Y_{i1}, Y_{i2}, \ldots, Y_{iT}, D_{i1}, D_{i2}, \ldots, D_{iT}, G_i \right\}_{i=1}^n$, where n denotes the sample size, which are independent and identically distributed. The outcomes are continuously distributed.

Assumption 2 (Treatment and Groups). All units in the same group experience the same treatment regime. That is, G_i fully determines $(D_{i1}, D_{i2}, \ldots, D_{iT})$. No units are treated in the first period, i.e., $D_{i1} = 0$ for all units.

Assumption 3 (No Anticipation). For all $t \in \{1, \ldots, T\}$ and for any unit i, $Y_{it}(\mathbf{d}) = Y_{it}(\mathbf{d}^t)$ where $\mathbf{d}^t = (d_1, \ldots, d_t)$

Assumption 1 says that we observe T periods of panel data. Assumption 2 connects groups and treatments. Assumption 3 says that potential outcomes in a particular period t can depend on the treatment status in the current period or previous periods, but they do not depend on future treatments.

Next, we provide an additional assumption that focuses the discussion in the main text, though we discuss how to relax it later in the paper.

Assumption 4 (Single Treated Group). *Only one group participates in the treatment. Without loss of generality, we use $G = 1$ to denote this group.*

Assumption 4 says that there is a single group that receives the treatment. Let t_g^* denote the time period that group g first participates in the treatment. Let $\bar{\mathcal{G}}$ denote the set of groups that do not participate in the treatment in any period. Under Assumption 4, for simplicity, we let $Y_{it}(1)$ denote the potential corresponding to the treatment regime of being untreated up through period $t_1^* - 1$, becoming treated in period t_1^*, and then following the same treatment regime experienced by units in group 1. We do not use the assumption of staggered treatment adoption, where treatment timing can vary across units, but once a unit becomes treated, it remains treated in subsequent periods. Staggered treatment adoption, however, is an important special case and has been widely considered in the literature on difference-in-differences (Goodman-Bacon (2021) and Callaway and Sant'Anna (2021)). Oftentimes in that literature, the groups are based on treatment timing. It is worth pointing out that our notion of groups can be more general than forming groups solely on the basis of treatment timing, as our setting allows for groups that could experience the same treatment regime (e.g., if we define groups on the basis of state, but several states become treated at the same point in time).

2.1.1 Sources of Multiple Comparison Groups

The approach that we consider relies on there being multiple possible comparison groups and that we observe underlying units with each group. Next, we provide some leading examples of where these groups could come from.

Example 2.1 (State-level policies). Consider the case where we are interested in the effect of a policy implemented by a particular state. In this case, $\mathcal{G}$ could be the set of all states, and $\bar{\mathcal{G}}$ could include all other states that did not implement the policy. We also suppose that we observe underlying units in each state, e.g., counties, individuals, or firms.

Example 2.2 (Staggered treatment adoption). Suppose that different units become treated at different points in time. Let G_i denote the time period when unit i becomes treated. In this case, $\mathcal{G}$ could include all timing-groups and $\bar{\mathcal{G}}$ could include all timing-groups that are not-yet-treated by period t_1^*.

2.1.2 Target Parameters

Following the majority of the literature on panel data causal inference, we target the average treatment effect on the treated (ATT). The DiD literature has often focused on "group-time average treatment effects"—ATT's that can vary by group and time period. In our setting, when there is a single treated group, we can simply index ATT's by time:

$$\text{ATT}_t = \mathbb{E}[Y_t(1) - Y_t(0) \mid G = 1]$$

which is the average treatment effect in period t for group 1. For the remainder of this section, we mainly focus on $\text{ATT}_{t_1^*}$—the average treatment effect for the treated group in the period it becomes treated—though it is straightforward to extend the arguments presented in this section

5

to recover ATT_t in other post-treatment periods. We discuss settings with more groups and more post-treatment periods in Section 7.

2.2 Overview of Identification Strategy

The central idea of our identification strategy is to choose a subset of the possible comparison groups based on being similar to the treated group in pre-treatment periods. This is much different from the way that panel data is typically used in modern approaches to causal inference. Instead of using panel data to somehow account for differences in unobserved heterogeneity between different groups, we use panel data to find very similar comparison groups. Then, under certain conditions, groups being "similar" in terms of pre-treatment outcomes can imply that they are also similar in terms of unobserved heterogeneity, further implying that they should be alike in post-treatment periods even if the mapping from unobservables to untreated potential outcomes changes over time. More generally, the upshot of having comparable groups is that our approach is robust to many different panel data causal inference identification strategies.

For our main results in this section, we consider the subset of comparison groups defined as follows:

$$\mathcal{G}^* = \left\{ g \in \bar{\mathcal{G}} : (Y_{t_1^*-1} \mid G = 1) \overset{d}{=} (Y_{t_1^*-1} \mid G = g) \right\} \tag{1}$$

which is the set of groups that have the same distribution of outcomes in the pre-treatment period $t_1^* - 1$ as the treated group has. We refer to groups that satisfy the condition in Equation (1) as *close comparison groups*. For any group $g \in \bar{\mathcal{G}}$, define

$$\tau_g = \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}[Y_{t_1^*} \mid G = g]$$

i.e., τ_g is the difference in mean outcomes between group 1 and group g. Our main result is that, for any group $g \in \mathcal{G}^*$,

$$\tau_g = \text{ATT}_{t_1^*}$$

if any of several leading panel data identification strategies hold. In other words, if we compare the average outcome in levels for the treated group to the average outcome in levels for group $g \in \mathcal{G}^*$ (i.e., to a close comparison group), then that comparison delivers the average treatment effect for the treated group. Because there may be multiple close comparison groups, the ultimate estimand that we use is

$$\tau = \mathbb{E}[Y_{t_1^*} \mid G = 1] - \sum_{g \in \mathcal{G}^*} w(g) \cdot \mathbb{E}[Y_{t_1^*} \mid G = g] \tag{2}$$

where $w(g)$ are weights, which can depend on things like the relative frequencies of each close comparison group.[2]

---

[2] One example of $w(g)$ is $w(g) = \text{P}(G = g \mid G \in \mathcal{G}^*)$—in this case, the right hand side of Equation (2) simplifies to $\mathbb{E}[Y_{t_1^*} \mid G \in \mathcal{G}^*]$, which is simply the average untreated potential outcome across all units in that are in any comparison group in $\mathcal{G}^*$. Most of our identification results below go through for any weighting function $w(g)$, so choosing the weights is often more

2.3 Five Leading Panel Data Causal Inference Identification Strategies

In this subsection, we briefly review five leading identification strategies for panel data causal inference. Later, we will connect these identification strategies to our approach of using close comparison groups.

Identification Strategy 1: Difference-in-Differences

The key assumption underlying a difference-in-differences identification strategy is the following parallel trends assumption:

Assumption DiD-1. For all $g \in \mathcal{G}$,

$$\mathbb{E}[\Delta Y_{t_1^*}(0) \mid G = g] = \mathbb{E}[\Delta Y_{t_1^*}(0)]$$

Assumption DiD-1 says that the trend in untreated potential outcomes over time is the same for all groups. It follows immediately that ATT is identified under the parallel trends assumption. In particular, for any $g \in \bar{\mathcal{G}}$,

$$\text{ATT}_{t_1^*} = \mathbb{E}[\Delta Y_{t_1^*} \mid G = 1] - \mathbb{E}[\Delta Y_{t_1^*} \mid G = g]$$

Identification Strategy 2: Change-in-Changes

The main assumption for a change-in-changes identification strategy (Athey and Imbens (2006)) is:

Assumption CiC-1. For any $y \in \text{support}(Y_{t_1^*-1}|G = 1)$

$$Q_{Y_{t_1^*}(0)|G=1}\big(\text{F}_{Y_{t_1^*-1}(0)|G=1}(y)\big) = Q_{Y_{t_1^*}(0)|G=g}\big(\text{F}_{Y_{t_1^*-1}(0)|G=g}(y)\big) \text{ for all } g \in \mathcal{G};$$

See, in particular, Theorem 3.1 of Athey and Imbens (2006).[3] Assumption CiC-1 can be thought of as a nonlinear way to account for trends in untreated potential outcomes over time—note that, like DiD, the left hand side is related to how untreated potential outcomes change over time for the treated group but is not identified by the sampling process, while the right hand side involves the observed change in untreated potential outcomes across time periods for an untreated comparison group. CiC identification strategies also require the following support condition:

Assumption CiC-2.

$$\text{support}(Y_{t_1^*-1} \mid G = 1) \subseteq \text{support}(Y_{t_1^*-1} \mid G = g)$$

---

of an estimation problem, with good choices of weights leading to more precise estimates of $ATT_{t_1^*}$.

[3]Athey and Imbens (2006) derive this condition from an underlying model where untreated potential outcomes are generated by a non-separable but invertible function of scalar unobservables whose distribution can be different across groups but does not change across time, and under a support assumption.

Under these assumptions, one can show that the ATT is identified and given by

$$\text{ATT}_{t_1^*} = \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}\left[Q_{Y_{t_1^*}|G=g}\left(\text{F}_{Y_{t_1^*-1}|G=g}(Y_{t_1^*-1})\right) \,\Big|\, G = 1\right]$$

Identification Strategy 3: Lagged Outcome Unconfoundedness

The main idea of a lagged outcome unconfoundedness identification strategy (see, e.g., Ding and Li (2019), Callaway and Li (2023), and Powell, Griffin, and Wolfson (2023)) is to compare post-treatment outcomes of treated and untreated units that have the same pre-treatment outcome. The main identifying assumption is:

Assumption LOU-1. $Y_{t_1^*}(0) \perp\!\!\!\perp G \mid Y_{t_1^*-1}(0)$

This is an unconfoundedness assumption, but where the conditioning variables are pre-treatment outcomes. For recovering $\text{ATT}_{t_1^*}$, an important implication of this assumption is that

$$\mathbb{E}[Y_{t_1^*}(0) \mid Y_{t_1^*-1}(0), G = 1] = \mathbb{E}[Y_{t_1^*}(0) \mid Y_{t_1^*-1}(0), G = g] \text{ for all } g \in \bar{\mathcal{G}} \tag{3}$$

The other important condition for lagged outcome unconfoundedness is the following overlap condition

Assumption LOU-2. For all $g \in \bar{\mathcal{G}}$ and $y \in \text{support}(Y_{t_1^*-1})$, $\text{P}(G = 1|Y_{t_1^*-1}(0) = y, G \in \{1, g\}) < 1$

This assumption says that, for all treated units, there exist units in group $g \in \bar{\mathcal{G}}$ that have the same pre-treatment outcomes. Under these assumptions, $\text{ATT}_{t_1^*}$ is identified and given by

$$\text{ATT}_{t_1^*} = \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}\left[\mathbb{E}[Y_{t_1^*} \mid Y_{t_1^*-1}, G = g] \,\big|\, G = 1\right]$$

Identification Strategy 4: Interactive Fixed Effects

Next, we consider an interactive fixed effects model for untreated potential outcomes based on the following assumption:

Assumption IFE-1. For $t = 1, \ldots, T$,

$$Y_{it}(0) = \theta_t + \eta_i + \lambda_i' F_t + e_{it} \tag{4}$$

where λ_i is an R-dimensional vector of time-invariant heterogeneity and F_t is an R-dimensional vector of fixed coefficients on the λ_i—F_t is often referred to as a factor and λ_i is often referred to as a factor loading. e_{it} satisfies the condition that $\mathbb{E}[e_t \mid G = g] = 0$ for all $g \in \mathcal{G}$.

This condition says that the groups can be systematically different from each other in terms of η_i and λ_i, but not in terms of the time-varying error term e_{it}. Interactive fixed effects models for untreated potential outcomes have been popular in papers that try to relax the parallel trends assumption (e.g., Abadie, Diamond, and Hainmueller (2010), Gobillon and Magnac (2016), Xu

(2017), and Callaway and Tsyawo (2023), among others). In general, in settings with only two time periods, $\text{ATT}_{t_1^*}$ is not identified if untreated potential outcomes are generated by the interactive fixed effects model in Equation (4).

For the discussion that follows, it is often convenient to stack untreated potential outcomes across time periods and take expectations conditional on group, i.e., let $\mathbf{Y}(0) = (Y_1(0), Y_2(0), \ldots, Y_T(0))'$, and then consider

$$\mathbb{E}[\mathbf{Y}(0) \mid G = g] = \boldsymbol{\theta} + \mathbf{F}\boldsymbol{\Lambda}_g \qquad (5)$$

where

$$\mathbb{E}[\mathbf{Y}(0) \mid G = g] := \begin{pmatrix} \mathbb{E}[Y_1(0) \mid G = g] \\ \vdots \\ \mathbb{E}[Y_T(0) \mid G = g] \end{pmatrix}, \boldsymbol{\theta} := \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_T \end{pmatrix}, \mathbf{F} := \begin{pmatrix} 1 & F_1' \\ \vdots & \vdots \\ 1 & F_T' \end{pmatrix}, \boldsymbol{\Lambda}_g := \begin{pmatrix} \mathbb{E}[\eta \mid G = g] \\ \mathbb{E}[\lambda \mid G = g] \end{pmatrix}$$

where $\boldsymbol{\Lambda}_g$ is an $R + 1$ dimensional vector. $\mathbf{F}$ is a $T \times (R + 1)$ matrix. We can also stack these across groups to get

$$\left[\mathbb{E}[\mathbf{Y}(0) \mid G = g]'\right]_{g \in \mathcal{G}} = \mathbf{1}_{|\mathcal{G}|}\boldsymbol{\theta}' + \boldsymbol{\Lambda}\mathbf{F}' \text{ where } \boldsymbol{\Lambda} := \left[\boldsymbol{\Lambda}_g'\right]_{g \in \mathcal{G}}$$

Next, define $\mathbf{Y}^{\text{pre}(t)}(0) = (Y_1(0), \ldots, Y_{t-1}(0))'$ to be the vector of untreated potential outcomes in periods before t, and consider

$$\mathbb{E}[\mathbf{Y}^{\text{pre}(t)}(0) \mid G = g] = \boldsymbol{\theta}^{\text{pre}(t)} + \mathbf{F}^{\text{pre}(t)}\boldsymbol{\Lambda}_g \qquad (6)$$

where

$$\boldsymbol{\theta}^{\text{pre}(t_1^*)} := \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_{t_1^*-1} \end{pmatrix}, \quad \mathbf{F}^{\text{pre}(t)} := \begin{pmatrix} 1 & F_1' \\ \vdots & \vdots \\ 1 & F_{t_1^*-1}' \end{pmatrix}$$

For our results on close comparison groups, we consider the following assumption.

Assumption IFE-2*. $\text{rank}\big(\mathbf{F}^{\text{pre}(t_1^*)}\big) = \text{rank}\big(\mathbf{F}\big)$.

We add the $*$ superscript to indicate that we do not maintain this assumption every time that we discuss the IFE identification strategy, but rather only invoke it in certain contexts. Assumption IFE-2* effectively requires that there be the same amount of variation in the factors in pre-treatment periods relative to the variation in the factors in post-treatment periods. A necessary condition for Assumption IFE-2* is that the number of pre-treatment periods be at least as large as the number of unobserved heterogeneity terms, i.e., $t_1^* - 1 \geq R + 1$. Most importantly, it rules out some factor "turning on" in post-treatment periods that did not affect outcomes in pre-treatment periods.

Additionally, notice that we can view Equations (5) and (6) as maps $\begin{pmatrix} \mathbb{E}[\eta \mid G = g] \\ \mathbb{E}[\lambda \mid G = g] \end{pmatrix} \mapsto$

$\mathbb{E}[\mathbf{Y}(0) \mid G = g]$ and $\begin{pmatrix} \mathbb{E}[\eta \mid G = g] \\ \mathbb{E}[\lambda \mid G = g] \end{pmatrix} \mapsto \mathbb{E}[\mathbf{Y}^{\mathrm{pre}(t)}(0) \mid G = g]$, respectively. The rank condition in Assumption IFE-2* implies that both maps are injective, and, hence, that groups that have the same mean of pre-treatment outcomes also have the same mean of η and λ.

Let $\bar{\mathbf{\Lambda}} = \left[\mathbf{\Lambda}'_g\right]_{g \in \bar{\mathcal{G}}}$. For our results on time homogeneity, we consider the following assumption.

Assumption IFE-3*. $\mathrm{rank}(\bar{\mathbf{\Lambda}}) = R + 1$.

Like Assumption IFE-2*, we do not maintain Assumption IFE-3* throughout the paper, but only invoke it in certain contexts. Assumption IFE-3* requires that there be enough independent variation in unobserved heterogeneity among the comparison groups relative to the number of interactive fixed effects.

Identification Strategy 5: Latent Unconfoundedness

An even more general identification strategy is to take the assumption of latent unconfoundedness as a starting point:

$$Y_t(0) \perp\!\!\!\perp G \mid \xi \tag{7}$$

where ξ is an L-dimensional vector of unobserved heterogeneity. Latent unconfoundedness can equivalently be stated as in the following assumption

Assumption Lat. Unc.-1. The following model for untreated potential outcomes holds

$$Y_{it}(0) = h_t(\xi_i) + e_{it}$$

where h_t is left unrestricted, the distribution of ξ can vary across groups, but $e_t \perp\!\!\!\perp G \mid \xi$.

The motivation for Assumption Lat. Unc.-1 is similar to that of the interactive fixed effects model discussed above, except that it does not impose the functional form assumption of the IFE model. Like for the case of IFE, in general, $\mathrm{ATT}_{t_1^*}$ is not identified under latent unconfoundedness with only two time periods. Let

$$\mathbf{H}(\xi) := \Big(h_1(\xi), \ldots, h_T(\xi)\Big), \qquad \mathbf{H}^{\mathrm{pre}(t)}(\xi) := \Big(h_1(\xi), \ldots, h_{t-1}(\xi)\Big)$$

which are T-dimensional and $t-1$-dimensional, respectively. Without loss of generality, we define ξ so that $\mathbf{H}(\xi)$ is injective—this effectively controls the dimension of ξ (where the maximum possible dimension of ξ is T, but it can be smaller), similar to our discussion above about $\mathrm{rank}(\mathbf{F})$ determining the number of factors.

