

Post-Matching Two-Way Fixed Effects Estimation*

Yihong Liu[†] and Gonzalo Vazquez-Bare[‡]

February 13, 2026

Abstract

When estimating treatment effects with two-way fixed effects (2WFE) models, researchers often use matching as a pre-processing step when the parallel trends assumption is thought to hold conditionally on covariates. Specifically, in a first step, each treated unit is matched to one or more untreated units based on observed time-invariant covariates. In the second step, treatment effects are estimated with a 2WFE regression in the matched sample, reweighting the untreated units by the number of times they are matched. We formally analyze this common practice and highlight two problems. First, when different treatment cohorts enter treatment in different time periods, the post-matching 2WFE estimator that pools all treated cohorts has an asymptotic bias, even when the treatment effect is constant across units and over time. Second, failing to account for the variability introduced by the matching procedure yields invalid standard error estimators, which can be biased upwards or downwards depending on the data generating process. We propose simple post-matching difference-in-differences estimators that compare each treated cohort to the never-treated separately, instead of pooling all treated cohorts. We provide conditions under which these estimators are consistent for well-defined causal parameters, and derive valid standard errors that account for the matching step. We illustrate our results with simulations and with an empirical application.

Keywords: two-way fixed effects, difference-in-differences, matching

JEL Classification: C14, C21, C23

*We thank Max Farrell and Doug Steigerwald for valuable comments and suggestions.

[†]Department of Economics, UC Santa Barbara.

[‡]Department of Economics, UC Santa Barbara.

1 Introduction

When estimating treatment effects with two-ways fixed effects (2WFE) regressions, researchers often use matching as a pre-processing step. In such settings, each treated unit is matched to an untreated unit using pre-treatment observed covariates. In a second step, a 2WFE specification is estimated in the matched sample, reweighting the untreated units by the number of times they are used as a match for the treated units. This approach is typically justified under a conditional parallel trends assumption, by which units that are similar in their observed covariates are more likely to have followed the same outcome trends in the absence of treatment. While several alternative methods are available for estimating treatment effects under a conditional parallel trends assumption, matching-based methods have the advantage of being intuitively appealing, easy to implement, computationally stable and fully nonparametric, not requiring the functional form assumptions often invoked in inverse-probability weighting or regression adjustment methods. This combination of matching and 2WFE is common in empirical work: Table A1 lists 22 papers published between 2013 and 2025 in top general interest and field journals in Economics and Political Science that use some version of this approach.

In this paper, we formally analyze this commonly employed strategy and highlight two problems that generally invalidate its findings. First, when different treatment cohorts enter treatment in different time periods, the post-matching 2WFE estimator that pools all treated cohorts converges to a linear combination of average treatment effects on the treated (ATTs) plus an asymptotic bias, which we refer to as “pooling bias”, even when treatment effects are constant across units and over time. To explain this result, we show that the post-matching 2WFE estimator can be decomposed into a linear combination of 2x2 (that is, two-group-two-period) reweighted difference-in-differences (DiD) estimators. These 2x2 estimators compare each treated cohort to the never-treated cohort, and each treatment cohort to other treatment cohorts, with a specific set of weights. Because each treated unit is matched to a never-treated unit in the first step, comparisons between treated and reweighted never-treated units consistently estimate ATTs for different cohorts at different time periods. When comparing units across different treatment cohorts, however, units are not matched to each other, and thus the group used as comparison is not reweighted. As a result, the two groups will generally exhibit different covariate distributions. When the parallel trends assumption holds conditionally, these unweighted comparisons between different treated cohorts are invalid, even when the comparison group consists of not-yet-treated units.

To address this issue, we propose post-matching difference-in-differences (DiD) estimators that compare each treated cohort to the never-treated units separately, instead of pooling all treated cohorts. We show that under the conditional parallel trends assumption and mild regularity conditions, these pairwise estimators are consistent for well-defined averages of ATTs over time.

The second problem is that matching introduces variability that needs to be accounted for when conducting inference, and failing to account for this additional variability results in invalid standard

errors. We characterize the probability limit of the naive cluster-robust variance estimator for the post-matching DiD estimator and show that failing to account for the matching step creates an asymptotic bias in the variance estimator, which can be positive or negative depending on the data generating process. We rely on recent advancements on measures of high-order Voronoi cells by [Chen and Han \(2024\)](#) to provide a closed-form representation of the correct limiting variance of our proposed estimators, a result previously unavailable in the literature, and propose consistent variance estimators. Based on these results, we also compare the true limiting variance to the semiparametric efficiency bound for a general dimension of the covariate vector. We show that the post-matching DiD estimators do not attain the bound when the number of matches is fixed, but approach it as the number of neighbors diverges under some conditions.