Assumption Lat. Unc.-2. $\mathbf{H}^{\mathrm{pre}(t_1^*)}$ is injective.

Assumption Lat. Unc.-2 play a similar role as Assumption IFE-2* discussed above for IFE. It mainly rules out components of ξ "turning on" in post-treatment periods. The main implication of Assumption Lat. Unc.-2 that we use below is that it implies that, for $t \geq t_1^*$, $h_t(\xi) = \phi_t\big(\mathbf{H}^{\mathrm{pre}(t_1^*)}(\xi)\big)$ for some function ϕ_t. Note that $\mathbf{H}^{\mathrm{pre}(t_1^*)}$ cannot be injective if $\dim(\xi) > t_1^* - 1$.

3 Robust Identification with Close Comparison Groups

This section contains our main results on the robustness of using close comparison groups. We start with the case where close comparison groups are based on having the same distribution of outcomes in the period immediately preceding the treatment. Then, we discuss extensions where close comparison groups are (i) based on having only the same mean outcomes in a single pre-treatment period, (ii) based on having the same mean outcomes over several pre-treatment periods, and (iii) based on having the same distribution of outcomes over several pre-treatment periods.

Theorem 1. Under Assumptions 1 to 4, if $\mathcal{G}^*$ is non-empty, and if any of the following identification strategies hold: (i) difference-in-differences, (ii) lagged outcome unconfoundedness, or (iii) change-in-changes, then

$$\tau = \text{ATT}_{t_1^*}$$

In contrast, under (iv) IFE, or (v) latent unconfoundedness, it is possible that $\tau \neq \text{ATT}_{t_1^*}$.

The proof of Theorem 1 is provided in Appendix A. The first part of Theorem 1 establishes the central result of our paper. It shows that, if we are able to find any comparison groups that satisfy the condition of having the same distribution of observed outcomes in the pre-treatment period, then we can simply compare the average outcome for the treated group to the average outcome for these groups to recover the ATT in a way that is robust across several different leading identification strategies. In other words, in applications where there are "close comparison groups", then it is possible to recover ATT without having to take a stand on the "correct" identification strategy, at least among DiD, CiC, and LOU—three leading identification strategies that are used in this context. This sort of strategy may not be feasible in all applications—because there may not be any available close comparison groups—but, in applications where close comparison groups are available, this result shows that recovering the ATT is substantially more robust than in applications that rely more strongly on auxiliary assumptions like parallel trends for identification.

The identification strategy discussed here is also transparent and simple to explain to non-experts. It suggests using pre-treatment data to find comparison groups that are very similar to the treated group, and then to compare the level of the outcomes for the treated group to the level of the outcome for very similar comparison groups. For example, in an application where a state implements a policy and where we have microdata from every state, this approach suggests using the pre-treatment data to find other states that have similar outcome distributions and to discard states that have different distributions of outcomes in the pre-treatment period. The simple intuition of our approach contrasts with most panel data approaches to causal inference. For example, at least arguably, comparing trends in outcomes over time for the treated group relative to untreated comparison states (as would be done with DiD) is unnatural outside of the context of linear fixed effects models for untreated potential outcomes and can be hard to explain or rationalize to non-economists. These strategies are also conceptually different in that they effectively use the pre-treatment data to "adjust for" differences in the levels of outcomes across

groups, even if the distributions of these outcomes are quite different across groups. Adjusting for differences between groups is a much harder task—and, as emphasized above, often requires making auxiliary assumptions that can be hard to rationalize—than comparing similar groups. Our approach is simply to check if we have access to similar groups and, if we do, to compare their outcomes in post-treatment periods.

The second part of Theorem 1 shows that forming $\mathcal{G}^*$ and comparing post-treatment outcomes for the treatment group relative to outcomes for groups in $\mathcal{G}^*$ is not a panacea—under IFE or latent unconfoundedness, this comparison is not guaranteed to deliver the ATT. The intuition for both of these results is that groups having the same distribution of outcomes in the pre-treatment period is not strong enough to guarantee that these groups have the same distribution of time-invariant unobservables, particularly in settings where the dimension of the time-invariant unobservables is greater than one. To foreshadow some of our results in the next section, it is interesting to notice that, as discussed above, in general, the ATT is not identified for either of these identification strategies when there are only two time periods. We will show below that, if there are more pre-treatment periods, at least under certain conditions, one can use an alternative definition of close comparison groups (involving groups have similar outcomes across multiple pre-treatment periods) that can rationalize the ATT as being equal to the difference between post-treatment outcomes for the treated group relative to post-treatment outcomes for groups in this alternative version of close comparison groups.

Remark 3.1 (Analogy to de-biasing using pre-treatment period). Difference-in-differences, change-in-changes, and lagged outcome unconfoundedness identification strategies all have representations like:

$$\text{ATT}_{t_1^*} = \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}[Y_{t_1^*} \mid G = g] - \text{de-biasing term from period } t_1^* - 1 \text{ differences}$$

where, for example, the de-biasing term for difference-in-differences is $\mathbb{E}[Y_{t_1^*-1} \mid G = 1] - \mathbb{E}[Y_{t_1^*-1} \mid G = g]$. The particular de-biasing terms for change-in-changes and lagged outcome unconfoundedness are different. Instead of de-biasing, the approach that we proposed above amounts to looking for groups where we do not need to do any de-biasing using the pre-treatment period, and our criteria for being a close comparison group implied that the de-biasing term for all three of these identification strategies is zero.

Remark 3.2 (Post-treatment placebo tests). When $|\mathcal{G}^*| > 1$, i.e., when there is more than one close comparison group, there are post-treatment testable implications of our approach. In particular, if any of the difference-in-differences, change-in-changes, or lagged outcome unconfoundedness identification strategies hold, it will be the case that

$$\mathbb{E}[Y_{t_1^*} \mid G = g] = \mathbb{E}[Y_{t_1^*} \mid G = g'] \text{ for all } g, g' \in \mathcal{G}^*$$

In other words, all close comparison groups should have the same average outcome in post-treatment periods.

Remark 3.3 (Single available comparison group). The discussion in this section has focused on the case where $|\bar{\mathcal{G}}| > 1$, i.e., where there are multiple possible comparison groups, but the

arguments can continue to go through when $|\bar{\mathcal{G}}| = 1$; in this case, our approach amounts to simply checking whether or not the one available comparison group is a close comparison group or not. The main value of having more comparison groups is that the researcher has more opportunities to find close comparison groups among them.

3.1 Alternative approaches to forming the set of close comparison groups

Above, we formed the set of close comparison groups on the basis of having the same distribution of outcomes as the treated group in the period immediately preceding the treatment. It is possible to form a set of close comparison groups using different criteria, and these different criteria lead to different robustness properties. We consider several different alternative definitions of close comparison groups below.

3.1.1 Single Pre-Treatment Period

One-period means: $\mathcal{G}^*_{\text{mean}}(1) := \left\{ g \in \bar{\mathcal{G}} : \mathbb{E}[Y_{t=1} \mid G = 1] = \mathbb{E}[Y_{t=1} \mid G = g] \right\}$

$\mathcal{G}^*_{\text{mean}}(1)$ collects all of the possible comparison groups that have the same mean pre-treatment outcome as the treated group. This is a weaker requirement than having the same distribution of outcomes in the pre-treatment period. Thus, $\mathcal{G}^* \subseteq \mathcal{G}^*_{\text{mean}}(1)$.

3.1.2 Multiple Pre-Treatment Periods

For some $S \in \{1, \ldots, t_1^* - 1\}$, we consider the following alternative definitions of close comparison groups.

Multiple-period means: $\mathcal{G}^*_{\text{mean}}(S) := \Big\{ g \in \bar{\mathcal{G}} : \mathbb{E}[Y_{t=t_1^*-j} \mid G = 1] = \mathbb{E}[Y_{t_1^*-j} \mid G = g] \text{ for } j = 1, \ldots, S \Big\}$

Multi-period distributions: $\mathcal{G}^*_{\text{dist}}(S) := \Big\{ g \in \bar{\mathcal{G}} : \big(Y_{t_1^*-1}, \ldots, Y_{t_1^*-S} \mid G = 1\big) \overset{d}{=} \big(Y_{t_1^*-1}, \ldots, Y_{t_1^*-S} \mid G = g\big) \Big\}$

$\mathcal{G}^*_{\text{mean}}(S)$ generalizes $\mathcal{G}^*_{\text{mean}}(1)$ so that a close comparison group must have the same mean of outcomes across S pre-treatment periods, which is a more stringent criteria when $S > 1$. Thus, $\mathcal{G}^*_{\text{mean}}(S) \subseteq \mathcal{G}^*_{\text{mean}}(1)$. The researcher can choose different values of S, up to $S = t_1^* - 1$ (which would amount to using all available pre-treatment periods), and larger values of S can only shrink the number of close comparison groups.

$\mathcal{G}^*_{\text{dist}}(S)$ is the most restrictive definition of a close comparison group that we consider in the paper. It generalizes $\mathcal{G}^*$—instead of requiring that a group has the same distribution of outcomes in the period immediately preceding treatment, it requires that a group has the same joint distribution of pre-treatment outcomes across S periods preceding treatment as the treated group has. This leads to the smallest set of close comparison groups among those that we have considered. Notice that $\mathcal{G}^*_{\text{dist}}(S) \subseteq \mathcal{G}^*$ and $\mathcal{G}^*_{\text{dist}}(S) \subseteq \mathcal{G}^*_{\text{mean}}(S)$. In addition, similar to the discussion above, larger values of S can only shrink the number of close comparison groups.

3.1.3 Identification under alternative definitions of close comparison groups

This section contains our second main result relating the five identification strategies discussed previously to the alternative definitions of a close comparison group discussed above. Before stating this result, we define the following estimands

$$\tau_{\text{mean}}^g(1) := \mathbb{E}[Y_{t_1^*}|G=1] - \mathbb{E}[Y_{t_1^*}|G=g] \qquad \text{for } g \in \mathcal{G}_{\text{mean}}^*(1)$$

$$\tau_{\text{mean}}^g(S) := \mathbb{E}[Y_{t_1^*}|G=1] - \mathbb{E}[Y_{t_1^*}|G=g] \qquad \text{for } g \in \mathcal{G}_{\text{mean}}^*(S)$$

$$\tau_{\text{dist}}^g(S) := \mathbb{E}[Y_{t_1^*}|G=1] - \mathbb{E}[Y_{t_1^*}|G=g] \qquad \text{for } g \in \mathcal{G}_{\text{dist}}^*(S)$$

and

$$\tau_{\text{mean}}(1) := \sum_{g \in \mathcal{G}_{\text{mean}}^*(1)} \tau_{\text{mean}}^g(1) \cdot w_{\text{mean}(1)}^g(g)$$

$$\tau_{\text{mean}}(S) := \sum_{g \in \mathcal{G}_{\text{mean}}^*(S)} \tau_{\text{mean}}^g(S) \cdot w_{\text{mean}(S)}^g(g)$$

$$\tau_{\text{dist}}(S) := \sum_{g \in \mathcal{G}_{\text{dist}}^*(S)} \tau_{\text{dist}}^g(S) \cdot w_{\text{dist}(S)}^g(g)$$

The next proposition collects a number of results on whether close comparison groups, defined in the ways discussed above, combined with particular identification strategies guarantee that the post-treatment comparison of means, $\mathbb{E}[Y_{t_1^*} \mid G=1] - \mathbb{E}[Y_{t_1^*} \mid G=g]$, where g is a close comparison group, recovers $\text{ATT}_{t_1^*}$.

Proposition 1. For IFE, suppose that Assumption IFE-2* holds, and, for latent unconfoundedness, suppose that $S \geq L$. Under Assumptions 1 to 4, the following table indicates when a group g satisfying a certain definition of being a close comparison group implies that $\mathbb{E}[Y_{t_1^*} \mid G=1] - \mathbb{E}[Y_{t_1^*} \mid G=g] = \text{ATT}_{t_1^*}$ under different identification strategies.

	$g \in \mathcal{G}_{\text{mean}}^*(1)$ $\implies \tau_{\text{mean}}^g(1) = \text{ATT}_{t_1^*}$	$g \in \mathcal{G}_{\text{mean}}^*(S)$ $\implies \tau_{\text{mean}}^g(S) = \text{ATT}_{t_1^*}$	$g \in \mathcal{G}_{\text{dist}}^*(S)$ $\implies \tau_{\text{dist}}^g(S) = \text{ATT}_{t_1^*}$
DiD	✓	✓	✓
CiC	✗	✗	✓
LOU	✗	✗	✓
IFE	✗	✓	✓
Lat. Unc.	✗	✗	✓

The proof of Proposition 1 is provided in Appendix A. There are several main takeaways from this proposition. First, among the identification strategies that we consider, a group g having the same mean outcome in a single pre-treatment period only guarantees that $\tau_{\text{mean}}(1) = \text{ATT}_{t_1^*}$ for difference-in-differences. The second column indicates that having the same mean outcomes across multiple pre-treatment periods is additionally sufficient to guarantee that $\tau_{\text{mean}}(S) = \text{ATT}_{t_1^*}$ for interactive fixed effects, as long as a rank condition related to the number of available pre-treatment time periods holds. On the other hand, it is not sufficient for change-in-changes,

lagged outcomes, or latent unconfoundedness. The intuition for this result is that DiD and IFE models are linear in unobserved heterogeneity, so groups with the same pre-treatment means of outcomes also have the same means of unobserved heterogeneity, as long as there are enough pre-treatment periods. In contrast, change-in-changes, lagged outcome unconfoundedness, and latent unconfoundedness are nonlinear in unobserved heterogeneity, so having the same pre-treatments means of outcomes is not sufficient to guarantee that the distribution of unobserved heterogeneity (or lagged outcomes) is the same between groups. Finally, the last column shows that having the same joint distribution of outcomes across multiple pre-treatment periods is sufficient to guarantee that $\tau_{\text{dist}}(S) = \text{ATT}_{t_1^*}$ for all five identification strategies considered here, as long as there are enough pre-treatment periods relative to the dimension of unobserved heterogeneity. Essentially, exhibiting the same joint distribution of outcomes over several pre-treatment periods guarantees that groups have the same distribution of unobserved heterogeneity, even if the model is nonlinear in unobserved heterogeneity. And having the same distribution of unobserved heterogeneity implies that post-treatment comparisons of means recover the ATT, even for nonlinear models that can vary over time in complicated ways.

3.2 Other identification strategies:

Our discussion above has focused on five important identification strategies for causal inference with panel data; however, it is important to note that our arguments apply to other identification strategies as well. We briefly mention two additional identification strategies here and conjecture that our arguments could be extended to others as well.

Unit-specific linear trends:

$$Y_{it}(0) = \theta_t + \eta_i + \lambda_i t + e_{it} \quad \text{with } \mathbb{E}[e_t \mid G = g] = 0 \text{ for all } g \in \mathcal{G}. \tag{8}$$

Dynamic panel data model:

$$Y_{it}(0) = \theta_t + \eta_i + \rho Y_{it-1}(0) + e_{it} \quad \text{with } \mathbb{E}[e_t \mid G = g] = 0 \text{ for all } g \in \mathcal{G}. \tag{9}$$

Unit-specific linear trends models are popular in empirical work and have been considered in Heckman and Hotz (1989) and Wooldridge (2005). Dynamic panel data models for untreated potential outcomes are somewhat less common in empirical work focused on causal inference (see Marx, Tamer, and Tang (2025) for a recent paper that considers this type of setting). In the next proposition, we relate these models to our approach of finding close comparison groups.

Proposition 2. Suppose that $S \geq 2$. Under Assumptions 1 to 4, the following table indicates when a group g satisfying a certain definition of being a close comparison group implies that $\mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}[Y_{t_1^*} \mid G = g] = \text{ATT}_{t_1^*}$ under different identification strategies.

	$g \in \mathcal{G}^*_{\text{mean}}(1)$	$g \in \mathcal{G}^*_{\text{dist}}(1)$	$g \in \mathcal{G}^*_{\text{mean}}(S)$	$g \in \mathcal{G}^*_{\text{dist}}(S)$
	$\implies \tau^g_{\text{mean}}(1) = \text{ATT}_{t^*_1}$	$\implies \tau^g_{\text{dist}}(1) = \text{ATT}_{t^*_1}$	$\implies \tau^g_{\text{mean}}(S) = \text{ATT}_{t^*_1}$	$\implies \tau^g_{\text{dist}}(S) = \text{ATT}_{t^*_1}$
Linear Trends	✗	✗	✓	✓
Dynamic Panel	✗	✗	✓	✓

The proof is provided in Appendix A. Proposition 2 shows that both linear trends and dynamic panel models for untreated potential outcomes behave very similarly to interactive fixed effects models with $R = 1$ (i.e., with one interactive fixed effect). Thus, defining the set of close comparison groups based on two periods of balance with respect to means is sufficient for $\tau^g_{\text{mean}}(2) = \text{ATT}_{t^*_1}$.

4 Robust Identification with Time Homogeneity

Remark 3.1 emphasized that all of the two-period identification strategies that we have considered involve a cross-group comparison de-biased by differences between the groups in pre-treatment periods (where the particular form of de-biasing depended on the identification strategy). Then, we tried to find groups where the de-biasing was unnecessary because the groups were already sufficiently similar.

A different, but related, viewpoint for all of these identification strategies is that they can be viewed as a de-biased before-after comparison where the de-biasing is to try to account for time effects on untreated potential outcomes. For example, for difference-in-differences:

$$\text{ATT}_{t^*_1} = \underbrace{\mathbb{E}[Y_{t^*_1} - Y_{t^*_1-1} \mid G = 1]}_{\text{before-after}} - \underbrace{\mathbb{E}[Y_{t^*_1} - Y_{t^*_1-1} \mid G = g]}_{\text{time de-biasing}}$$

where, as above, the exact form of time de-biasing depends on the identification strategy. In line with our approach above, this suggests that an alternative way to provide an identification-strategy-robust approach is to look for periods where outcomes do not change over time for the comparison groups, which is what we pursue here.

We define "periods where outcomes do not change over time for the comparison groups" in terms of there being no change in the mean outcome for each group across time periods (though the mean outcome need not be the same for each group). Consider

$$t^{\text{TH}}_{\text{mean}} := \max\left\{s \in \{t^*_1, \ldots, T\} : \mathbb{E}[Y_{t=s'}(0) \mid G = g] = \mathbb{E}[Y_{t=t^*_1-1}(0) \mid G = g] \text{ for all } s' \leq s \text{ and } g \in \bar{\mathcal{G}}\right\}$$

which is the maximum post-treatment period for which the mean outcome did not change relative to the period immediately preceding treatment for any comparison group. We follow the convention of setting $t^{\text{TH}}_{\text{mean}}$ to 0 if the criteria is not met for any period.

Next, we define the following sets of time periods:

$$\mathcal{T}_{\text{mean}} = \left\{t \in \{t^*_1, \ldots, T\} : t \leq t^{\text{TH}}_{\text{mean}}\right\}$$

which is the set of post-treatment time periods for which mean time homogeneity holds. Note

that this set can be empty if there are no time periods where there are no observed periods in which time homogeneity holds.

For this section, we consider the following estimand:

$$\tau_t^{\mathrm{TH}} := \mathbb{E}[Y_t - Y_{t_1^*-1} \mid G = 1]$$

for some period $t \in \mathcal{T}_{\mathrm{mean}}$. τ_t^{TH} is the average change in outcomes from period $Y_{t_1^*-1}$, the period right before group 1 becomes treated, to period t for the treated group.

The following proposition contains our main result on the robustness of τ_t^{TH} under mean time homogeneity.

Proposition 3. For IFE, suppose that Assumption IFE-3* holds. Under Assumptions 1 to 4, and if $\mathcal{T}_{\mathrm{mean}} \neq \emptyset$, then, for any $t \in \mathcal{T}_{\mathrm{mean}}$, the following table indicates when $\tau_t^{\mathrm{TH}} = \mathrm{ATT}_t$ under different identification strategies:

	$\tau_t^{\mathrm{TH}} = \mathrm{ATT}_t$
DiD	✓
CiC	✗
LOU	✗
IFE	✓
Lat. Unc.	✗

The proof is provided in Appendix A. The proposition shows that difference-in-differences and interactive fixed effects identification strategies both imply that $\tau_t^{\mathrm{TH}} = \mathrm{ATT}_t$ for any t in the set of time periods where mean time homogeneity holds. That mean time homogeneity is sufficient for $\tau_t^{\mathrm{TH}} = \mathrm{ATT}_t$ under an IFE identification strategy is quite interesting, as typically it would not be possible to identify ATT_t under IFE with only two time periods, where quasi-differencing strategies would typically require at least $(R+1)$ pre-treatment periods. However, under an IFE model for untreated potential outcomes, if we have access to a sufficient number of "distinct" comparison groups, then mean time homogeneity effectively implies that the factors, F_t, are constant over the relevant time periods, which then implies that untreated potential outcomes for the treated group are time homogeneous as well.

In contrast, change-in-changes, lagged outcome unconfoundedness, and latent unconfoundedness do not guarantee that $\tau_t^{\mathrm{TH}} = \mathrm{ATT}_t$ even when there are no time effects for comparison groups. The intuition for this result is that the nonlinearity of each of these identification strategies leads to mean time homogeneity not being sufficient to guarantee that untreated potential outcomes for the treated group are time homogeneous as well. That being said, Proposition 6 in Appendix A counterbalances this result, at least to some extent, by showing that mean time homogeneity can be sufficient for $\tau_t^{\mathrm{TH}} = \mathrm{ATT}_t$ under lagged outcome unconfoundedness under a "flexible" parametric model for $\mathbb{E}[Y_t(0) \mid Y_{t_1^*-1}(0), G = g]$, where the amount of flexibility that can be allowed for is connected to the number of available comparison groups.

5 Multiply Robust Identification

As discussed above, comparisons (i)between close comparison groups and (ii) based on time homogeneity can both be used to identify $\text{ATT}_{t_1^*}$. In both cases, these comparisons are robust to multiple identification strategies. In some applications, it may be the case that the researcher has access to both close comparison groups and a setting where time homogeneity holds. Given that the robustness is non-nested for each approach, it suggests combining these two approaches to provide an even more robust estimator of $\text{ATT}_{t_1^*}$. To do so, we consider the following estimand:

$$\tau_{\text{mr}} := \mathbb{E}[Y_{t_1^*} \mid G = 1] - \sum_{g \in \mathcal{G}^*} w_g \mathbb{E}[Y_{t_1^*} \mid G = g] - \left(\mathbb{E}[Y_{t_1^*-1} \mid G = 1] - \sum_{g \in \mathcal{G}^*} w_g \mathbb{E}[Y_{t_1^*-1} \mid G = g] \right)$$

τ_{mr} is the difference between the post-treatment comparison of means between the treated group and close comparison groups adjusted by the analogous pre-treatment comparison of means. The following discussion illustrates that τ_{mr} identifies $\text{ATT}_{t_1^*}$ if either close comparison groups identification or time homogeneity identification holds.

Case 1: Suppose that close comparison groups identification works, i.e., $\mathbb{E}[Y_{t_1^*}(0) \mid G = 1] = \sum_{g \in \mathcal{G}^*} w_g \mathbb{E}[Y_{t_1^*}(0) \mid G = g]$, but time homogeneity does not hold, i.e., $\mathbb{E}[Y_{t_1^*}(0) \mid G = 1] \neq \mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1]$. Then,

$$\tau_{\text{mr}} = \left(\mathbb{E}[Y_{t_1^*}(1) \mid G = 1] - \sum_{g \in \mathcal{G}^*} w_g \mathbb{E}[Y_{t_1^*}(0) \mid G = g] \right) + \left(\mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1] - \sum_{g \in \mathcal{G}^*} w_g \mathbb{E}[Y_{t_1^*-1}(0) \mid G = g] \right)$$

$$= \text{ATT}_{t_1^*} + 0 = \text{ATT}_{t_1^*}$$

where the first equality holds by writing τ_{mr} in terms of potential outcomes, and the second equality holds because close comparison groups holds—the first term is equal to $\text{ATT}_{t_1^*}$ because close comparison groups holds, and the second term is equal to zero because, under any of our definitions of close comparison groups, the treated group and all comparison groups have the same mean outcome in the pre-treatment period $t_1^* - 1$.

Case 2: Suppose that time homogeneity identification holds, i.e., $\mathbb{E}[Y_{t_1^*}(0) \mid G = 1] = \mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1]$, but close comparison groups does not hold, i.e., $\mathbb{E}[Y_{t_1^*}(0) \mid G = 1] \neq \sum_{g \in \mathcal{G}^*} w_g \mathbb{E}[Y_{t_1^*}(0) \mid G = g]$. Then,

$$\tau_{\text{mr}} = \left(\mathbb{E}[Y_{t_1^*}(1) \mid G = 1] - \sum_{g \in \mathcal{G}^*} w_g \mathbb{E}[Y_{t_1^*}(0) \mid G = g] \right) + \left(\mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1] - \sum_{g \in \mathcal{G}^*} w_g \mathbb{E}[Y_{t_1^*-1}(0) \mid G = g] \right)$$

$$= \left(\mathbb{E}[Y_{t_1^*}(1) \mid G = 1] - \mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1] \right) - \sum_{g \in \mathcal{G}^*} w_g \left(\mathbb{E}[Y_{t_1^*}(0) \mid G = g] - \mathbb{E}[Y_{t_1^*-1}(0) \mid G = g] \right)$$

$$= \text{ATT}_{t_1^*} + 0 = \text{ATT}_{t_1^*}$$

where the first equality holds by writing τ_{mr} in terms of potential outcomes, the second equality holds by rearranging terms, and the last line holds because time homogeneity holds—the first term is equal to $\text{ATT}_{t_1^*}$ because time homogeneity holds, and the second term is equal to zero

because time homogeneity holds among all comparison groups.

6 Estimation and Inference

In this section, we propose an estimator for $\text{ATT}_{t_1^*}$. The key challenge for our estimation procedure is that the set of close comparison groups is not known and must be estimated. Given an estimate of this, we effectively then plug it into the identification formula from Theorem 1 to obtain an estimate of $\text{ATT}_{t_1^*}$. We also propose a method for conducting valid inference, taking into account the uncertainty in choosing the set of close comparison groups.

Slightly abusing notation, in this section, we use the notation $\mathcal{G}^*$ to indicate the set of close comparison groups. This is the same notation we used above for choosing the set of close comparison groups based on having the same distribution of outcomes in the period immediately before the treatment; however, the arguments in this section apply to any of the notions of forming close comparison groups earlier in the paper. We consider the following estimator for $\text{ATT}_{t_1^*}$.

$$\widehat{\text{ATT}}_{t_1^*}(h_n) = \bar{Y}_{1,t_1^*} - \sum_{g \in \mathcal{G}} \widehat{w}_g(h_n) \cdot \bar{Y}_{g,t_1^*}$$

where

$$\bar{Y}_{g,t_1^*} := \frac{1}{n_g} \sum_{i=1}^{n} \mathbf{1}\{G_i = g\} Y_{it_1^*} \quad \text{and} \quad \widehat{w}_g(h_n) := \frac{\text{K}\big(\hat{\text{d}}_g/h_n\big) \cdot \widehat{p}_g}{\displaystyle\sum_{g' \in \mathcal{G}} \text{K}\big(\hat{\text{d}}_{g'}/h_n\big) \cdot \widehat{p}_{g'}}$$

and where K is a compactly supported kernel function, h_n is a bandwidth, and $\hat{\text{d}}_g$ is an estimate of the distance between group 1 and group g. Together, having a compactly supported kernel and the bandwidth result in a form of model selection, where only certain groups receive any positive weight. We use the notation $\widehat{\mathcal{G}}^*$ to denote the implied estimator of the set of close comparison groups. An important special case is when K is a uniform kernel, i.e., $\text{K}(u) = \mathbf{1}\{|u| \leq 1\}$, in which case the estimator reduces to

$$\widehat{\text{ATT}}_{t_1^*} = \frac{1}{n_1} \sum_{i=1}^{n} \mathbf{1}\{G_i = 1\} Y_{it_1^*} - \frac{1}{n_{\widehat{\mathcal{G}}^*}} \sum_{i=1}^{n} \mathbf{1}\{G_i \in \widehat{\mathcal{G}}^*\} Y_{it_1^*}$$

where n_1 is the number of units in the treated group, and $n_{\widehat{\mathcal{G}}^*}$ is the number of units in the estimated close comparison groups. This estimator is simply the difference in average outcomes for the treated group relative to the average outcomes in the estimated close comparison groups in the post-treatment period.

The key estimation task is to measure how "close" each untreated group g is to the treated group in pre-treatment periods. Next, we give two leading examples for distance metrics, though we note that our theory applies to any distance metric that satisfies fairly minimal properties discussed below, and the appropriate choice of distance metric mainly comes down to the definition of close comparison groups that the researcher chooses. For example, if the researcher chooses close comparison groups based on having the same distribution of outcomes as the treated group

19

in the period immediately preceding treatment, then a natural distance metric is

$$\mathsf{d}_g^{\text{dist}} := \frac{\int_0^1 \left(\mathrm{Q}_1(u) - \mathrm{Q}_g(u)\right)^2 du}{\sqrt{(\sigma_1^2 + \sigma_g^2)/2}}$$

where $\mathrm{Q}_g(u) := \mathrm{Q}_{Y_{t_1^*-1}|G=g}(u)$ is the quantile function of the outcome distribution for group g in the period immediately preceding treatment, and $\sigma_g^2 := \text{Var}(Y_{t_1^*-1} \mid G = g)$. $\mathsf{d}_g^{\text{dist}}$ is a standardized 2-Wasserstein distance between the treated and group g pre-treatment outcome distributions. The distance is zero if and only if the pre-treatment outcome distributions match, i.e., if group g belongs to the set of close comparison groups.

Similarly, if close comparison groups are based on having the same mean outcomes as the treated group across several pre-treatment periods, then a natural distance metric is

$$\mathsf{d}_g^{\text{mean}} := \sqrt{(\boldsymbol{\mu}_g - \boldsymbol{\mu}_1)'\left((\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_g)/2\right)^{-1}(\boldsymbol{\mu}_g - \boldsymbol{\mu}_1)}.$$

where $\boldsymbol{\mu}_g := (\mathbb{E}[Y_{t_1^*-1} \mid G = g], \ldots, \mathbb{E}[Y_{t_1^*-S} \mid G = g])'$ is the vector of mean outcomes for group g for the S periods preceding treatment, and $\boldsymbol{\Sigma}_g$ is the covariance matrix of outcomes for group g across the same S periods. $\mathsf{d}_g^{\text{mean}}$ is a multivariate standardized distance between the treated and group g pre-treatment mean outcome vectors—it reduces to the standardized difference in Imbens and Rubin (2015) in the case where $S = 1$. The distance is zero if and only if the pre-treatment mean outcome vectors match, i.e., if group g belongs to the set of close comparison groups.

Whatever the distance metric that is ultimately used, it needs to be estimated. For the examples given above, natural estimators are

$$\widehat{\mathsf{d}}_g^{\text{dist}} := \frac{\dfrac{1}{J}\sum_{j=1}^{J} \left(\widehat{\mathrm{Q}}_1(u_j) - \widehat{\mathrm{Q}}_g(u_j)\right)^2}{\sqrt{(\widehat{\sigma}_1^2 + \widehat{\sigma}_g^2)/2}} \quad \text{and} \quad \widehat{\mathsf{d}}_g^{\text{mean}} := \sqrt{(\widehat{\boldsymbol{\mu}}_g - \widehat{\boldsymbol{\mu}}_1)'\left((\widehat{\boldsymbol{\Sigma}}_1 + \widehat{\boldsymbol{\Sigma}}_g)/2\right)^{-1}(\widehat{\boldsymbol{\mu}}_g - \widehat{\boldsymbol{\mu}}_1)}$$

where $\mathcal{U} = \{u_1, \ldots, u_J\}$ is an equally spaced grid on $(0,1)$, $\widehat{\mathrm{Q}}_g(u)$ is the empirical quantile function of the outcome distribution for group g in the period immediately preceding treatment, $\widehat{\sigma}_g^2$ is the sample variance of outcomes for group g in the same period, $\widehat{\boldsymbol{\mu}}_g$ is the vector of sample mean outcomes for group g across the S periods preceding treatment, and $\widehat{\boldsymbol{\Sigma}}_g$ is the sample covariance matrix of outcomes for group g across the same S periods.

We make the following additional assumptions for estimation and inference.

Assumption 5 (Regularity Conditions).

(i) $\mathcal{G}^*$ is non-empty

(ii) $p_g > 0$ for all $g \in \mathcal{G}$.

(iii) $\mathbb{E}\left[||\mathbf{Y}||^2\right] < \infty$ where $\mathbf{Y} = (Y_1, \ldots, Y_T)'$.

Assumption 6 (Group Separation). $\mathsf{d}_g = 0$ for all $g \in \mathcal{G}^*$. Let $\mathsf{d}_{\min} = \min_{g \notin \mathcal{G}^*} \mathsf{d}_g$. $\mathsf{d}_{\min} > 0$.

Assumption 7 (Bandwidth)**.** The bandwidth h_n satisfies $h_n \downarrow 0$ and $\sqrt{n_{\min}}\, h_n^2 \to \infty$, as $n_{\min} \to \infty$.

Assumption 8 (Kernel)**.** The kernel function $K : \mathbb{R} \to \mathbb{R}$ satisfies:

  (i) K is symmetric: $K(u) = K(-u)$ for all $u \in \mathbb{R}$;

 (ii) K has compact support: $K(u) = 0$ for $|u| > 1$;

(iii) $K(0) > 0$;

(iv) K is twice continuously differentiable in a neighborhood of 0;

 (v) K is bounded.

Assumption 5 provides several mild regularity conditions. Assumption 6 says that, in population, the d_g should be equal to zero for close comparison groups and be non-zero for groups that are not close comparison groups.

Assumption 7 provides the needed properties of the bandwidth. It should converge to zero, but not too fast relative to the sample size. The condition that $\sqrt{n_{\min}}h_n^2 \to \infty$ can be weakened to $\sqrt{n_{\min}}h_n \to \infty$ to establish consistency, but we use the stronger condition to establish asymptotic normality—it can be considered as a form of undersmoothing where we effectively impose a more stringent requirement on the allowable distance between group g and the treated group in pre-treatment periods for g to be considered a close comparison group.

Assumption 8, particularly the part about the kernel having compact support, rules out some kernels such as the Gaussian kernel, but the assumption allows for a uniform kernel as well as other kernels with compact support, such as an Epanechnikov kernel. Moreover, using a compactly supported kernel provides a natural way to define $\widehat{\mathcal{G}}^* = \left\{ g \in \mathcal{G} : K\left(\widehat{\mathsf{d}}_g/h_n\right) > 0 \right\}$, i.e., the estimated set of close comparison groups is given by all the groups having $K\left(\widehat{\mathsf{d}}_g/h\right) > 0$. We use Assumption 8(iv) to establish asymptotic normality of our estimator, but the differentiability requirement can be relaxed to continuity at $u = 0$ for consistency.

Assumption 9 (Asymptotic linearity of distance estimators)**.** For each $g \in \mathcal{G}$, the distance estimator $\widehat{\mathsf{d}}_g$ admits an asymptotically linear representation

$$\sqrt{n_g}\left(\widehat{\mathsf{d}}_g - \mathsf{d}_g\right) = \frac{1}{\sqrt{n_g}} \sum_{i=1}^{n} \mathbf{1}\{G_i = g\}\, \psi_{\mathsf{d},g}(W_i) + o_p(1),$$

where W_i denotes the observed data for unit i, $\mathbb{E}[\psi_{\mathsf{d},g}(W) \mid G = g] = 0$, and $\mathrm{Var}\left(\psi_{\mathsf{d},g}(W) \mid G = g\right) < \infty$.