We illustrate our results with simulations and with a reanalysis of the National Supported Work (NSW) Demonstration. Using the experimental benchmark from the randomized evaluation following [LaLonde \(1986\)](#), we find that the post-matching 2WFE estimator yields estimates that are much closer to the experimental estimates than the unmatched 2WFE, and accounting for the matching step increases the standard errors.

Finally, while our main results focus on nearest-neighbor matching with replacement, we discuss how they extend to matching on discrete covariates, nearest-neighbor matching without replacement and propensity score matching.

Our paper contributes to a large and growing literature on DiD and 2WFE methods in staggered designs, that is, designs where treated units exhibit variation in their treatment timing (see [de Chaisemartin and D’Haultfoeuille, 2022](#); [Roth, Sant’Anna, Bilinski, and Poe, 2023](#), for surveys). In particular, studies have found that in staggered designs, 2WFE estimators generally recover linear combinations of ATTs with weights that are hard to interpret and possibly negative when the treatment effects are heterogeneous ([Athey and Imbens, 2022](#); [Goodman-Bacon, 2021](#); [De Chaisemartin and d’Haultfoeuille, 2020](#); [Imai and Kim, 2021](#); [Sun and Abraham, 2021](#); [Borusyak, Jaravel, and Spiess, 2024](#)). These studies focus on unmatched 2WFE estimators, mainly under the unconditional version of the parallel trends assumption. We complement this literature by showing that these issues remain for post-matching 2WFE estimators when the parallel trends assumption holds conditionally. Furthermore, we show that pooling different treatment cohorts in the matching step creates an additional bias. Specifically, our [Theorem 1](#) decomposes the probability limit of the post-matching 2WFE estimator into a linear combination of cohort-specific ATTs and three types of bias: one due to the heterogeneity of treatment effects over time, one due to the heterogeneity of treatment effects across cohorts and a bias due to differences in covariate distributions across different treated cohorts. Because the third bias, which is specific to our setting, is caused by differences in covariate distributions and not by treatment effect heterogeneity, this bias remains even when the treatment effect is constant.

Another closely related strand of the literature is the one analyzing DiD and 2WFE under a conditional parallel trends assumption ([Abadie, 2005](#); [Sant’Anna and Zhao, 2020](#); [Callaway and](#)

Sant’ Anna, 2021). These studies focus on inverse-probability weighting and doubly-robust estimators. We build on their framework to analyze alternative estimators that control for covariates employing fully non-parametric nearest-neighbor matching methods.

Our paper also contributes to the literature on cross-sectional matching methods (Heckman, Ichimura, and Todd, 1998; Abadie and Imbens, 2006, 2008, 2011, 2012, 2016; Abadie and Spiess, 2022) by generalizing the results to 2WFE estimators with panel data. Specifically, our Propositions 1 and 2 show that while the post-matching 2WFE that pools all treated cohorts is a complicated linear combination of 2x2 DiD estimators, the estimator that compares one treated cohort to the never-treated can be recast as a cross-sectional matching estimator using a simple transformation of the outcome of interest, and thus existing matching methods can be used to characterize the probability limit and asymptotic distribution of this estimator. In addition, our closed-form characterization of the asymptotic variance based on the results by Chen and Han (2024) allows us to compare it to the semiparametric efficiency bound for a general dimension of the covariate vector, a result previously available only for the scalar case. In line with existing results for the average treatment effect (ATE) in cross-sectional settings (Abadie and Imbens, 2006; Lin, Ding, and Han, 2023), we find that the asymptotic variance is larger than the semiparametric efficiency bound, but approaches it under some conditions as the number of neighbors diverges.

Finally, among the papers at the intersection of matching and panel data methods, Heckman, Ichimura, and Todd (1997) propose a DiD-matching estimator in a two-period setting and derive its limiting distribution when matching is conducted using kernel and local linear regression methods. We complement their findings by considering nearest-neighbor matching with multiple periods and a staggered design, a setting that is very common in empirical work. Imai, Kim, and Wang (2023) propose matching methods for panel data, matching units based on their treatment history. They also show that their estimators can be written as weighted 2WFE regressions, but their inference is conducted conditionally on the matching weights and does not account for the variability introduced by the matching step.

The remainder of the paper is organized as follows. Section 2 describes the setup, notation and provides identification conditions for the parameters of interest. Section 3 characterizes the estimators of interest and their probability limits. Section 4 studies the asymptotic distribution of the estimators, provides a closed-form formula for their limiting variance, discusses efficiency and the inconsistency of the naive cluster-robust variance estimator that ignores the matching step. Section 5 contains simulation studies and Section 6 illustrates our results using data from the NSW program. Section 7 discusses extensions of our results to other types of matching, and Section 8 provides concluding remarks.