Assumption 9 imposes a high-level condition on the distance metric and its estimator. This assumption is likely to be weak in practice as all common distance metrics should satisfy this condition, though it may require additional regularity conditions.

The next proposition provides our first main estimation result.

Proposition 4. Under Assumptions 1 to 9, the estimator $\widehat{\mathcal{G}}^*$ of $\mathcal{G}^*$ is consistent. That is, $P(\widehat{\mathcal{G}}^* = \mathcal{G}^*) \to 1$ as $n_{\min} \to \infty$.

21

The proof is provided in Appendix A. Proposition 4 says that our implied estimator of the close comparison groups, $\widehat{\mathcal{G}}^*$, is consistent for the true set of close comparison groups, $\mathcal{G}^*$, under the assumptions discussed above.

Next, we move towards establishing the asymptotic distribution of our estimator. Towards this end, consider

$$\overset{\circ}{\mathrm{ATT}}_{t_1^*} = \bar{Y}_{1,t_1^*} - \sum_{g \in \mathcal{G}^*} \frac{\widehat{p}_g}{\sum_{g' \in \mathcal{G}^*} \widehat{p}_{g'}} \cdot \bar{Y}_{g,t_1^*}$$

which is the oracle estimator of $\mathrm{ATT}_{t_1^*}$ that knows the true set of close comparison groups $\mathcal{G}^*$.

Finally, the following result provides the limiting distribution of our estimator.

Theorem 2. Under Assumptions 1 to 9,

$$\sqrt{n}\left(\widehat{\mathrm{ATT}}_{t_1^*}(h_n) - \mathrm{ATT}_{t_1^*}\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(W_i) + o_p(1)$$
$$\overset{d}{\to} \mathcal{N}(0, V)$$

where

$$\psi(W) = \frac{\mathbf{1}\{G = 1\}}{p_1}\left(Y_{t_1^*} - \mathbb{E}[Y_{t_1^*} \mid G = 1]\right) - \frac{\mathbf{1}\{G \in \mathcal{G}^*\}}{p_{\mathcal{G}^*}}\left(Y_{t_1^*} - \mathbb{E}[Y_{t_1^*} \mid G \in \mathcal{G}^*]\right)$$
$$\text{and } V = \mathbb{E}\big[\psi(W)^2\big] = \frac{\mathrm{Var}(Y_{t_1^*} \mid G = 1)}{p_1} + \frac{\mathrm{Var}(Y_{t_1^*} \mid G \in \mathcal{G}^*)}{p_{\mathcal{G}^*}}$$

The proof is provided in Appendix A. Our argument to show this result proceeds by showing that, under the conditions discussed above, our estimator is asymptotically equivalent to the oracle estimator above where $\mathcal{G}^*$ is known. Then, our estimator amounts to the difference in average outcomes between units in the treated group and units in any of the treated group's close comparison groups.

7 Extensions

This section discusses several extensions to our main framework. We first show how to incorporate observed covariates. Second, we discuss how to handle settings with multiple treated groups and variation in treatment timing.

7.1 Covariates

All of the arguments discussed above can be extended to hold conditional on covariates. For example, for some $x \in \mathrm{support}(X)$, we can define

$$\mathcal{G}^*(x) := \big\{g \in \bar{\mathcal{G}} \mid (Y_{t_1^*-1} \mid X = x, G = 1) \overset{d}{=} (Y_{t_1^*-1} \mid X = x, G = g)\big\}$$

where $\mathcal{G}^*(x)$ is the conditional-on-covariates version of $\mathcal{G}^*$ from earlier in the paper, which is based on having the same distribution of outcomes in the period immediately preceding treatment;

notice that we could define alternative versions of conditional-on-covariates close comparison groups based on means and/or multiple periods. Then, defining

$$\tau_g(x) := \mathbb{E}[Y_{t_1^*}|X = x, G = 1] - \mathbb{E}[Y_{t_1^*} \mid X = x, G = g] \quad \text{for } g \in \mathcal{G}^*(x),$$

$$\text{and} \quad \tau(x) := \sum_{g \in \mathcal{G}^*(x)} \tau_g(x)w_g(x)$$

for some weighting function $w_g(x)$—e.g., $w_g(x) = \mathrm{P}\big(G = g \mid X = x, G \in \mathcal{G}^*(x)\big)$. Then, using the same arguments as discussed above, it is possible to show conditional-on-covariates versions of the results in Theorem 1 and Propositions 1 and 2, under which

$$\tau(x) = \mathrm{ATT}(x)$$

where $\mathrm{ATT}(x) := \mathbb{E}[Y_{t_1^*}(1) - Y_{t_1^*}(0) \mid X = x, G = 1]$, is the conditional average treatment effect on the treated for $G = 1$. Moreover, it immediately follows that

$$\mathbb{E}[\tau(X) \mid G = 1] = \mathrm{ATT}$$

There are two empirical issues worth briefly discussing. First, the groups composing $\mathcal{G}^*(x)$ could change for different values of the covariates. This does not create any issues except in the case where $\mathcal{G}^*(x)$ is empty for some values of the covariates, in which case $\tau(x)$ is not defined. In this case, a natural approach is to redefine the target parameter to be

$$\widetilde{\mathrm{ATT}} := \mathbb{E}[Y_{t_1^*}(1) - Y_{t_1^*}(0) \mid X \in \bar{\mathcal{X}}, G = 1]$$

where $\bar{\mathcal{X}} := \{x \in \mathrm{support}(X) : \mathcal{G}^*(x) \neq \emptyset\}$. $\widetilde{\mathrm{ATT}}$ is the average treatment effect on treated across only values of the covariates for which we are able to find at least one close comparison group. Re-defining the target parameter based on trimming out certain values of the covariates where there are no available comparison groups is conceptually similar to the idea of trimming when there are no available comparison units due to violations of common support in settings with a single comparison group (Crump, Hotz, Imbens, and Mitnik (2009)).

Second, we leave dealing with high-dimensional covariates (or even continuous covariates) in estimation for future work; however, we note that our estimation results essentially apply immediately for low-dimensional, discrete covariates.

7.2 Multiple Treated Groups and Variation in Treatment Timing

Our approach extends straightforwardly to settings when multiple groups participate in the treatment and where treatment timing can vary across groups. In these settings, it is useful to define the group-time average treatment effect, $ATT(g, t)$, the average treatment effect for group g in time period t. For each (g, t) pair where $t \geq t_g^*$ (i.e., periods after group g first becomes treated), comparison groups can be formed using the procedures outlined in the previous sections. The key difference relative to the case emphasized above with a single treated group is that the set of available comparison groups may vary across different (g, t) pairs. For instance,

in later time periods, the pool of not-yet-treated groups may shrink as more groups become treated. Despite these time-varying comparison group structures, the core logic of our approach applies period-by-period and group-by-group: for each $ATT(g,t)$, one can construct a set of close comparison groups and compare average outcomes in time period t for group g relative to its close comparison groups. Once the $ATT(g,t)$ parameters are estimated, they can be aggregated into more interpretable summary parameters. Common aggregations include event study parameters, which trace out treatment effect dynamics relative to the time of treatment adoption, and an overall average treatment effect that averages across all groups and time periods. The specific aggregation scheme depends on the ultimate target of the analysis; see Callaway and Sant'Anna (2021) and Callaway (2023) for detailed discussions of aggregation strategies and their interpretation.

8   Conclusion

This paper has considered an approach to panel data causal inference based on finding close comparison groups. Rather than using panel data to adjust for differences in unobserved heterogeneity between treated and comparison groups, we use it to search for groups that are similar to the treated group in terms of their pre-treatment outcomes. We demonstrate that this approach is identification-strategy-robust, recovering the average treatment effect on the treated under many different non-nested panel data identification strategies. Our approach also avoids the auxiliary assumptions that are required by the most common approaches to causal inference with panel data, such as difference-in-differences, change-in-changes, and lagged outcome unconfoundedness. The key insight is that comparison groups satisfying relatively simple criteria in pre-treatment periods provide valid comparisons under multiple identification strategies simultaneously, substantially increasing the credibility of causal inferences when such groups exist.

Our approach asks more of the research design by requiring the existence of close comparison groups, which is not necessary for most existing panel data approaches. This is the main downside of our approach. However, when close comparison groups exist, researchers can avoid relying on functional form assumptions and identification strategies that may be difficult to justify. While this restricts the set of applications where our method applies relative to approaches that impose stronger assumptions, it offers substantially more credible inferences in settings where multiple potential comparison groups are available and researchers have access to underlying microdata within each group.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010). "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program". *Journal of the American Statistical Association* 105.490, pp. 493–505.

Arkhangelsky, Dmitry, Susan Athey, David A Hirshberg, Guido W Imbens, and Stefan Wager (2021). "Synthetic difference-in-differences". *American Economic Review* 111.12, pp. 4088–4118.

Arkhangelsky, Dmitry and Guido W Imbens (2022). "Doubly robust identification for causal panel data models". *The Econometrics Journal* 25.3, pp. 649–674.

Arkhangelsky, Dmitry, Guido W Imbens, Lihua Lei, and Xiaoman Luo (2024). "Design-robust two-way-fixed-effects regression for panel data". *Quantitative Economics* 15.4, pp. 999–1034.

Athey, Susan and Guido Imbens (2006). "Identification and inference in nonlinear difference-in-differences models". *Econometrica* 74.2, pp. 431–497.

Athey, Susan, Guido Imbens, Zhaonan Qu, and Davide Viviano (2025). "Triply Robust Panel Estimators". Working Paper.

Callaway, Brantly (2023). "Difference-in-differences for policy evaluation". *Handbook of Labor, Human Resources and Population Economics*. Ed. by Zimmermann, Klaus F. Springer International Publishing, pp. 1–61.

Callaway, Brantly and Tong Li (2023). "Evaluating policies early in a pandemic: Bounding policy effects with nonrandomly missing data". Working Paper.

Callaway, Brantly and Pedro HC Sant'Anna (2021). "Difference-in-differences with multiple time periods". *Journal of Econometrics* 225.2, pp. 200–230.

Callaway, Brantly and Emmanuel Selorm Tsyawo (2023). "Treatment effects in staggered adoption designs with non-parallel trends". Working Paper.

Cattaneo, Matias D., Carlos Diaz, and Rocio Titiunik (2024). "Randomization inference for before-and-after studies with multiple units: An application to a criminal procedure reform in Uruguay". Working Paper.

Chen, Xiaohong, Pedro HC Sant'Anna, and Haitian Xie (2025). "Efficient Difference-in-Differences and Event Study Estimators". Working Paper.

Chernozhukov, Victor, Ivan Fernandez-Val, Jinyong Hahn, and Whitney Newey (2013). "Average and quantile effects in nonseparable panel models". *Econometrica* 81.2, pp. 535–580.

Crump, Richard K, V Joseph Hotz, Guido W Imbens, and Oscar A Mitnik (2009). "Dealing with limited overlap in estimation of average treatment effects". *Biometrika* 96.1, pp. 187–199.

Ding, Peng and Fan Li (2019). "A bracketing relationship between difference-in-differences and lagged-dependent-variable adjustment". *Political Analysis* 27.4, pp. 605–615.

Gobillon, Laurent and Thierry Magnac (2016). "Regional policy evaluation: Interactive fixed effects and synthetic controls". *Review of Economics and Statistics* 98.3, pp. 535–551.

Goodman-Bacon, Andrew (2021). "Difference-in-differences with variation in treatment timing". *Journal of Econometrics* 225.2, pp. 254–277.

Gunsilius, Florian F (2023). "Distributional synthetic controls". *Econometrica* 91.3, pp. 1105–1117.

— (2025). "A primer on optimal transport for causal inference with observational data". Working Paper.

Hausman, Catherine and David S Rapson (2018). "Regression discontinuity in time: Considerations for empirical applications". *Annual Review of Resource Economics* 10.1, pp. 533–552.

Heckman, James and V Joseph Hotz (1989). "Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training". *Journal of the American Statistical Association* 84.408, pp. 862–874.

Imbens, Guido W and Donald B Rubin (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Kellogg, Maxwell, Magne Mogstad, Guillaume A Pouliot, and Alexander Torgovitsky (2021). "Combining matching and synthetic control to tradeoff biases from extrapolation and interpolation". *Journal of the American Statistical Association* 116.536, pp. 1804–1816.

Leavitt, Thomas and Laura A Hatfield (2025). "Averaged Prediction Models (APM): Identifying causal effects in controlled pre-post settings with application to gun policy". *The Annals of Applied Statistics* 19.3, pp. 1826–1846.

Liu, Yiqi (2025). "Synthetic Parallel Trends". Working Paper.

Manski, Charles F and John V Pepper (2018). "How do right-to-carry laws affect crime rates? Coping with ambiguity using bounded-variation assumptions". *Review of Economics and Statistics* 100.2, pp. 232–244.

Marcus, Michelle and Pedro HC Sant'Anna (2021). "The role of parallel trends in event study settings: An application to environmental economics". *Journal of the Association of Environmental and Resource Economists* 8.2, pp. 235–275.

Marx, Philip, Elie Tamer, and Xun Tang (2025). "Heterogeneous Treatment Effects via Linear Dynamic Panel Data Models". Working Paper.

Powell, David, Beth Ann Griffin, and Tal Wolfson (2023). "Estimating policy effects using lagged outcome values to impute counterfactuals". Working Paper.

Rambachan, Ashesh and Jonathan Roth (2023). "A more credible approach to parallel trends". *Review of Economic Studies* 90.5, pp. 2555–2591.

Rincón, Ratzanyel and Kyungchul Song (2025). "Causal Inference with Groupwise Matching". Working Paper.

Sun, Yixiao, Haitian Xie, and Yuhang Zhang (2025). "Difference-in-Differences Meets Synthetic Control: Doubly Robust Identification and Estimation". Working Paper.

Wooldridge, Jeffrey M (2005). "Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models". *Review of Economics and Statistics* 87.2, pp. 385–390.

Xu, Yiqing (2017). "Generalized synthetic control method: Causal inference with interactive fixed effects models". *Political Analysis* 25.1, pp. 57–76.

A   Proofs

A.1   Close Comparison Group Proofs

Proof of Theorem 1. We consider each identification strategy in turn.

(i) Difference-in-Differences: For any group $g \in \mathcal{G}^*$

$$
\begin{aligned}
\tau_g &= \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}[Y_{t_1^*} \mid G = g] \\
&= \mathbb{E}[Y_{t_1^*} \mid G = 1] - \underbrace{\mathbb{E}[Y_{t_1^*-1} \mid G = 1]}_{} - \left(\mathbb{E}[Y_{t_1^*} \mid G = g] - \underbrace{\mathbb{E}[Y_{t_1^*-1} \mid G = g]}_{} \right) \\
&= \mathbb{E}[Y_{t_1^*} - Y_{t_1^*-1} \mid G = 1] - \mathbb{E}[Y_{t_1^*}(0) - Y_{t_1^*-1}(0) \mid G = g] \\
&= \mathbb{E}[Y_{t_1^*} - Y_{t_1^*-1} \mid G = 1] - \mathbb{E}[Y_{t_1^*}(0) - Y_{t_1^*-1}(0) \mid G = 1] \\
&= \mathbb{E}[Y_{t_1^*}(1) - Y_{t_1^*}(0) \mid G = 1] \\
&= \text{ATT}_{t_1^*}
\end{aligned}
$$

where the first equality holds by the definition of τ_g, the two underlined terms in the second line cancel because $g \in \mathcal{G}^*$, the third equality holds by rearranging terms, the fourth equality holds by Assumption DiD-1, the fifth equality holds by rearraning, and the last equality holds by the definition of $\text{ATT}_{t_1^*}$.

This argument holds for all $g \in \mathcal{G}^*$, implying that $\tau_g = \tau = ATT_{t_1^*}$.

(ii) Change-in-Changes: The proof uses the probability integral transform, i.e.:

$$F_Y(Y) \sim \text{Uniform}[0,1] \tag{10}$$

and the inverse probability transform, i.e., for $U \sim \text{Uniform}[0,1]$:

$$Q_Y(U) \sim F_Y \tag{11}$$

For any group $g \in \mathcal{G}^*$,

$$
\begin{aligned}
\tau_g &= \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}[Y_{t_1^*}(0) \mid G = g] \\
&= \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}[Q_{Y_{t_1^*}(0)\mid G=g}(U) \mid G = g] \\
&= \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}\left[Q_{Y_{t_1^*}(0)\mid G=g} \circ F_{Y_{t_1^*-1}(0)\mid G=g}\left(Y_{t_1^*-1}(0)\right) \,\middle|\, G = g\right] \\
&= \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}\left[Q_{Y_{t_1^*}(0)\mid G=1} \circ F_{Y_{t_1^*-1}(0)\mid G=1}\left(Y_{t_1^*-1}(0)\right) \,\middle|\, G = g\right] \\
&= \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}\left[Q_{Y_{t_1^*}(0)\mid G=1} \circ F_{Y_{t_1^*-1}(0)\mid G=1}\left(Y_{t_1^*-1}(0)\right) \,\middle|\, G = 1\right] \\
&= \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}[Q_{Y_{t_1^*}(0)\mid G=1}(U) \mid G = 1] \\
&= \mathbb{E}[Y_{t_1^*}(1) \mid G = 1] - \mathbb{E}[Y_{t_1^*}(0) \mid G = 1] \\
&= \mathrm{ATT}_{t_1^*}
\end{aligned}
$$

where the first equality holds by the definition of τ_g and because group g does not participate in the treatment, the second equality uses Equation (11), the third equality uses Equation (10), the fourth equality holds by Assumption CiC-1, the fifth equality uses that $g \in \mathcal{G}^*$, the sixth equality uses Equation (11), the seventh equality rearranges terms, and the last equality holds by the definition of $\mathrm{ATT}_{t_1^*}$.

(iii) Lagged Outcome Unconfoundedness: For any group $g \in \mathcal{G}^*$

$$
\begin{aligned}
\tau_g &= \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}[Y_{t_1^*} \mid G = g] \\
&= \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}\left[\mathbb{E}\left[Y_{t_1^*}(0) \mid Y_{t_1^*-1}(0), G = g\right]\middle| G = g\right] \\
&= \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}\left[\mathbb{E}\left[Y_{t_1^*}(0) \mid Y_{t_1^*-1}(0), G = g\right]\middle| G = 1\right] \\
&= \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}\left[\mathbb{E}\left[Y_{t_1^*}(0) \mid Y_{t_1^*-1}(0), G = 1\right]\middle| G = 1\right] \\
&= \mathbb{E}\left[Y_{t_1^*}(1) - Y_{t_1^*}(0) \mid G = 1\right] \\
&= \mathrm{ATT}_{t_1^*}
\end{aligned}
$$

where the first equality holds by the definition of τ_g, the second equality holds because observed outcomes for group g are untreated potential outcomes and by the law of iterated expectations, the third equality uses that $g \in \mathcal{G}^*$, the fourth equality holds by Equation (3), the fifth equality rearranges terms, and the last equality holds by the definition of $\mathrm{ATT}_{t_1^*}$. This argument holds for all $g \in \mathcal{G}^*$, implying that $\tau_g = \tau = ATT_{t_1^*}$.

(iv) Interactive Fixed Effects: We show that $g \in \mathcal{G}^*$ is not sufficient for $\tau_g = \mathrm{ATT}_{t_1^*}$ by counterexample. Consider the case where $T = 2$ and $t_1^* = 2$. Suppose that $Y_{t=1} \mid G = 1 \sim \mathcal{N}(1,1)$, and that $g \in \mathcal{G}^*$, hence, $Y_{t=1} \mid G = g \sim \mathcal{N}(1,1)$. Furthermore, suppose that $R = 1$, $F_1 = 1$, $F_2 = 2$, $\mathbb{E}[\eta \mid G = 1] = 1$, $\mathbb{E}[\lambda \mid G = 1] = 0$, $\mathbb{E}[\eta \mid G = g] = 0$, and $\mathbb{E}[\lambda \mid G = g] = 1$. In this case, although group g has the same distribution of untreated potential outcomes in the first period, $1 = \mathbb{E}[Y_{t=2}(0) \mid G = 1] \neq \mathbb{E}[Y_{t=2}(0) \mid G = g] = 2$, which implies that $\tau_g \neq \mathrm{ATT}$.

(v) Latent Unconfoundedness: The same counterexample for IFE works for latent unconfound-

edness. □

Proof of Proposition 1. We organize the proof by each row in the table in Proposition 1.

(i) Difference-in-differences: For $g \in \mathcal{G}^*_{\mathrm{mean}}(1)$, notice that the proof of part (i) of Theorem 1 only relied on the means of pre-treatment outcomes being the same for the treated group as for group g; therefore, the same proof applies in this case.

Next, if $g \in \mathcal{G}^*_{\mathrm{mean}}(S)$, it also follows that $g \in \mathcal{G}^*_{\mathrm{mean}}(1)$, which implies the result.

Likewise, if $g \in \mathcal{G}^*_{\mathrm{dist}}(S)$, it also follows that $g \in \mathcal{G}^*_{\mathrm{mean}}(1)$, which implies the result.

(ii) Change-in-Changes: For $g \in \mathcal{G}^*_{\mathrm{mean}}(1)$, consider the following counterexample. Suppose that $Y_{t^*_1-1} \mid G = g \sim \mathrm{Uniform}[0,1]$, $Y_{t^*_1-1} \mid G = 1 \sim \mathrm{Triangular}(0,1,0.5)$, and that $Y_{t^*_1}(0) \mid G = g \sim \chi^2_1$. Then,

$$
\begin{aligned}
\tau^g_{\mathrm{mean}}(1) &= \mathbb{E}[Y_{t^*_1} \mid G = 1] - \mathbb{E}[Y_{t^*_1} \mid G = g] \\
&= \mathrm{ATT}_{t^*_1} + \mathbb{E}[Y_{t^*_1}(0) \mid G = 1] - \mathbb{E}[Y_{t^*_1}(0) \mid G = g] \\
&= \mathrm{ATT}_{t^*_1} + \mathbb{E}\Big[Q_{Y_{t^*_1}(0)|G=g}\big(\mathrm{F}_{Y_{t^*_1-1}(0)|G=g}(Y_{t^*_1-1})\big) \,\Big|\, G = 1\Big] - \mathbb{E}\Big[Q_{Y_{t^*_1}(0)|G=g}\big(\mathrm{F}_{Y_{t^*_1-1}(0)|G=g}(Y_{t^*_1-1})\big) \,\Big|\, G = g\Big] \\
&= \mathrm{ATT}_{t^*_1} + \mathbb{E}\Big[Q_{Y_{t^*_1}(0)|G=g}(Y_{t^*_1-1}) \,\Big|\, G = 1\Big] - \mathbb{E}\Big[Q_{Y_{t^*_1}(0)|G=g}(Y_{t^*_1-1}) \,\Big|\, G = g\Big] \\
&\approx \mathrm{ATT}_{t^*_1} + 0.91 - 1 \\
&\neq \mathrm{ATT}_{t^*_1}
\end{aligned}
$$

where the first equality holds by the definition of $\tau^g_{\mathrm{mean}}(1)$, the second equality by adding and subtracting $\mathbb{E}[Y_{t^*_1}(0) \mid G = 1]$, the third equality holds by the change-in-changes identification strategy, the simplification in the fourth equality holds by the setup of the counterexample (noting that it is easy to come up with many counterexamples given the previous expression and we choose a particular simple one), the fifth equality holds by approximating the quantiles of a χ^2_1 random variable evaluated at draws from a triangular distribution as in the counterexample and because the mean of a χ^2_1 random variable is 1.

For $g \in \mathcal{G}^*(S)$, essentially the same as the immediately preceding counterexample can be used as it is invariant to the means of the outcomes being equal across additional pre-treatment periods.

Finally, if $g \in \mathcal{G}^*_{\mathrm{dist}}(S)$, it implies that $g \in \mathcal{G}^*$ from Theorem 1, and, therefore, the result follows immediately from the proof of part (ii) of Theorem 1.

(iii) Lagged Outcome Unconfoundedness: For $g \in \mathcal{G}^*_{\mathrm{mean}}(1)$, consider the following counterexample. Suppose that $Y_{t^*_1-1} \mid G = 1 \sim \mathcal{N}(0,\sigma^2_1)$, $Y_{t^*_1-1} \mid G = g \sim \mathcal{N}(0,\sigma^2_g)$, and that

$\mathbb{E}[Y_{t_1^*}(0) \mid Y_{t_1^*-1}(0), G] = Y_{t_1^*-1}^2$. Then,

$$
\begin{aligned}
\tau_{\text{mean}}^g(1) &= \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}[Y_{t_1^*} \mid G = g] \\
&= \text{ATT}_{t_1^*} + \mathbb{E}[Y_{t_1^*}(0) \mid G = 1] - \mathbb{E}[Y_{t_1^*}(0) \mid G = g] \\
&= \text{ATT}_{t_1^*} + \mathbb{E}\big[\mathbb{E}[Y_{t_1^*}(0) \mid Y_{t_1^*-1}(0), G = 1] \mid G = 1\big] - \mathbb{E}\big[\mathbb{E}[Y_{t_1^*}(0) \mid Y_{t_1^*-1}(0), G = g] \mid G = g\big] \\
&= \text{ATT}_{t_1^*} + \mathbb{E}\big[Y_{t_1^*-1}(0)^2 \mid G = 1\big] - \mathbb{E}\big[Y_{t_1^*-1}(0)^2 \mid G = g\big] \\
&= \text{ATT}_{t_1^*} - \sigma_1^2 + \sigma_g^2 \\
&\neq \text{ATT}_{t^*-1}
\end{aligned}
$$

where the first equality holds by the definition of $\tau_{\text{mean}}^g(1)$, the second equality by adding and subtracting $\mathbb{E}[Y_{t_1^*}(0) \mid G = 1]$, the third equality holds by the law of iterated expectations, the fourth equality holds by the setup of the counterexample, and the fifth and sixth equalities hold immediately.

For $g \in \mathcal{G}^*(S)$, essentially the same as the immediately preceding counterexample can be used as it is invariant to the means of the outcomes being equal across additional pre-treatment periods.

Finally, if $g \in \mathcal{G}_{\text{dist}}^*(S)$, it implies that $g \in \mathcal{G}^*$ from Theorem 1, and, therefore, the result follows immediately from the proof of part (iii) of Theorem 1.

(iv) Interactive Fixed Effects: For $g \in \mathcal{G}_{\text{mean}}^*(1)$, the same counterexample as in part (iv) of Theorem 1 implies that this is not a sufficient condition to guarantee that $\tau_{\text{mean}}^g(1) = \text{ATT}_{t_1^*}$.

For $g \in \mathcal{G}_{\text{mean}}^*(S)$, notice that

$$
\begin{pmatrix} \mathbb{E}[Y_{t_1^*-1} \mid G = 1] \\ \vdots \\ \mathbb{E}[Y_{t_1^*-S} \mid G = 1] \end{pmatrix} = \boldsymbol{\theta}^{\text{pre}(t_1^*)} + \mathbf{F}^{\text{pre}(t_1^*)'}\boldsymbol{\Lambda}_1 \quad \text{and} \quad \begin{pmatrix} \mathbb{E}[Y_{t_1^*-1} \mid G = g] \\ \vdots \\ \mathbb{E}[Y_{t_1^*-S} \mid G = g] \end{pmatrix} = \boldsymbol{\theta}^{\text{pre}(t_1^*)} + \mathbf{F}^{\text{pre}(t_1^*)'}\boldsymbol{\Lambda}_g
$$

Then, because $S \geq (R + 1)$ and by Assumption IFE-2*, it follows that $\boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_g$. Next, notice that

$$
\begin{aligned}
\tau_{\text{mean}}^g(S) &= \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}[Y_{t_1^*} \mid G = g] \\
&= \text{ATT}_{t_1^*} + \mathbb{E}[Y_{t_1^*}(0) \mid G = 1] - \mathbb{E}[Y_{t_1^*}(0) \mid G = g] \\
&= \text{ATT}_{t_1^*} + \boldsymbol{\Lambda}_1' \mathbf{F}_{t_1^*} - \boldsymbol{\Lambda}_g' \mathbf{F}_{t_1^*} \\
&= \text{ATT}_{t_1^*}
\end{aligned}
$$

where the first equality holds by the definition of $\tau_{\text{mean}}^g(S)$, the second equality by adding and subtracting $\mathbb{E}[Y_{t_1^*}(0) \mid G = 1]$, the third equality from Assumption IFE-1, and the last equality from the discussion above that showed that $g \in \mathcal{G}_{\text{mean}}^*(S)$ implied that $\boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_g$.

Finally, if $g \in \mathcal{G}_{\text{dist}}^*(S)$, it also follows that $g \in \mathcal{G}_{\text{mean}}^*(S)$, which implies the result using the same argument as above.

(v) Latent Unconfoundedness: For $g \in \mathcal{G}^*_{\text{dist}}(S)$, notice that,

$$
\begin{aligned}
\tau^g_{\text{dist}}(S) &= \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}[Y_{t_1^*} \mid G = g] \\
&= \text{ATT}_{t_1^*} + \mathbb{E}[Y_{t_1^*}(0) \mid G = 1] - \mathbb{E}[Y_{t_1^*}(0) \mid G = g] \\
&= \text{ATT}_{t_1^*} + \mathbb{E}[h_{t_1^*}(\xi) \mid G = 1] - \mathbb{E}[h_{t_1^*}(\xi) \mid G = g] \\
&= \text{ATT}_{t_1^*}
\end{aligned}
$$

where the first equality comes from the definition of $\tau^g_{\text{dist}}(S)$, the second equality holds by adding and subtracting $\mathbb{E}[Y_{t_1^*}(0) \mid G = 1]$, the third equality holds by Assumption Lat. Unc.-1, and the last equality also holds by Assumption Lat. Unc.-2—the last two terms are equal to each other because injectivity implies that groups having the same distribution of observed outcomes $(Y_{t_1^*-1}, \ldots, Y_{t_1^*-S})$ also have the same distribution of unobserved heterogeneity ξ. $\qquad\square$

Proof of Proposition 2. We organize the proof by each row in the table in Proposition 2.

Linear Trends: To start with, notice that the linear trends model for untreated potential outcomes in Equation (8) is a special case of the interactive fixed effects model in Assumption IFE-1 with $F_t = t$ for all t. Moreover, for $g \in \mathcal{G}^*_{\text{mean}}(1)$ and $g \in \mathcal{G}^*_{\text{dist}}(1)$, notice that the same counterexample and proof as in Theorem 1(iv) applies here as it used involved $F_t = t$.

Next, for $g \in \mathcal{G}^*_{\text{mean}}(S)$, let $\mathbf{Y}(0)^{\text{pre}(t_1^*,S)} = \left(Y_{t_1^*-S}(0), \ldots, Y_{t_1^*-1}(0)\right)'$, $\boldsymbol{\theta}^{\text{pre}(t_1^*,S)} = \left(\theta_{t_1^*-S}, \ldots, \theta_{t_1^*-1}\right)'$ and notice that

$$
\mathbb{E}\left[\mathbf{Y}(0)^{\text{pre}(t_1^*,S)} \mid G = 1\right] = \boldsymbol{\theta}^{\text{pre}(t_1^*,S)} + \begin{pmatrix} 1 & t_1^*-1 \\ \vdots & \vdots \\ 1 & t_1^*-S \end{pmatrix} \begin{pmatrix} \mathbb{E}[\eta \mid G = 1] \\ \mathbb{E}[\lambda \mid G = 1] \end{pmatrix}, \text{ and}
$$

$$
\mathbb{E}\left[\mathbf{Y}(0)^{\text{pre}(t_1^*,S)} \mid G = g\right] = \boldsymbol{\theta}^{\text{pre}(t_1^*,S)} + \begin{pmatrix} 1 & t_1^*-1 \\ \vdots & \vdots \\ 1 & t_1^*-S \end{pmatrix} \begin{pmatrix} \mathbb{E}[\eta \mid G = g] \\ \mathbb{E}[\lambda \mid G = g] \end{pmatrix}
$$

The matrix $\begin{pmatrix} 1 & t_1^*-1 \\ \vdots & \vdots \\ 1 & t_1^*-S \end{pmatrix}$ has full rank by construction, which implies that, if $g \in \mathcal{G}^*_{\text{mean}}(S)$, then

$$
\begin{pmatrix} \mathbb{E}[\eta \mid G = 1] \\ \mathbb{E}[\lambda \mid G = 1] \end{pmatrix} = \begin{pmatrix} \mathbb{E}[\eta \mid G = g] \\ \mathbb{E}[\lambda \mid G = g] \end{pmatrix}. \tag{12}
$$

Next, notice that

$$
\begin{aligned}
\tau_{\text{mean}}^g(S) &= \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}[Y_{t_1^*} \mid G = g] \\
&= \text{ATT}_{t_1^*} + \mathbb{E}[Y_{t_1^*}(0) \mid G = 1] - \mathbb{E}[Y_{t_1^*}(0) \mid G = g] \\
&= \text{ATT}_{t_1^*} + \left(\theta_t + \mathbb{E}[\eta \mid G = 1] + \mathbb{E}[\lambda \mid G = 1]t_1^* \right) - \left(\theta_t + \mathbb{E}[\eta \mid G = g] + \mathbb{E}[\lambda \mid G = g]t_1^* \right) \\
&= \text{ATT}_{t_1^*}
\end{aligned}
$$

where the first equality holds by the definition of $\tau_{\text{mean}}^g(S)$, the second equality holds by adding and subtracting $\mathbb{E}[Y_{t_1^*}(0) \mid G = 1]$, the third equality holds from Equation (8), and the last equality holds by Equation (12).

Finally, if $g \in \mathcal{G}_{\text{dist}}^*(S)$, it implies that $g \in \mathcal{G}_{\text{mean}}^*(S)$, and, therefore, the result follows from the discussion above.

Dynamic Panel: For $g \in \mathcal{G}_{\text{mean}}^*(1)$, consider the following counterexample. Suppose that $\theta_{t_1^*} = \theta_{t_1^*-1} = \theta_{t_1^*-2} = 0$, $Y_{t_1^*-2}(0) \mid \eta, G = 1 \sim \mathcal{N}(0,1)$, $Y_{t_1^*-2}(0) \mid \eta, G = g \sim \mathcal{N}(2,1)$, $\eta \mid G = 1 \sim \mathcal{N}(1,1)$, $\eta \mid G = g \sim \mathcal{N}(0,1)$, $e_{t_1^*-1} \mid \eta, Y_{t_1^*-2}(0), G \sim \mathcal{N}(0,1)$, $e_{t_1^*} \mid \eta, Y_{t_1^*-1}(0), G \sim \mathcal{N}(0,1)$, and $\rho = 0.5$. Notice that $\mathbb{E}[\eta \mid G = 1] = 1$, $\mathbb{E}[\eta \mid G = g] = 0$, $\mathbb{E}[Y_{t_1^*-2}(0) \mid G = 1] = 0$, $\mathbb{E}[Y_{t_1^*-2}(0) \mid G = g] = 2$, and $\rho = 0.5$, which suffices for g to be in $\mathcal{G}_{\text{mean}}^*(1)$. Then, notice that

$$
\begin{aligned}
\tau_{\text{mean}}^g(1) &= \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}[Y_{t_1^*} \mid G = g] \\
&= \text{ATT}_{t_1^*} + \mathbb{E}[Y_{t_1^*}(0) \mid G = 1] - \mathbb{E}[Y_{t_1^*}(0) \mid G = g] \\
&= \text{ATT}_{t_1^*} + \left(\theta_t + \mathbb{E}[\eta \mid G = 1] + \rho\mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1] \right) - \left(\theta_t + \mathbb{E}[\eta \mid G = g] + \rho\mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1] \right) \\
&= \text{ATT}_{t_1^*} + \left(1 + 0.5 \times 1 \right) - \left(0 + 0.5 \times 1 \right) \\
&\neq \text{ATT}_{t_1^*}
\end{aligned}
$$

For $g \in \mathcal{G}_{\text{dist}}^*(1)$, we can use the same counterexample. Notice that $Y_{t_1^*-1}(0) \mid G = 1 \sim \mathcal{N}(1, 2.25) \overset{d}{=} Y_{t_1^*-1}(0) \mid G = g$, which implies that $g \in \mathcal{G}_{\text{dist}}^*(1)$ in the counterexample, and the same argument as above implies that $\tau_{\text{dist}}^g(1) \neq \text{ATT}_{t_1^*}$.