2 Setup and Identification

Consider panel data following n units $i = 1, \dots, n$ over T time periods $t = 1, 2, \dots, T$. Let t_i^* be a random variable with support $\mathcal{S} = \{s \in \{2, \dots, T, \infty\} : \mathbb{P}[t_i^* = s] > 0\}$ that indicates the period in which unit i receives the treatment for the first time. The values of t_i^* identify different treatment cohorts. We assume that there is at least one pure pre-treatment period (and possibly more) in which no unit is treated, $\mathbb{P}[t_i^* = 1] = 0$, and denote the never-treated units by $t_i^* = \infty$. We denote the population proportion of each treatment cohort by $p_s = \mathbb{P}[t_i^* = s]$. We assume that the treatment is an absorbing state, so that once unit i is treated, it remains treated through period T . Let $D_{it} = \mathbb{1}(t \geq t_i^*)$ be the treatment indicator in each period.

For $t' > t$, let $\mathbf{d}_{t:t'} = (d_t, d_{t+1}, \dots, d_{t'})$ be a vector of treatments between periods t and t' , where $d_t \in \{0, 1\}$ and $d_{t+1} \geq d_t$ for all t , and let $\mathbf{0}_{t:t'}$ and $\mathbf{1}_{t:t'}$ be vectors with all their entries being equal to zero or one, respectively. We use the shorthand notation $\mathbf{d}_{1:t} = \mathbf{d}_t$. Without further assumptions, the potential outcome for each unit in each time period may be a function of the whole treatment path, $Y_{it}(\mathbf{d}_T)$. We introduce the following assumption, standard in the DiD literature, which ensures that potential outcomes do not depend on future treatments.

Assumption 1 (No treatment anticipation) *For all t , $Y_{it}(\mathbf{d}_T) = Y_{it}(\mathbf{d}_t)$.*

We use the shorthand notation $Y_{it}(\mathbf{0}_t) = Y_{it}(0)$ to denote the potential outcome of a unit not receiving treatment up to period t and, for any $t \geq t_i^*$, we write $Y_{it}(t_i^*) = Y_{it}(\mathbf{0}_{t_i^*-1}, \mathbf{1}_{t_i^*:t})$, which is the potential outcome in period t for a unit that enters treatment in period t_i^* . The unit-level treatment effect is defined as $\tau_{it} = Y_{it}(t_i^*) - Y_{it}(0)$. The main parameters of interest are the cohort-specific ATTs, $\mathbb{E}[\tau_{it}|t_i^* = s]$.

Let X_i be a time-invariant vector of covariates of dimension $q \geq 1$, measured at a pure pre-treatment period $t < \min\{t_i^*\}$. To justify the matching procedure, we assume that the parallel trends assumption holds conditional on covariates in the following way.

Assumption 2 (Conditional parallel trends) *For all t ,*

$$\mathbb{E}[Y_{it}(0) - Y_{it-1}(0)|t_i^*, X_i] = \mathbb{E}[Y_{it}(0) - Y_{it-1}(0)|X_i] \quad (a.s.).$$

This assumption states that for each value of X_i , all treatment cohorts experience the same outcome trend in the absence of treatment.

Finally, let $e_s(X_i) = \mathbb{P}[t_i^* = s|X_i]$ be the cohort-specific propensity score, that is, the probability of entering treatment in period s given covariates X_i . We impose the following conditions which guarantees a non-zero mass of never treated and common support.

Assumption 3 (Overlap) $\mathbb{P}[t_i^* = \infty] > 0$ and for all $s \in \mathcal{S}$, $e_s(X_i) < 1 - \eta$ (a.s.) for some $\eta > 0$.

The following identification result is standard in the literature, and we include it here for reference.

Lemma 1 *Under Assumptions 1, 2 and 3, for any $t' < s \leq t < s'$,*

$$\mathbb{E}[\tau_{it}|t_i^* = s] = \mathbb{E}[Y_{it} - Y_{it'}|t_i^* = s] - \mathbb{E}\{\mathbb{E}[Y_{it} - Y_{it'}|t_i^* = s', X_i]|t_i^* = s\}.$$

3 Post-matching 2WFE estimation

Consider a sample $(Y_{i1}, \dots, Y_{iT}, t_i^*, X_i)_{i=1}^n$. Let $N_s = \sum_i \mathbb{1}(t_i^* = s)$ be the number of units in treatment cohort $t_i^* = s < \infty$ and $N_0 = \sum_i \mathbb{1}(t_i^* = \infty)$ be the number of never-treated units in the sample, with corresponding sample proportions $\hat{p}_s = N_s/n$ and $\hat{p}_\infty = N_0/n$. We assume observations are drawn independently from the same distribution.