For $g \in \mathcal{G}_{\text{mean}}^*(S)$, we have that

$$
\begin{pmatrix} \mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1] \\ \mathbb{E}[Y_{t_1^*-2}(0) \mid G = 1] \end{pmatrix} = \begin{pmatrix} \mathbb{E}[Y_{t_1^*-1}(0) \mid G = g] \\ \mathbb{E}[Y_{t_1^*-2}(0) \mid G = g] \end{pmatrix} \tag{13}
$$

Moreover, notice that

$$
\mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1] = \theta_{t_1^*-1} + \mathbb{E}[\eta \mid G = 1] + \rho\mathbb{E}[Y_{t_1^*-2}(0) \mid G = 1]
$$

which implies that

$$\mathbb{E}[\eta \mid G = 1] = \mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1] - \theta_{t_1^*-1} - \rho\mathbb{E}[Y_{t_1^*-2}(0) \mid G = 1]$$
$$= \mathbb{E}[Y_{t_1^*-1}(0) \mid G = g] - \theta_{t_1^*-1} - \rho\mathbb{E}[Y_{t_1^*-2}(0) \mid G = g]$$
$$= \mathbb{E}[\eta \mid G = g] \tag{14}$$

where the first equality holds by Equation (9), the second equality holds by Equation (13), and the last equality holds by using Equation (9) again. Then,

$$\tau_{\text{mean}}^g(S) = \mathbb{E}[Y_{t_1^*} \mid G = 1] - \mathbb{E}[Y_{t_1^*} \mid G = g]$$
$$= \text{ATT}_{t_1^*} + \mathbb{E}[Y_{t_1^*}(0) \mid G = 1] - \mathbb{E}[Y_{t_1^*}(0) \mid G = g]$$
$$= \text{ATT}_{t_1^*} + \left(\theta_t + \mathbb{E}[\eta \mid G = 1] + \rho\mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1]\right) - \left(\theta_t + \mathbb{E}[\eta \mid G = g] + \rho\mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1]\right)$$
$$= \text{ATT}_{t_1^*}$$

where the first equality holds by the definition of $\tau_{\text{mean}}^g(S)$, the second equality holds by adding and subtracting $\mathbb{E}[Y_{t_1^*}(0) \mid G = 1]$, the third equality holds by Equation (9), and the last equality holds by Equations (13) and (14).

Finally, if $g \in \mathcal{G}^*(S)$, it implies that $g \in \mathcal{G}_{\text{mean}}^*(S)$, and, therefore, the result follows from the discussion above. $\square$

A.2 Time Homogeneity Proofs

Proof of Proposition 3. We organize the proof by each row in the table in Proposition 3.

(i) Difference-in-Differences: Notice that, for any group $g \in \bar{\mathcal{G}}$,

$$\tau_t^{\text{TH}} = \tau_t^{\text{TH}} - \mathbb{E}[Y_t(0) - Y_{t_1^*-1}(0) \mid G = g]$$
$$= \tau_t^{\text{TH}} - \mathbb{E}[Y_t(0) - Y_{t_1^*-1}(0) \mid G = 1]$$
$$= \mathbb{E}[Y_t(1) - Y_t(0) \mid G = 1]$$
$$= \text{ATT}_t$$

where the first equality holds because $t \in \mathcal{T}_{\text{mean}}$, the second equality holds by Assumption DiD-1, the third equality holds by adding and subtracting $\mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1]$ and by the definition of τ_t^{TH}, and the last equality holds by the definition of ATT_t.

(ii) Change-in-Changes: Consider the following counterexample. For $g = 1, \ldots, |\mathcal{G}|$, suppose that $Y_{t_1^*-1}(0) \mid G = g \sim \mathcal{N}(g, 1)$. We suppose that, for all groups, $Q_{Y_t(0)|G=g} \circ F_{Y_{t_1^*-1}(0)|G=g}(y) = y + \varepsilon\, q(y)$, where ε is small and q is a weighted sum of normal density functions placed at the group means:

$$q(y) = c_1\phi(y - 1) + \cdots c_{|\mathcal{G}|}\phi(y - |\mathcal{G}|).$$

where we set $c_1 = 1$, but allow $c_g, g = 2, \ldots, |\mathcal{G}|$ to be constructed in order to allow for mean

32

time homogeneity. In particular, in order for mean time homogeneity to hold for all comparison groups, it must be the case that $\int q(y) f_g(y) \, dy = 0$ for all $g \in \bar{\mathcal{G}}$, where $f_g(y) = \phi(y - g)$ denotes the density of $Y_{t_1^*-1}(0) \mid G = g$. Specifically, let

$$\mathbf{A} = \left[\int \phi(y - g)\phi(y - 1) \, dy \quad \int \phi(y - g)\phi(y - 2) \, dy \quad \cdots \quad \int \phi(y - g)\phi(y - |\mathcal{G}|) \, dy \right]_{g \in \bar{\mathcal{G}}}$$

which is a $|\bar{\mathcal{G}}| \times |\mathcal{G}|$ matrix, so that mean time homogeneity amounts to $\mathbf{A}\mathbf{c} = 0$ where $\mathbf{c} = (c_1, \ldots, c_{|\mathcal{G}|})'$. Since $|\bar{\mathcal{G}}| < |\mathcal{G}|$. This is a system of $|\bar{\mathcal{G}}|$ linear equations with $|\mathcal{G}|$ unknowns, which we can re-write as $\mathbf{A}_{-1}\mathbf{c}_{-1} = -\mathbf{A}_1$ where $\mathbf{A}_{-1}$ is the matrix $\mathbf{A}$ with the first column removed, $\mathbf{c}_{-1} = (c_2, \ldots, c_{|\mathcal{G}|})'$, and $\mathbf{A}_1$ is the first column of $\mathbf{A}$. Then, if we set $\mathbf{c}_{-1} = -(\mathbf{A}_{-1})^{-1}\mathbf{A}_1$, then we have found a q that satisfies mean time homogeneity for all comparison groups.

Next, suppose that f_1 is not a linear combination of f_g's, $g = 2, \ldots, |\mathcal{G}|$. Notice that $\mathbb{E}[Y_{t_1^*}(0) - Y_{t_1^*-1}(0) \mid G = 1] = \varepsilon \int q(y) f_1(y) \, dy = \varepsilon(1 + c_2 \int \phi(y - 1)\phi(y - 2) \, dy + \cdots + c_{|\mathcal{G}|} \int \phi(y - 1)\phi(y - |\mathcal{G}|) \, dy)$, which is generically nonzero since $c_1 = 1$ and the other terms are small (due to the small overlap between normal densities with different means). Thus, $\tau_t^{\mathrm{TH}} \neq \mathrm{ATT}_t$ even though mean time homogeneity holds for all comparison groups.

To give a specific counterexample, consider $|\mathcal{G}| = 3$ with two comparison groups $g \in \{2, 3\}$). Then $f_1(y) = \phi(y - 1)$, $f_2(y) = \phi(y - 2)$, and $f_3(y) = \phi(y - 3)$. The matrix $\mathbf{A}$ is 2×3:

$$\mathbf{A} = \begin{bmatrix} \int \phi(y - 2)\phi(y - 1) \, dy & \int \phi(y - 2)\phi(y - 2) \, dy & \int \phi(y - 2)\phi(y - 3) \, dy \\ \int \phi(y - 3)\phi(y - 1) \, dy & \int \phi(y - 3)\phi(y - 2) \, dy & \int \phi(y - 3)\phi(y - 3) \, dy \end{bmatrix}.$$

Using the fact that $\int \phi(y - j)\phi(y - k) \, dy = \frac{1}{\sqrt{4\pi}} \exp\left(-\frac{(j-k)^2}{4}\right)$, we have:

$$\mathbf{A} = \frac{1}{\sqrt{4\pi}} \begin{bmatrix} e^{-1/4} & 1 & e^{-1/4} \\ e^{-1} & e^{-1/4} & 1 \end{bmatrix}.$$

With $c_1 = 1$, the system $\mathbf{A}_{-1}\mathbf{c}_{-1} = -\mathbf{A}_1$ becomes:

$$\frac{1}{\sqrt{4\pi}} \begin{bmatrix} 1 & e^{-1/4} \\ e^{-1/4} & 1 \end{bmatrix} \begin{bmatrix} c_2 \\ c_3 \end{bmatrix} = -\frac{1}{\sqrt{4\pi}} \begin{bmatrix} e^{-1/4} \\ e^{-1} \end{bmatrix}.$$

Solving yields $c_2 \approx -0.93$ and $c_3 \approx -0.15$. For the treated group:

$$\mathbb{E}[Y_{t_1^*}(0) - Y_{t_1^*-1}(0) \mid G = 1] = \varepsilon \int q(y) \, \phi(y - 1) \, dy$$

$$= \varepsilon \left(\int \phi(y - 1)^2 \, dy + c_2 \int \phi(y - 1)\phi(y - 2) \, dy + c_3 \int \phi(y - 1)\phi(y - 3) \, dy \right)$$

$$= \frac{\varepsilon}{\sqrt{4\pi}} \left(1 - 0.93 \cdot e^{-1/4} - 0.15 \cdot e^{-1} \right)$$

$$\approx \varepsilon \times 0.062 \neq 0.$$

Thus, $\tau_t^{\mathrm{TH}} \neq \mathrm{ATT}_t$ even though mean time homogeneity holds for both comparison groups.

(iii) Lagged Outcome Unconfoundedness: Suppose that

$$\mathbb{E}[Y_{t_1^*}(0) \mid Y_{t_1^*-1}(0), G = g] = \psi(Y_{t_1^*-1}(0))'\beta$$

for $\psi(y)$ a J-dimensional vector of basis functions. Further, let

$$\boldsymbol{\Psi} = \left[\mathbb{E}\big[\psi(Y_{t_1^*-1}(0)) \mid G = g\big]'\right]_{g \in \bar{\mathcal{G}}}$$

which is a $|\bar{\mathcal{G}}| \times J$ matrix. Suppose $\operatorname{rank}(\boldsymbol{\Psi}) < J$ (e.g., the number of comparison groups could be less than the complexit of $\mathbb{E}[Y_{t_1^*}(0) \mid Y_{t_1^*-1}(0), G = g]$ or the comparison groups do not contain enough independent information relative to the complexist of $\mathbb{E}[Y_{t_1^*}(0) \mid Y_{t_1^*-1}(0), G = g]$), so that $\boldsymbol{\Psi}$ does not have full rank. Notice that $t \in \mathcal{T}_{\text{mean}}$ implies that $\boldsymbol{\Psi}\beta = \left[\mathbb{E}[Y_{t_1^*-1}(0) \mid G = g]\right]_{g \in \bar{\mathcal{G}}}$. However, because $\boldsymbol{\Psi}$ does not have full rank, there are many solutions to this system of equations, and, in particular, there exists a solution β^* such that $\mathbb{E}[Y_{t_1^*}(0) \mid G = 1] = \mathbb{E}\big[\psi(Y_{t_1^*-1}(0)) \mid G = 1\big]'\beta \neq \mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1]$, implying $\tau_t^{\text{TH}} \neq \text{ATT}_t$.

Consider the following specific counterexample. For some constants γ and k, suppose that $Y_t(0) = Y_{t_1^*-1}(0) + \gamma(Y_{t_1^*-1}(0)^2 - k) + e$ where $e \mid Y_{t_1^*-1}(0), G = g \sim \mathcal{N}(0, \sigma_e^2)$. Notice that this implies that $Y_t(0) \perp\!\!\!\perp G \mid Y_{t_1^*-1}(0)$, so that lagged outcome unconfoundedness holds. Further, suppose that $Y_{t_1^*-1}(0) \mid G = 1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, and for $g \in \bar{\mathcal{G}}$, $Y_{t_1^*-1}(0) \mid G = g \sim \mathcal{N}(\mu_g, \sigma_g^2)$ and satisfying $\mu_g^2 + \sigma_g^2 = k$, but $\mu_1^2 + \sigma_g^2 \neq k$. In this case,

$$\boldsymbol{\Psi} = \begin{bmatrix} 1 & \mu_2 & \mu_2^2 + \sigma_2^2 \\ 1 & \mu_3 & \mu_3^2 + \sigma_3^2 \\ \vdots & \vdots & \vdots \\ 1 & \mu_{|\mathcal{G}|} & \mu_{|\mathcal{G}|}^2 + \sigma_{|\mathcal{G}|}^2 \end{bmatrix} = \begin{bmatrix} 1 & \mu_2 & k \\ 1 & \mu_3 & k \\ \vdots & \vdots & \vdots \\ 1 & \mu_{|\mathcal{G}|} & k \end{bmatrix}$$

which has rank at most $2 < J = 3$. And, more specifically, it holds that

$$
\begin{aligned}
\tau_t^{\text{TH}} &= \mathbb{E}[Y_t(1) \mid G = 1] - \mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1] \\
&= \text{ATT}_t + \mathbb{E}[Y_t(0) \mid G = 1] - \mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1] \\
&= \text{ATT}_t + \mathbb{E}\big[\mathbb{E}[Y_t(0) \mid Y_{t_1^*-1}(0), G = 1] \mid G = 1\big] - \mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1] \\
&= \text{ATT}_t + \mathbb{E}\big[\mathbb{E}[Y_t(0) \mid Y_{t_1^*-1}(0), G = g] \mid G = 1\big] - \mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1] \\
&= \text{ATT}_t + \mathbb{E}\Big[Y_{t_1^*-1}(0) + \gamma(Y_{t_1^*-1}(0)^2 - k) \Big| G = 1\Big] - \mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1] \\
&= \text{ATT}_t + \gamma(\mu_1^2 + \sigma_1^2 - k) \\
&\neq \text{ATT}_t
\end{aligned}
$$

where the first equality holds by the definition of τ_t^{TH}, the second equality holds by adding and subtracting $\mathbb{E}[Y_t(0) \mid G = 1]$, the third equality holds by the law of iterated expectations, the fourth equality holds by the lagged outcome unconfoundedness identification strategy, the fifth equality holds by the setting of the counterexample, and the last two equalities are immediate simplifications of the previous line.

<u>(iv) Interactive Fixed Effects</u>: Define

$$\bar{\mathbf{\Lambda}} = \begin{bmatrix} 1 & \mathbb{E}[\lambda'|G=g] \end{bmatrix}_{g \in \bar{\mathcal{G}}} \tag{15}$$

where we use the notation $\begin{bmatrix} \cdot \end{bmatrix}_{g \in \bar{\mathcal{G}}}$ to indicate a matrix with one row for each $g \in \bar{\mathcal{G}}$; thus, $\bar{\mathbf{\Lambda}}$ is a $|\bar{\mathcal{G}}| \times (R+1)$ dimensional matrix, where R continues to be defined as the number of factors in Equation (4). We assume that $\text{rank}(\bar{\mathbf{\Lambda}}) = R+1$, which amounts to assuming that the groups are "different enough" from each other in terms of the mean of λ. This also implies that the number of groups is larger than the number of interactive fixed effects. That $t_1^* \in \mathcal{T}_{\text{mean}}(1)$ implies that

$$\mathbf{0}_{|\bar{\mathcal{G}}|} = \begin{bmatrix} \mathbb{E}[\Delta Y_{t_1^*}|G=g] \end{bmatrix}_{g \in \bar{\mathcal{G}}}$$
$$= \bar{\mathbf{\Lambda}} \begin{bmatrix} \Delta \theta_{t_1^*} \\ \Delta F_{t_1^*} \end{bmatrix}$$

Since $\bar{\mathbf{\Lambda}}$ has full column rank, this implies that

$$\mathbf{0}_{R+1} = \begin{bmatrix} \Delta \theta_{t_1^*} \\ \Delta F_{t_1^*} \end{bmatrix} \tag{16}$$

Thus,

$$\begin{aligned}
\text{ATT}_{t_1^*} &= \mathbb{E}[Y_{t_1^*}(1) - Y_{t_1^*}(0) \mid G = 1] \\
&= \tau^{\text{TH}} - \mathbb{E}[Y_{t_1^*}(0) - Y_{t_1^*-1}(0) \mid G = 1] \\
&= \tau^{\text{TH}} + \left(\Delta \theta_{t_1^*} + \mathbb{E}[\lambda' \mid G = 1] \Delta F_{t_1^*} \right) \\
&= \tau^{\text{TH}}
\end{aligned}$$

where the first equality holds by the definition of $\text{ATT}_{t_1^*}$, the second equality holds by adding and subtracting $\mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1]$, the third equality holds from the interactive fixed effects model for untreated potential outcomes in Equation (4) and by Assumption IFE-1, and the last equality holds by Equation (16).

<u>(v) Latent Unconfoundedness:</u> Consider the following counterexample. Suppose that $h_t(\eta_i) = \theta_t + \eta_i + \lambda_i' F_t + e_{it}$ where $R+1 > |\bar{\mathcal{G}}|$ (i.e., the number of latent factors plus one exceeds the number of comparison groups). Then, mean stability for all comparison groups implies $\text{rank}(\bar{\mathbf{\Lambda}}) < R+1$ where $\bar{\mathbf{\Lambda}}$ is defined as in Equation (15). Therefore, there exists a non-zero vector $\begin{bmatrix} \Delta \theta_t \\ \Delta F_t \end{bmatrix}'$ such that $\bar{\mathbf{\Lambda}} \begin{bmatrix} \Delta \theta_t \\ \Delta F_t \end{bmatrix} = \mathbf{0}_{|\bar{\mathcal{G}}|}$. This implies that mean stability for all comparison groups does not force $\mathbb{E}[Y_{t_1^*}(0) - Y_{t_1^*-1}(0) \mid G = 1] = 0$, and, hence, does not imply that $\tau_{t_1^*}^{\text{TH}} = \text{ATT}_{t_1^*}$.

$\square$

A.2.1 Additional Results for Change-in-Changes

In line with the discussion in Section 4, define

$$t_{\text{dist}}^{\text{TH}} := \max\left\{ s \in \{t_1^*, \ldots, T\} \mid (Y_{s'}(0)|G = g) \overset{d}{=} (Y_{t_1^*-1}(0)|G = g) \text{ for all } s' \leq s \text{ and } g \in \bar{\mathcal{G}} \right\}$$

which is the last period after treatment adoption where the distribution of untreated potential outcomes is the same as in period $t_1^* - 1$ for all groups in $\bar{\mathcal{G}}$. Also, define

$$\mathcal{T}_{\text{dist}} = \left\{ t \in \{t_1^*, \ldots, T\} \mid t \leq t_{\text{dist}}^{\text{TH}} \right\}$$

which is the set of periods after treatment adoption where the distribution of untreated potential outcomes is the same as in period $t_1^* - 1$ for all groups in $\bar{\mathcal{G}}$. Then, we have the following result for τ_t^{TH} if the change-in-changes identification strategy and under "distributional time homogeneity".

Proposition 5. Suppose that the change-in-changes identification strategy holds. Then, for any $t \in \mathcal{T}_{\text{dist}}$

$$\tau_t^{\text{TH}} = \text{ATT}_t$$

Proof. Notice that, for any group $g \in \bar{\mathcal{G}}$,

$$
\begin{aligned}
\tau_t^{\text{TH}} &= \mathbb{E}[Y_t(1) \mid G = 1] - \mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1] \\
&= \mathbb{E}[Y_t(1) \mid G = 1] - \mathbb{E}\left[Q_{Y_t(0)|G=g} \circ \text{F}_{Y_{t_1^*-1}(0)|G=g}\big(Y_{t_1^*-1}(0)\big) \;\Big|\; G = 1 \right] \\
&= \mathbb{E}[Y_t(1) \mid G = 1] - \mathbb{E}\left[Q_{Y_t(0)|G=1} \circ \text{F}_{Y_{t_1^*-1}(0)|G=1}\big(Y_{t_1^*-1}(0)\big) \;\Big|\; G = 1 \right] \\
&= \mathbb{E}[Y_t(1) \mid G = 1] - \mathbb{E}\left[Q_{Y_t(0)|G=1}(U) \;\Big|\; G = 1 \right] \\
&= \mathbb{E}[Y_t(1) \mid G = 1] - \mathbb{E}[Y_t(0) \mid G = 1] \\
&= \text{ATT}_t
\end{aligned}
$$

where the first equality holds by the definition of τ_t^{TH}, the second equality holds because $t \in \mathcal{T}_{\text{dist}}$ (i.e., the distribution of untreated potential outcomes is the same in period t and period $t_1^* - 1$ for group g), the third equality holds by the CiC identification strategy (Assumption CiC-1 in particular), the fourth equality holds because $\text{F}_{Y_{t_1^*-1}(0)|G=1}(Y_{t_1^*-1}(0))|G = 1$ follows a Uniform$[0, 1]$ distribution, the fifth equality holds by the probability inverse transform with $U \mid G = 1 \sim \text{Uniform}[0, 1]$, and the last equality holds by the definition of ATT_t. $\square$

A.2.2 Additional Results for Lagged Outcome Unconfoundedness

The next result shows that under lagged outcome unconfoundedness and with a flexible parametric model for $\mathbb{E}[Y_{t_1^*}(0) \mid Y_{t_1^*-1}(0), G = g]$, if time there are enough comparison groups that exhibit enough independent variation, then time homogeneity among the comparison groups implies that time homogeneity also holds for the treated group.

Proposition 6. Suppose that $\mathbb{E}[Y_{t_1^*}(0) \mid Y_{t_1^*-1}(0), G = g] = \psi\big(Y_{t_1^*-1}(0)\big)'\beta$ where $\psi(y)$ is a J-

dimensional vector of basis functions that includes the identity function, i.e., $\psi_j(y) = y$ for some j (this requirement is met by typical polynomial, spline, or other flexible basis functions). Further, let $\boldsymbol{\Psi}$ be the $|\bar{\mathcal{G}}| \times J$ matrix with rows $\mathbb{E}[\psi(Y_{t_1^*-1}(0)) \mid G = g]'$ for $g \in \bar{\mathcal{G}}$, and suppose that $\boldsymbol{\Psi}$ has full column rank J. If $\text{rank}(\boldsymbol{\Psi}) = J$, then time homogeneity for the comparison groups implies that time mean time homogeneity also applies for the treated group, i.e., $\mathbb{E}[Y_{t_1^*}(0) \mid G = 1] = \mathbb{E}[Y_{t_1^*-1}(0) \mid G = 1]$, and, hence, $\tau_{t_1^*}^{\text{TH}} = \text{ATT}_{t_1^*}$.

Proof. Without loss of generality, assume that the first element of $\psi(y)$ is the identity function, i.e., $\psi_1(y) = y$. Next, notice that time homogeneity implies that

$$\boldsymbol{\Psi}\beta = \Big[\mathbb{E}[Y_{t_1^*-1}(0) \mid G = g]\Big]_{g \in \bar{\mathcal{G}}}$$

Since $\boldsymbol{\Psi}$ has full column rank, this system has a unique solution for β. Moreover, notice that

$$\boldsymbol{\Psi}\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix} = \Big[\mathbb{E}[Y_{t_1^*-1}(0) \mid G = g] \quad \mathbb{E}[\psi_2(Y_{t_1^*-1}(0)) \mid G = g] \quad \cdots \quad \mathbb{E}[\psi_J(Y_{t_1^*-1}(0)) \mid G = g]\Big]_{g \in \bar{\mathcal{G}}}\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}$$

$$= \Big[\mathbb{E}[Y_{t_1^*-1}(0) \mid G = g]\Big]_{g \in \bar{\mathcal{G}}}$$

This implies that the unique solution to the system is $\beta = \begin{pmatrix}1 & 0 & \cdots & 0\end{pmatrix}'$, and, hence,

$$\mathbb{E}[Y_{t_1^*}(0) \mid Y_{t_1^*-1}(0), G = g] = \psi(Y_{t_1^*-1}(0))'\beta = Y_{t_1^*-1}(0)$$

for all $g \in \bar{\mathcal{G}}$. Therefore,

$$\mathbb{E}[Y_{t_1^*}(0) \mid G = 1] = \mathbb{E}\big[\mathbb{E}[Y_{t_1^*}(0) \mid Y_{t_1^*-1}(0), G = 1] \mid G = 1\big]$$
$$= \mathbb{E}\big[Y_{t_1^*-1}(0) \mid G = 1\big]$$

which implies that $\tau_{t_1^*}^{\text{TH}} = \text{ATT}_{t_1^*}$. $\qquad\square$

The takeaway is that (i) in general, mean stability for multiple comparison groups does not guarantee mean stability for the treated group under lagged outcome unconfoundedness, but (ii) under a flexible parametric model for the transition function $\mathbb{E}[Y_{t_1^*}(0) \mid Y_{t_1^*-1}(0)]$ can restore mean stability for the treated group, and more comparison groups allow for more flexibility in $\mathbb{E}[Y_{t_1^*}(0) \mid Y_{t_1^*-1}(0)]$ while still delivering mean stability for the treated group.

A.3 Estimation and Inference Proofs

Lemma 1 (Distance convergence rate). Under Assumption 9,

$$\max_{g \in \mathcal{G}} |\widehat{d}_g - d_g| = O_p(n_{\min}^{-1/2}).$$

where $n_{\min} := \min_{g \in \mathcal{G}} n_g$.

Proof of Lemma 1. This result is an immediate implication of Assumption 9 because $\mathcal{G}$ is a finite set. □

Lemma 2 (No false inclusions)**.** Under Assumptions 1 to 9, if $g \notin \mathcal{G}^*$, then $g \notin \widehat{\mathcal{G}^*}$ as $n_{\min} \to \infty$. That is, for all $g \notin \mathcal{G}^*$, $\mathrm{P}\big(\mathrm{K}(\widehat{\mathsf{d}}_g/h_n) = 0\big) \to 1$ as $n_{\min} \to \infty$, which implies $\mathrm{K}\big(\widehat{\mathsf{d}}_g/h_n\big) \xrightarrow{p} 0$ as $n_{\min} \to \infty$.

Proof of Lemma 2. Suppose that $g \notin \mathcal{G}^*$. By Assumption 6, $\mathsf{d}_g \geq \mathsf{d}_{\min} > 0$. Take any $\epsilon \in (0, \mathsf{d}_{\min})$. Notice that:

$$
\begin{aligned}
\mathrm{P}(\widehat{\mathsf{d}}_g \geq \mathsf{d}_{\min} - \epsilon) &= 1 - \mathrm{P}(\widehat{\mathsf{d}}_g < \mathsf{d}_{\min} - \epsilon) \\
&\geq 1 - \mathrm{P}(\widehat{\mathsf{d}}_g < \mathsf{d}_g - \epsilon) \\
&= 1 - \mathrm{P}(|\widehat{\mathsf{d}}_g - \mathsf{d}_g| > \epsilon) \\
&\to 1, \text{ as } n_{\min} \to \infty
\end{aligned}
\tag{17}
$$

where the first line holds by taking the complement; the second line holds since $\mathsf{d}_g \geq \mathsf{d}_{\min}$; the third line holds by rearranging terms and taking the absolute value; and the last line holds by Assumption 9, which implies that $|\widehat{\mathsf{d}}_g - \mathsf{d}_g| = o_p(1)$.

Next, take the ratio $\widehat{\mathsf{d}}_g/h_n$. On the event $\{\widehat{\mathsf{d}}_g \geq \mathsf{d}_{\min} - \epsilon\}$, which occurs with probability approaching one as $n_{\min} \to \infty$ by Equation (17), we have that

$$
\frac{\widehat{\mathsf{d}}_g}{h_n} \geq \frac{\mathsf{d}_{\min} - \epsilon}{h_n} > 1
\tag{18}
$$

where the second inequality holds for sufficiently large n since $h_n \to 0$ and $\mathsf{d}_{\min}$ and ϵ are fixed with $\mathsf{d}_{\min} > \epsilon > 0$. Next, notice that

$$
\begin{aligned}
\mathrm{P}\big(\mathrm{K}(\widehat{\mathsf{d}}_g/h_n) = 0\big) &\geq \mathrm{P}\big(\widehat{\mathsf{d}}_g/h_n > 1\big) \\
&\geq \mathrm{P}\big(\widehat{\mathsf{d}}_g/h_n \geq (\mathsf{d}_{\min} - \epsilon)/h_n\big) \\
&= \mathrm{P}\big(\widehat{\mathsf{d}}_g \geq \mathsf{d}_{\min} - \epsilon\big) \\
&\to 1, \text{ as } n_{\min} \to \infty
\end{aligned}
\tag{19}
$$

where the first line holds since $u > 1 \implies \mathrm{K}(u) = 0$ and because $\widehat{\mathsf{d}}_g/h_n > 1$ by construction, the second line holds by Equation (18) for n large enough, the third line holds immediately by canceling the h_n terms, and the fourth line holds by Equation (17) above. Thus, for any $g \notin \mathcal{G}^*$, $\mathrm{P}\big(\mathrm{K}(\widehat{d}_g/h_n) = 0\big) \to 1$ as $n_{\min} \to \infty$.

□

Lemma 3 (Exclusion of non-close comparison groups from $\widehat{\mathcal{G}}^*$)**.** Under Assumptions 1 to 9, the probability that any $g \notin \mathcal{G}^*$ is incorrectly included in the estimated set of close comparison groups, $\widehat{\mathcal{G}}^*$, converges to zero as $n_{\min} \to \infty$. That is, $\mathrm{P}\big(\widehat{\mathcal{G}}^* \subseteq \mathcal{G}^*\big) \to 1$ as $n_{\min} \to \infty$.

Proof of Lemma 3. Consider the probability that there exists at least some group $g \notin \mathcal{G}^*$ that receives positive weight:

$$
\begin{aligned}
\mathrm{P}\left(\bigcup_{g \notin \mathcal{G}^*} \mathrm{K}(\widehat{\mathsf{d}}_g/h_n) > 0 \right) &\leq \sum_{g \notin \mathcal{G}^*} \mathrm{P}\big(\mathrm{K}(\widehat{\mathsf{d}}_g/h_n) > 0\big) \\
&= \sum_{g \notin \mathcal{G}^*} \Big(1 - \mathrm{P}\big(\mathrm{K}(\widehat{\mathsf{d}}_g/h_n) = 0\big) \Big) \\
&\to \sum_{g \notin \mathcal{G}^*} (1 - 1) \text{ as } n_{\min} \to \infty \\
&= 0
\end{aligned}
\tag{20}
$$

where the first line holds by Boole's inequality; the second line holds by taking the complement and that $\mathrm{K}(\cdot)$ is non-negative; the third line holds by Equation (19); and the fourth line holds trivially. That is, for all groups $g \notin \mathcal{G}^*$, the probability that non-credible groups receive zero weight converges to 1:

$$
\mathrm{P}\left(\bigcap_{g \notin \mathcal{G}^*} \mathrm{K}(\widehat{\mathsf{d}}_g/h_n) = 0 \right) = 1 - \mathrm{P}\left(\bigcup_{g \notin \mathcal{G}^*} \mathrm{K}(\widehat{\mathsf{d}}_g/h_n) > 0 \right) \to 1 \text{ as } n_{\min} \to \infty
$$

where the equality holds by De Morgan's law, and the convergence holds by Equation (20). Hence, $\mathrm{P}\big(\{g \in \mathcal{G}^* : \mathrm{K}(\widehat{\mathsf{d}}_g/h_n) > 0\} \subseteq \mathcal{G}^*\big) = \mathrm{P}(\widehat{\mathcal{G}}^* \subseteq \mathcal{G}^*) \to 1$, as $n \to \infty$. □

Lemma 4 (No false exclusions)**.** Under Assumptions 1 to 9, for all $g \in \mathcal{G}^*$, $\mathrm{K}(\widehat{\mathsf{d}}_g/h_n) \xrightarrow{p} \mathrm{K}(0)$ as $n_{\min} \to \infty$. Moreover, $\mathrm{P}\big(\mathrm{K}(\widehat{\mathsf{d}}_g/h_n) > 0\big) \to 1$ as $n_{\min} \to \infty$.

Proof of Lemma 4. For any $g \in \mathcal{G}^*$, by definition, $d_g = 0$. From Assumption 9, this implies $\widehat{d}_g = O_p(n_{\min}^{-1/2})$. For bandwidth $h_n > 0$:

$$
\frac{\widehat{d}_g}{h_n} = \frac{O_p(n_{\min}^{-1/2})}{h_n} = O_p\left(\frac{1}{\sqrt{n_{\min}}h_n} \right) \xrightarrow{p} 0
$$

where convergence in probability follows from Assumption 7, which ensures $\sqrt{n_{\min}}h_n \to \infty$.

Since K is continuous at 0 by Assumption 8, the continuous mapping theorem implies:

$$
\mathrm{K}\big(\widehat{d}_g/h_n\big) \xrightarrow{p} \mathrm{K}(0)
$$

which completes the first part of the proof. Next, for any $\epsilon \in \big(0, \mathrm{K}(0)\big)$, where $\mathrm{K}(0) > 0$ by Assumption 8, it follows that

$$
\mathrm{P}\Big(\mathrm{K}(\widehat{d}_g/h_n) > 0\Big) \geq \mathrm{P}\big(\mathrm{K}(\widehat{d}_g/h_n) > \epsilon\big) \to 1 \text{ as } n_{\min} \to \infty.
$$

□

Lemma 5 (Inclusion of all close comparison groups in $\widehat{\mathcal{G}}^*$). Under Assumptions 1 to 9, the probability that any group $g \in \mathcal{G}^*$ receives positive weight converges to one as $n_{\min} \to \infty$. Thus, $\mathrm{P}(\widehat{\mathcal{G}}^* \supseteq \mathcal{G}^*) \to 1$ as $n_{\min} \to \infty$.