Assumption 4 (Random sampling) *$(Y_{i1}, \dots, Y_{iT}, t_i^*, X_i)_{i=1}^n$ is a sample of independent and identically distributed observations from a population $(Y_1, \dots, Y_T, t^*, X')$.*

We study the following two-step estimation strategy commonly used in empirical practice. In the first step, each treated unit is matched with replacement to $M \geq 1$ units from the pool of never-treated units based on their covariates X_i . In the second step, treatment effects are estimated using a 2WFE regression with weights defined by the matching procedure, as described in detail below.

To analyze the matching step, we rely on the setup from [Abadie and Imbens \(2006\)](#). As in their paper, we consider the case in which all the covariates are continuous, and discuss matching on discrete covariates in Section 7. Let $\|x\| = \sqrt{x'x}$. For each unit i , we define $j_m(i, s')$ as the index $j \in \{1, \dots, n\}$ of units in comparison cohort $s' \neq t_i^*$ such that:

$$\sum_{l:t_l^*=s'} \mathbb{1}\{\|X_l - X_i\| \leq \|X_j - X_i\|\} = m,$$

that is, the index of the m -th closest unit to i in covariate values among the comparison cohort s' . Let $\mathcal{J}_M(i, s') = \{j_1(i, s'), j_2(i, s'), \dots, j_M(i, s')\}$ be the set of indices for the M nearest neighbors in cohort s' for unit i . Finally, let:

$$K_M(i, s) = \sum_{l:t_l^*=s} \mathbb{1}\{i \in \mathcal{J}_M(l, t_i^*)\}$$

be the number of times unit i is used as a match for units in cohort s . We also let:

$$K_M(i) = \sum_{s \neq t_i^*} K_M(i, s)$$

the total number of times unit i is used as a match for units in any other cohort.

Once the matching is conducted, treatment effects are estimated using a 2WFE regression on the matched sample that includes the treated units and their nearest neighbors. We consider the following general weighted 2WFE estimator. Let w_i be (possibly random) unit-level weights, collected in a vector $w = (w_1, \dots, w_n)$. For each time period $t = 1, \dots, T$, let $\lambda_t \in \{0, 1\}$ be a nonrandom indicator of whether the period is included in the analysis, and collect these indicators in a vector $\lambda = (\lambda_1, \dots, \lambda_T)$ where $\sum_t \lambda_t \geq 2$ so there are at least two periods. The weighted 2WFE estimator is defined as the estimator of τ from the 2WFE regression:¹

$$Y_{it} = \alpha_i + \delta_t + \tau D_{it} + u_{it} \quad (1)$$

using weights $w_i \lambda_t$, and can be written as:

$$\hat{\tau}(w, \lambda) = \frac{\sum_i \sum_t w_i \lambda_t Y_{it} (D_{it} - \bar{D}_i - \tilde{D}_t + \bar{D})}{\sum_i \sum_t w_i \lambda_t D_{it} (D_{it} - \bar{D}_i - \tilde{D}_t + \bar{D})} \quad (2)$$

where

$$\bar{D}_i = \frac{\sum_t \lambda_t D_{it}}{\sum_t \lambda_t}, \quad \tilde{D}_t = \frac{\sum_i w_i D_{it}}{\sum_i w_i}, \quad \bar{D} = \frac{\sum_i \sum_t w_i \lambda_t D_{it}}{\sum_i w_i \sum_t \lambda_t}.$$

Lemma B1 in the appendix provides a general representation of this estimator for an arbitrary set of weights.

In what follows, we focus on the empirical approach described above, which assigns a weight equal to one to all treated units, pooling all treatment cohorts, and reweight the never-treated units by the number of times they are used as a match, scaled by the number of neighbors M . We refer to this estimator as the full-sample matched 2WFE estimator. The following result characterizes this estimator.