Proof of Lemma 5. Consider the probability that all groups in $\mathcal{G}^*$ get positive weight

$$
\begin{aligned}
\mathrm{P}\left(\bigcap_{g \in \mathcal{G}^*} \mathrm{K}(\widehat{\mathsf{d}}_g/h_n) > 0 \right) &= 1 - \mathrm{P}\left(\bigcup_{g \in \mathcal{G}^*} \mathrm{K}(\widehat{\mathsf{d}}_g/h_n) = 0 \right) \\
&\geq 1 - \sum_{g \in \mathcal{G}^*} \mathrm{P}\left(\mathrm{K}(\widehat{\mathsf{d}}_g/h_n) = 0 \right) \\
&\to 1 - \sum_{g \in \mathcal{G}^*} 0 \text{ as } n_{\min} \to \infty \\
&= 1
\end{aligned}
$$

where the first line holds by De Morgan's law—taking the complement and since the kernel function is non-negative; the second line holds by Boole's inequality; and the third line holds by Lemma 4. This result states that the probability that all groups that belong in $\mathcal{G}^*$ are included in the estimated $\widehat{\mathcal{G}}^*$ converges to one as $n_{\min} \to \infty$, i.e., that $\mathrm{P}(\widehat{\mathcal{C}^*} \supseteq \mathcal{C}^*) \to 1$, as $n_{\min} \to \infty$. $\square$

Proof of Proposition 4. The result holds immediately by Lemmas 3 and 5. $\square$

Proposition 7. Under Assumptions 1 to 9, our estimator of $\mathrm{ATT}_{t_1^*}$ is consistent, i.e.,

$$
\widehat{\mathrm{ATT}}_{t_1^*}(h_n) \xrightarrow{p} \mathrm{ATT}_{t_1^*}
$$

Proof of Proposition 7. Notice that, for each $g \in \mathcal{G}$,

$$
\widehat{p}_g \cdot \mathrm{K}(\widehat{d}_g/h_n) \xrightarrow{p} \begin{cases} p_g \cdot \mathrm{K}(0), & g \in \mathcal{G}^*, \\ 0, & g \notin \mathcal{G}^*, \end{cases} \tag{21}
$$

and further that

$$
\widehat{p}_g \cdot \mathrm{K}(\widehat{d}_g/h_n) \cdot \bar{Y}_{g,t_1^*} \xrightarrow{p} \begin{cases} p_g \cdot \mathrm{K}(0) \cdot \mathbb{E}[Y_{t_1^*} \mid G = g], & g \in \mathcal{G}^*, \\ 0, & g \notin \mathcal{G}^*, \end{cases} \tag{22}
$$

which holds by Lemmas 2 and 4 and Slutsky's theorem. This implies that

$$\widehat{\mathrm{ATT}}_{t_1^*}(h_n) = \bar{Y}_{1,t_1^*} - \sum_{g \in \mathcal{G}} \frac{\widehat{p}_g \cdot \mathrm{K}(\widehat{\mathsf{d}}_g/h_n)}{\sum_{g' \in \mathcal{G}} \widehat{p}_{g'} \cdot \mathrm{K}(\widehat{\mathsf{d}}_{g'}/h_n)} \cdot \bar{Y}_{g,t_1^*}$$

$$\xrightarrow{p} \mathbb{E}[Y_{t_1^*} \mid G = 1] - \sum_{g \in \mathcal{G}^*} \left(\frac{p_g \cdot K(0)}{\sum_{g' \in \mathcal{G}^*} p_{g'} \cdot K(0)} \right) \cdot \mathbb{E}[Y_{t_1^*} \mid G = g]$$

$$= \mathbb{E}[Y_{t_1^*} \mid G = 1] - \sum_{g \in \mathcal{G}^*} \left(\frac{p_g}{\sum_{g' \in \mathcal{G}^*} p_{g'}} \right) \cdot \mathbb{E}[Y_{t_1^*} \mid G = g]$$

$$= \mathrm{ATT}_{t_1^*}$$

$\square$

Lemma 6 (Kernel convergence). Under Assumptions 1 to 9,

$$\sqrt{n_{\min}} \sum_{g \in \mathcal{G}} \widehat{p}_g \cdot \left(\mathrm{K}(\widehat{d}_g/h_n) - \mathrm{K}(0)\mathbf{1}\{g \in \mathcal{G}^\star\} \right) \cdot \bar{Y}_{g,t_1^*} \xrightarrow{p} 0.$$

Proof of Lemma 6. Notice that we can decompose the sum into two parts:

$$\sum_{g \in \mathcal{G}} \widehat{p}_g \cdot \left(\mathrm{K}(\widehat{\mathsf{d}}_g/h_n) - \mathrm{K}(0)\mathbf{1}\{g \in \mathcal{G}^*\} \right) \cdot \bar{Y}_{g,t_1^*}$$

$$= \underbrace{\sum_{g \in \mathcal{G}^\star} \widehat{p}_g \cdot \left(\mathrm{K}(\widehat{\mathsf{d}}_g/h_n) - \mathrm{K}(0) \right) \cdot \bar{Y}_{g,t_1^*}}_{\mathrm{I}_n} + \underbrace{\sum_{g \notin \mathcal{G}^\star} \widehat{p}_g \cdot \left(\mathrm{K}(\widehat{\mathsf{d}}_g/h_n) - 0 \right) \cdot \bar{Y}_{g,t_1^*}}_{\mathrm{II}_n}$$

and we continue by showing both terms are $o_p(1)$.

For term I_n, by Assumption 8, K is a symmetric function that is twice continuously differentiable at 0, which implies that $\mathrm{K}'(0) = 0$. By a Taylor expansion for each $g \in \mathcal{G}^*$, we have:

$$\mathrm{K}\left(\widehat{\mathsf{d}}_g/h_n\right) = \mathrm{K}(0) + \mathrm{K}'(0) \cdot \left(\frac{\widehat{\mathsf{d}}_g}{h_n} - 0 \right) + \frac{1}{2}\mathrm{K}''(0) \cdot \left(\frac{\widehat{\mathsf{d}}_g}{h_n} - 0 \right)^2 + o_p\left(\left(\frac{\widehat{\mathsf{d}}_g}{h_n} - 0 \right)^2 \right)$$

which implies

$$\mathrm{K}\left(\widehat{\mathsf{d}}_g/h_n\right) - \mathrm{K}(0) = \mathrm{K}'(0) \cdot \left(\widehat{\mathsf{d}}_g/h_n\right) + \frac{1}{2}\mathrm{K}''(0) \cdot \left(\widehat{\mathsf{d}}_g/h_n\right)^2 + o_p\left((\widehat{\mathsf{d}}_g/h_n)^2\right)$$

$$= \frac{1}{2}\mathrm{K}''(0) \cdot \left(\widehat{\mathsf{d}}_g/h_n\right)^2 + o_p\left((\widehat{\mathsf{d}}_g/h_n)^2\right)$$

$$= O_p\left((\widehat{\mathsf{d}}_g/h_n)^2\right) + o_p\left((\widehat{\mathsf{d}}_g/h_n)^2\right)$$

$$= O_p\left((\widehat{\mathsf{d}}_g/h_n)^2\right) \tag{23}$$

where the first equality holds immediately, the second equality holds since $\mathrm{K}'(0) = 0$, the third and fourth equalities hold by properties of $O_p(\cdot)$.

Recall that, by Lemma 1, $\widehat{\mathsf{d}}_g = O_p(n_{\min}^{-1/2})$; and, by Lemma 4, that $\widehat{\mathsf{d}}_g/h_n = O_p\left(\frac{1}{\sqrt{n_{\min}}h_n}\right)$.

Hence, by properties of $O_p(\cdot)$, $(\widehat{\mathsf{d}}_g/h_n)^2 = O_p\left(\frac{1}{n_{\min}h_n^2}\right)$. Since $(1/2) \cdot K''(0)$ is some constant, from Equation (23), we have:

$$\mathrm{K}\left(\widehat{\mathsf{d}}_g/h_n\right) - \mathrm{K}(0) = O_p\left(\frac{1}{n_{\min}h_n^2}\right) \tag{24}$$

Next, $|\mathcal{G}| < \infty$ and $|\mathcal{G}^\star| \neq 0$ by Assumption 5. For each $g \in \mathcal{G}$, $\widehat{p}_g = O_p(1)$ and $\bar{Y}_{g,t_1^*} = O_p(1)$ by the WLLN. Then, this implies:

$$
\begin{aligned}
\sum_{g \in \mathcal{G}^\star} \widehat{p}_g \cdot \bar{Y}_{g,t_1^*} \cdot \left(\mathrm{K}(\widehat{\mathsf{d}}_g/h_n) - \mathrm{K}(0)\right) &= \sum_{g \in \mathcal{G}^\star} O_p(1) \cdot O_p(1) \cdot O_p\left(\frac{1}{n_{\min}h_n^2}\right) \\
&= \sum_{g \in \mathcal{G}^\star} O_p\left(\frac{1}{n_{\min}h_n^2}\right) \\
&= O_p\left(\frac{1}{n_{\min}h_n^2}\right)
\end{aligned}
\tag{25}
$$

where the first equality holds by Equation (24); and the second and third equalities hold by properties of $O_p(\cdot)$. Lastly, by multiplication of Equation (25) by $\sqrt{n_{min}}$, we have:

$$\sqrt{n_{min}} \cdot \sum_{g \in \mathcal{G}^\star} \widehat{p}_g \cdot \bar{Y}_{g,t_1^*} \cdot \left(\mathrm{K}(\widehat{d}_g/h_n) - \mathrm{K}(0)\right) = O_p\left(\frac{1}{\sqrt{n_{\min}}h_n^2}\right)$$

by properties of $O_p(\cdot)$. Given Assumption 7 which states $\sqrt{n_{\min}}h_n^2 \to \infty$, it holds that

$$\sqrt{n_{min}} \sum_{g \in \mathcal{G}^\star} \widehat{p}_g \cdot \bar{Y}_{g,t_1^*} \cdot \left(K(\widehat{d}_g/h_n) - K(0)\right) = o_p(1) \tag{26}$$

as desired.

For term II_n, we consider only groups $g \notin \mathcal{G}^\star$. Lemma 2 states that for each group $g \notin \mathcal{G}^\star$, $\mathrm{K}(\widehat{\mathsf{d}}_g/h_n) = 0$ with probability tending to one. Since the number of groups is finite, we have that the sum

$$\sum_{g \notin \mathcal{G}^\star} \widehat{p}_g \cdot \bar{Y}_{g,t_1^*} \cdot \mathrm{K}(\widehat{\mathsf{d}}_g/h_n) = 0$$

with probability tending to one. This implies:

$$\sqrt{n_{\min}} \sum_{g \notin \mathcal{G}^\star} \widehat{p}_g \cdot \bar{Y}_{g,t_1^*} \cdot \mathrm{K}(\widehat{\mathsf{d}}_g/h_n) = o_p(1) \tag{27}$$

as desired.

Finally, the result follows from Equation (26) and Equation (27):

$$\sqrt{n_{\min}} \sum_{g \in \mathcal{G}} \widehat{p}_g \cdot \bar{Y}_{g,t_1^*} \cdot \left(\mathrm{K}(\widehat{\mathsf{d}}_g/h_n) - \mathrm{K}(0)\mathbf{1}\{g \in \mathcal{G}^\star\}\right) = o_p(1).$$

$\square$

Lemma 7 (Asymptotic equivalence to oracle estimator). Under Assumptions 1 to 9,

$$\sqrt{n_{\min}}\left(\widehat{\mathrm{ATT}}_{t_1^*} - \mathring{\mathrm{ATT}}_{t_1^*}\right) \xrightarrow{p} 0.$$

Proof of Lemma 7. Notice that

$$\sqrt{n_{\min}}\left(\widehat{\mathrm{ATT}}_{t_1^*}(h_n) - \mathring{\mathrm{ATT}}_{t_1^*}\right)$$

$$= -\sqrt{n_{\min}} \sum_{g\in\mathcal{G}} \frac{\widehat{p}_g \cdot \mathrm{K}\left(\widehat{\mathsf{d}}_g/h_n\right)}{\sum_{g'\in\mathcal{G}} \widehat{p}_{g'} \cdot \mathrm{K}\left(\widehat{\mathsf{d}}_{g'}/h_n\right)} \cdot \bar{Y}_{g,t_1^*} + \sqrt{n_{\min}} \sum_{g\in\mathcal{G}^*} \frac{\widehat{p}_g}{\sum_{g'\in\mathcal{G}^*} \widehat{p}_{g'}} \cdot \bar{Y}_{g,t_1^*}$$

$$= -\underbrace{\sqrt{n_{\min}} \left(\sum_{g\in\mathcal{G}} \frac{\widehat{p}_g \cdot \left(\mathrm{K}\left(\widehat{\mathsf{d}}_g/h_n\right) - \mathrm{K}(0)\mathbf{1}\{g\in\mathcal{G}^*\}\right)}{\sum_{g'\in\mathcal{G}} \widehat{p}_{g'} \cdot \mathrm{K}\left(\widehat{\mathsf{d}}_{g'}/h_n\right)} \cdot \bar{Y}_{g,t_1^*} \right)}_{\mathrm{I}}$$

$$+ \underbrace{\sqrt{n_{\min}} \left(\sum_{g\in\mathcal{G}} \frac{\widehat{p}_g \cdot \mathrm{K}(0)\mathbf{1}\{g\in\mathcal{G}^*\}}{\sum_{g'\in\mathcal{G}} \widehat{p}_{g'} \cdot \mathrm{K}(0)\mathbf{1}\{g'\in\mathcal{G}^*\}} \cdot \bar{Y}_{g,t_1^*} - \sum_{g\in\mathcal{G}} \frac{\widehat{p}_g \cdot \mathrm{K}(0)\mathbf{1}\{g\in\mathcal{G}^*\}}{\sum_{g'\in\mathcal{G}} \widehat{p}_{g'} \cdot \mathrm{K}\left(\widehat{\mathsf{d}}_{g'}/h_n\right)} \cdot \bar{Y}_{g,t_1^*} \right)}_{\mathrm{II}}$$

where the first equality holds by the definitions of $\widehat{\mathrm{ATT}}_{t_1^*}(h_n)$ and $\mathring{\mathrm{ATT}}_{t_1^*}$; and the second equality holds by adding and subtracting $\sqrt{n_{\min}} \sum_{g\in\mathcal{G}} \frac{\widehat{p}_g \cdot \mathrm{K}(0)\mathbf{1}\{g\in\mathcal{G}^*\}}{\sum_{g'\in\mathcal{G}} \widehat{p}_{g'} \cdot \mathrm{K}\left(\widehat{\mathsf{d}}_{g'}/h_n\right)} \cdot \bar{Y}_{g,t_1^*}$. Term I converges to 0 by Lemma 6 and the continuous mapping theorem. For Term II, notice that

$$\mathrm{II} = \sqrt{n_{\min}} \sum_{g\in\mathcal{G}} \frac{\widehat{p}_g \cdot \mathrm{K}(0) \cdot \mathbf{1}\{g\in\mathcal{G}^*\}}{\sum_{g'\in\mathcal{G}} \widehat{p}_{g'} \cdot \mathrm{K}(0)\mathbf{1}\{g'\in\mathcal{G}^*\}} \cdot \bar{Y}_{g,t_1^*} \cdot \left(1 - \frac{\sum_{g'\in\mathcal{G}} \widehat{p}_{g'} \cdot \mathrm{K}(0)\mathbf{1}\{g'\in\mathcal{G}^*\}}{\sum_{g'\in\mathcal{G}} \widehat{p}_{g'} \cdot \mathrm{K}\left(\widehat{\mathsf{d}}_{g'}/h_n\right)}\right)$$

$$= \sqrt{n_{\min}} \, \mathring{\mathrm{ATT}}_{t_1^*} \cdot \left(\frac{\sum_{g\in\mathcal{G}} \widehat{p}_g \cdot \left(\mathrm{K}\left(\widehat{\mathsf{d}}_g/h_n\right) - \mathrm{K}(0)\mathbf{1}\{g\in\mathcal{G}^*\}\right)}{\sum_{g'\in\mathcal{G}} \widehat{p}_{g'} \cdot \mathrm{K}\left(\widehat{\mathsf{d}}_{g'}/h_n\right)}\right) \xrightarrow{p} 0$$

where the first equality holds by rearranging terms; the second equality holds by the definition of $\mathring{\mathrm{ATT}}_{t_1^*}$; and the convergence to 0 holds by Lemma 6, the continuous mapping theorem, and the fact that the denominator converges to a positive constant, as shown in the proof of Proposition 7. $\square$

Lemma 8 (Asymptotic normality of oracle estimator). Under Assumptions 1 to 9,

$$\sqrt{n}\left(\mathring{\mathrm{ATT}}_{t_1^*} - \mathrm{ATT}_{t_1^*}\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(W_i) + o_p(1)$$

$$\xrightarrow{d} \mathcal{N}(0, V)$$

where

$$\psi(W) = \frac{\mathbf{1}\{G=1\}}{p_1}\left(Y_{t_1^*} - \mathbb{E}[Y_{t_1^*} \mid G=1]\right) - \frac{\mathbf{1}\{G \in \mathcal{G}^*\}}{p_{\mathcal{G}^*}}\left(Y_{t_1^*} - \mathbb{E}[Y_{t_1^*} \mid G \in \mathcal{G}^*]\right)$$

$$\text{and } V = \mathbb{E}\left[\psi(W)^2\right] = \frac{\text{Var}(Y_{t_1^*} \mid G=1)}{p_1} + \frac{\text{Var}(Y_{t_1^*} \mid G \in \mathcal{G}^*)}{p_{\mathcal{G}^*}}$$

Proof. Recall that

$$\overset{\circ}{\text{ATT}}_{t_1^*} = \frac{1}{n_1}\sum_{i=1}^n \mathbf{1}\{G_i = 1\}Y_{it_1^*} - \frac{1}{n_{\mathcal{G}^*}}\sum_{i=1}^n \mathbf{1}\{G_i \in \mathcal{G}^*\}Y_{it_1^*}$$

where $n_{\mathcal{G}^*}$ is the number of units in any group in $\mathcal{G}^*$, and this implies that

$$\sqrt{n}\left(\overset{\circ}{\text{ATT}}_{t_1^*} - \text{ATT}_{t_1^*}\right)$$

$$= \frac{n}{n_1} \cdot \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{1}\{G_i = 1\}\left(Y_{it_1^*} - \mathbb{E}[Y_{t_1^*} \mid G=1]\right)$$

$$\quad - \frac{n}{n_{\mathcal{G}^*}} \cdot \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{1}\{G_i \in \mathcal{G}^*\}\left(Y_{it_1^*} - \mathbb{E}[Y_{t_1^*} \mid G \in \mathcal{G}^*]\right)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\frac{\mathbf{1}\{G_i = 1\}}{p_1}\left(Y_{it_1^*} - \mathbb{E}[Y_{t_1^*} \mid G=1]\right) - \frac{\mathbf{1}\{G_i \in \mathcal{G}^*\}}{p_{\mathcal{G}^*}}\left(Y_{it_1^*} - \mathbb{E}[Y_{t_1^*} \mid G \in \mathcal{G}^*]\right)\right) + o_p(1)$$

where the first equality holds by the definition of $\text{ATT}_{t_1^*}$ and by multiplying and dividing by $\sqrt{n}$, and the second equality holds by the law of large numbers, continuous mapping theorem, and combining terms. The result then follows by applying the central limit theorem. $\qquad\square$

Proof of Theorem 2. The result follows immediately from Lemma 7 and Lemma 8. $\qquad\square$