Proposition 1 *Let*

$$w_i^{\text{pool}} = \sum_{s \neq \infty} \mathbb{1}(t_i^* = s) + \mathbb{1}(t_i^* = \infty) \frac{K_M(i)}{M}$$

and $w^{\text{pool}} = (w_1^{\text{pool}}, \dots, w_n^{\text{pool}})$. *With these weights, the weighted 2WFE estimator (2) can be written as:*

$$\hat{\tau}(w^{\text{pool}}, \lambda) = \frac{\hat{\tau}_{\text{num}}^{\text{pool}}(\lambda)}{\hat{\tau}_{\text{den}}^{\text{pool}}(\lambda)}$$

¹We assume that the weights (w, λ) are chosen in such a way that this estimator is well defined, which requires avoiding perfect collinearity between the regressor D_{it} and the unit and time effects. This rules out obvious cases where, for example, no included cohort experiences a treatment change between the included periods, or all included cohorts experience the same treatment change.

where $\Lambda_s = \sum_{t \geq s} \lambda_t / \sum_t \lambda_t$,

$$\begin{aligned} \hat{\tau}_{\text{num}}^{\text{pool}}(\lambda) &= \sum_{s < \infty} \hat{p}_s \left\{ \sum_{t \geq s} \sum_{t' < s} \lambda_t \lambda_{t'} \hat{\tau}_{\infty}^s(t, t') + \sum_{s' \neq s, \infty} \frac{\hat{p}_{s'}}{1 - \hat{p}_{\infty}} \sum_{t \geq s} \sum_{t' < s} \lambda_t \lambda_{t'} \hat{\tau}_{s'}^s(t, t') \right\} \\ \hat{\tau}_{\text{den}}^{\text{pool}}(\lambda) &= \sum_{s < \infty} \hat{p}_s \Lambda_s \left\{ (1 - \Lambda_s) + \frac{\sum_{s' < s} \hat{p}_{s'} (\Lambda_{s'} - \Lambda_s)}{1 - \hat{p}_{\infty}} \right\} + \sum_{s < s' < \infty} \frac{\sum_{s < \infty} \hat{p}_s \hat{p}_{s'}}{1 - \hat{p}_{\infty}} (1 - \Lambda_s) (\Lambda_s - \Lambda_{s'}) \\ &\quad \times \left(\sum_t \lambda_t \right)^2 \\ \hat{\tau}_{\infty}^s(t, t') &= \frac{1}{N_s} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)}{M} \right) (Y_{it} - Y_{it'}) \\ \hat{\tau}_{s'}^s(t, t') &= \frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = s) (Y_{it} - Y_{it'}) - \frac{1}{N_{s'}} \sum_i \mathbb{1}(t_i^* = s') (Y_{it} - Y_{it'}), \quad s' \neq \infty. \end{aligned}$$

Proposition 1 shows that the full-sample matched 2WFE estimator can be decomposed as a linear combination of all possible 2x2 estimators, comparing each cohort s to the never-treated cohort in different post- and pre-treatment periods, $\hat{\tau}_{\infty}^s(t, t')$, and comparing each treated cohort s to other eventually treated cohorts, $\hat{\tau}_{s'}^s(t, t')$ for $s' \neq \infty$. As pointed out in the literature on unmatched 2WFE estimators (De Chaisemartin and d'Haultfoeuille, 2020; Athey and Imbens, 2022; Goodman-Bacon, 2021; Imai and Kim, 2021; Sun and Abraham, 2021; Borusyak, Jaravel, and Spiess, 2024), the second set of comparisons is problematic because it involves some comparison groups that are already treated, often known as “forbidden comparisons”. Importantly, while the 2x2 estimators using the never-treated, $\hat{\tau}_{\infty}^s(t, t')$, use the matching weights, the 2x2 estimators comparing different eventually treated cohorts, $\hat{\tau}_{s'}^s(t, t')$ for $s' \neq \infty$, are unweighted DiD comparisons. These unweighted comparisons generally bias the estimator, as we show next. To characterize the probability limit of the estimator, we introduce the following regularity conditions.

Assumption 5 (Regularity conditions)

1. The conditional densities of $X_i | t_i^* = \infty$ and $X_i | t_i^* = s$ for all $s \in \mathcal{S} \setminus \{\infty\}$, $f_0(x)$ and $f_s(x)$ respectively, are continuous and have convex and compact support, and $f_0(x)$ is bounded away from zero.
2. For all s, t and any treatment path \mathbf{d}_t , $\mathbb{E}[Y_{it}(\mathbf{d}_t) | t_i^* = s, X_i = x]$ is Lipschitz-continuous.
3. For all s, t and any treatment path \mathbf{d}_t , $\mathbb{V}[Y_{it}(\mathbf{d}_t) | t_i^* = s, X_i = x]$ is Lipschitz-continuous and bounded away from zero.
4. For all s, t and any treatment path \mathbf{d}_t , $\mathbb{E}[Y_{it}(\mathbf{d}_t)^4 | t_i^* = s, X_i = x]$ is uniformly bounded.

Theorem 1 Under Assumptions 1-5,

$$\hat{\tau}(w^{\text{pool}}, \lambda) \rightarrow_{\mathbb{P}} \sum_{s < \infty} \phi_1(s, \lambda) \frac{\sum_{t \geq s} \lambda_t \mathbb{E}[\tau_{it} | t_i^* = s]}{\sum_{t \geq s} \lambda_t} + \sum_{s < \infty} \sum_{s < s' < \infty} \phi_2(s, s', \lambda) \frac{\sum_{s \leq t < s'} \lambda_t \mathbb{E}[\tau_{it} | t_i^* = s]}{\sum_{s \leq t < s'} \lambda_t} \\ + B_{\text{cohort}}^{\text{het}} - B_{\text{time}}^{\text{het}} + B_{\text{pool}}$$

where $B_{\text{cohort}}^{\text{het}}$, $B_{\text{time}}^{\text{het}}$ and B_{pool} are bias terms defined as:

$$B_{\text{cohort}}^{\text{het}} = \sum_{s < \infty} \sum_{s' \neq s, \infty} \eta_1(s, s', \lambda) \frac{\sum_{t \geq \max\{s, s'\}} \lambda_t (\mathbb{E}[\tau_{it} | t_i^* = s] - \mathbb{E}[\tau_{it} | t_i^* = s'])}{\sum_{t \geq \max\{s, s'\}} \lambda_t} \\ B_{\text{time}}^{\text{het}} = \sum_{s < \infty} \sum_{s' < s} \eta_2(s, s', \lambda) \frac{\sum_{t \geq s} \sum_{s' \leq t' < s} \lambda_t \lambda_{t'} \mathbb{E}[\tau_{it} - \tau_{it'} | t_i^* = s']}{\sum_{t \geq s} \sum_{s' \leq t' < s} \lambda_t \lambda_{t'}} \\ B_{\text{pool}} = \sum_{s < \infty} \sum_{s' \neq s, \infty} \eta_3(s, s', \lambda) \frac{\sum_{t \geq s} \sum_{t' < s} \lambda_t \lambda_{t'} B_{s'}^s(t, t')}{\sum_{t \geq s} \sum_{t' < s} \lambda_t \lambda_{t'}} \\ B_{s'}^s(t, t') = \mathbb{E}[Y_{it}(0) - Y_{it'}(0) | t_i^* = s] - \mathbb{E}[Y_{it}(0) - Y_{it'}(0) | t_i^* = s']$$

and where, letting $\Lambda_s = \sum_{t \geq s} \lambda_t / \sum_t \lambda_t$, the weights are:

$$\phi_1(s, \lambda) = \frac{p_s \Lambda_s [(1 - \Lambda_s) + \sum_{s' < s} p_{s'} (\Lambda_{s'} - \Lambda_s) / (1 - p_\infty)]}{\sum_{s < \infty} \left\{ p_s \Lambda_s \left[(1 - \Lambda_s) + \frac{\sum_{s' < s} p_{s'} (\Lambda_{s'} - \Lambda_s)}{1 - p_\infty} \right] + \sum_{s < s' < \infty} \frac{p_s p_{s'}}{1 - p_\infty} (1 - \Lambda_s) (\Lambda_s - \Lambda_{s'}) \right\}} \\ \phi_2(s, s', \lambda) = \frac{p_s p_{s'} (1 - \Lambda_s) (\Lambda_s - \Lambda_{s'}) / (1 - p_\infty)}{\sum_{s < \infty} \left\{ p_s \Lambda_s \left[(1 - \Lambda_s) + \frac{\sum_{s' < s} p_{s'} (\Lambda_{s'} - \Lambda_s)}{1 - p_\infty} \right] + \sum_{s < s' < \infty} \frac{p_s p_{s'}}{1 - p_\infty} (1 - \Lambda_s) (\Lambda_s - \Lambda_{s'}) \right\}} \\ \eta_1(s, s', \lambda) = \frac{p_s p_{s'} \Lambda_{\max\{s, s'\}} (1 - \Lambda_{\min\{s, s'\}}) / (1 - p_\infty)}{\sum_{s < \infty} \left\{ p_s \Lambda_s \left[(1 - \Lambda_s) + \frac{\sum_{s' < s} p_{s'} (\Lambda_{s'} - \Lambda_s)}{1 - p_\infty} \right] + \sum_{s < s' < \infty} \frac{p_s p_{s'}}{1 - p_\infty} (1 - \Lambda_s) (\Lambda_s - \Lambda_{s'}) \right\}} \\ \eta_2(s, s', \lambda) = \frac{p_s p_{s'} \Lambda_s (\Lambda_{s'} - \Lambda_s) / (1 - p_\infty)}{\sum_{s < \infty} \left\{ p_s \Lambda_s \left[(1 - \Lambda_s) + \frac{\sum_{s' < s} p_{s'} (\Lambda_{s'} - \Lambda_s)}{1 - p_\infty} \right] + \sum_{s < s' < \infty} \frac{p_s p_{s'}}{1 - p_\infty} (1 - \Lambda_s) (\Lambda_s - \Lambda_{s'}) \right\}} \\ \eta_3(s, s', \lambda) = \frac{p_s p_{s'} \Lambda_s (1 - \Lambda_s) / (1 - p_\infty)}{\sum_{s < \infty} \left\{ p_s \Lambda_s \left[(1 - \Lambda_s) + \frac{\sum_{s' < s} p_{s'} (\Lambda_{s'} - \Lambda_s)}{1 - p_\infty} \right] + \sum_{s < s' < \infty} \frac{p_s p_{s'}}{1 - p_\infty} (1 - \Lambda_s) (\Lambda_s - \Lambda_{s'}) \right\}}.$$

Theorem 1 shows that in a staggered design, the post-matching 2WFE estimator does not generally recover a causally interpretable parameter. The probability limit of this estimator can be decomposed into four components. The first component, in the first line of the expression for $\hat{\tau}(w^{\text{pool}}, \lambda)$ above, is a weighted average of ATTs across cohorts and over the included time periods, where the weights are non-negative and sum up to one, $\sum_{s < \infty} \phi_1(s, \lambda) + \sum_{s < \infty} \sum_{s < s' < \infty} \phi_2(s, s', \lambda) = 1$. Although this is a proper weighted average, the weights may be highly unequal across ATTs, and this term is generally hard to interpret.

In addition to this weighted average of ATTs, the full-sample matched 2WFE estimator contains three bias terms. The first bias, $B_{\text{cohort}}^{\text{het}}$, involves differences in ATTs in a given time period between different cohorts, $\mathbb{E}[\tau_{it}|t_i^* = s] - \mathbb{E}[\tau_{it}|t_i^* = s']$. These terms are non-zero whenever treatment effects are heterogeneous across cohorts. Notice that because this term consists of differences in ATTs, some ATTs enter negatively. This is an illustration of the negative weights issue pointed out in the literature on unmatched DiD with staggered adoption.

The second bias, $B_{\text{time}}^{\text{het}}$, involves the change in the ATTs for a given cohort over time, $\mathbb{E}[\tau_{it} - \tau_{it'}|t_i^* = s']$. This term appears because the 2WFE estimator uses already-treated units as comparison units, resulting in some ATTs entering with a negative sign in the probability limit of the estimator. Notice that the bias terms $B_{\text{cohort}}^{\text{het}}$ and $B_{\text{time}}^{\text{het}}$ have been discussed in existing literature analyzing unmatched DiD estimators. Theorem 1 shows that these issues persist when using matched 2WFE estimators.

The third bias, B_{pool} , on the other hand, is specific to the post-matching setting when the matching is conducted as explained above. This ‘‘pooling bias’’ involves differences in untreated outcome trends across eventually-treated cohorts, $B_{s'}^s(t, t')$. To understand the intuition behind this bias, notice that these terms can be rewritten as:

$$B_{s'}^s(t, t') = \mathbb{E} \left\{ \mathbb{E} [Y_{it}(0) - Y_{it'}(0) | t_i^* = s, X_i] \left(\frac{f_s(X_i) - f_{s'}(X_i)}{f(X_i)} \right) \right\}$$

where $f(x)$ is the unconditional density of X_i and $f_s(x)$ and $f_{s'}(x)$ are the conditional densities of X_i for cohorts s and s' , respectively. Thus, this bias is driven by the difference in covariate distributions across eventually-treated cohorts. As shown in Proposition 1, the full-sample matched 2WFE estimator involves comparisons between different treated cohorts. Although each treated unit is matched in the first step to a never-treated unit, ensuring that the distribution of covariates for the treated is identical to the reweighted distribution of the never-treated, the different treated cohorts are not matched to each other, and thus their covariate distributions will differ in general. Because the parallel trends assumption only holds conditionally on X_i , different treated cohorts may have different unconditional trends in their untreated outcomes, which creates this bias. In other words, when the parallel trends assumption holds conditionally, any unmatched comparison between eventually treated cohorts is invalid, even those involving not-yet-treated units as comparisons.

When the treatment effect is constant both across units and over time $\tau_{it} = \tau$, $\mathbb{E}[\tau_{it}|t_i^* = s] - \mathbb{E}[\tau_{it}|t_i^* = s'] = \mathbb{E}[\tau_{it} - \tau_{it'}|t_i^* = s'] = 0$, so the biases due to treatment effect heterogeneity equal zero, $B_{\text{time}}^{\text{het}} = B_{\text{cohort}}^{\text{het}} = 0$. However, the bias due to the difference in covariate distributions, B_{pool} , remains regardless of whether treatment effects are homogeneous or heterogeneous. This bias equals zero when either the parallel trends assumption holds unconditionally, $\mathbb{E} [Y_{it}(0) - Y_{it'}(0) | t_i^* = s, X_i] = \mathbb{E} [Y_{it}(0) - Y_{it'}(0) | t_i^* = s]$ for all s , so matching is unnecessary, or when the distribution of covariates across cohorts are the same, $f_s(x) = f_{s'}(x)$, which is generally

not true even when the treatment effect is constant.

Thus, the commonly used full-sample matched 2WFE estimator requires stringent homogeneity conditions on treatment effects and covariate distributions to recover causally interpretable parameters when there is variation in treatment timing. To avoid this problem, a staggered design can be split into multiple pairwise designs, as suggested by [Callaway and Sant'Anna \(2021\)](#); [Sun and Abraham \(2021\)](#); [Steigerwald, Vazquez-Bare, and Maier \(2021\)](#), among others, in unmatched panel designs. We now consider an estimator that assigns weight equal to one to units in a specific cohort $t_i^* = s$, weights equal to $K_M(i, s)/M$ to never-treated units, and weight equal to zero to all other cohorts. We refer to these estimators as pairwise matched DiD estimators. The following proposition characterizes this estimator.

Proposition 2 *For a given $s \in \mathcal{S} \setminus \{\infty\}$, let:*

$$w_i^s = \mathbb{1}(t_i^* = s) + \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)}{M},$$

with $w^s = (w_1^s, \dots, w_n^s)$. With these weights, the weighted 2WFE estimator (2) can be written as:

$$\hat{\tau}(w^s, \lambda) = \frac{1}{N_s} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)}{M} \right) (\bar{Y}_i^{\text{post},s}(\lambda) - \bar{Y}_i^{\text{pre},s}(\lambda)) \quad (3)$$

where

$$\bar{Y}_i^{\text{post},s}(\lambda) = \frac{\sum_{t \geq s} \lambda_t Y_{it}}{\sum_{t \geq s} \lambda_t}, \quad \bar{Y}_i^{\text{pre},s}(\lambda) = \frac{\sum_{t < s} \lambda_t Y_{it}}{\sum_{t < s} \lambda_t}$$

are the average outcome for each unit in the included post- and pre-treatment periods, respectively.

Proposition 2 shows that the pairwise matched DiD estimator can be recast as cross-sectional matching estimator like the ones studied by [Abadie and Imbens \(2006\)](#), where the outcome is the difference in average post- and pre-treatment periods. As an illustration, for a given s , when $\lambda_t^* = \mathbb{1}(t \in \{\ell', \ell\})$ for a pair of periods such that $\ell' < s \leq \ell$,

$$\hat{\tau}(w^s, \lambda^*) = \frac{1}{N_s} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)}{M} \right) (Y_{i\ell} - Y_{i\ell'})$$

which is a 2x2 (i.e. two-group-two-period) matched DiD estimator.

Remark 1 (Normalized weights) *Because weighted OLS automatically normalizes the weights to sum to one, and since the sum of the number of times each untreated unit is used as a match, $\sum_i \mathbb{1}(t_i^* = \infty) K_M(i, s)$, equals the total number of matches, $N_s M$, it follows that Proposition 2 also holds for normalized weights:*

$$\tilde{w}_i^s = \frac{\mathbb{1}(t_i^* = s)}{N_s} + \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)}{N_s M}$$

which replicate those in [Abadie and Imbens \(2006\)](#).

The pairwise matched DiD estimator is consistent for an average of the ATTs for cohort s over the included post-treatment periods $t \geq s$, as we show below.

Theorem 2 *Under Assumptions 1-5,*

$$\hat{\tau}(w^s, \lambda) \rightarrow_{\mathbb{P}} \frac{\sum_{t \geq s} \lambda_t \mathbb{E}[\tau_{it} | t_i^* = s]}{\sum_{t \geq s} \lambda_t}.$$

As an illustration, if $\lambda_t = \mathbb{1}(t \in \{\ell', \ell\})$ with $\ell' < s \leq \ell$, we have that $\hat{\tau}(w^s, \lambda) \rightarrow_{\mathbb{P}} \mathbb{E}[\tau_{i\ell} | t_i^* = s]$ which is the ATT for cohort s in period ℓ . On the other hand, when all periods are included, so that $\lambda_t = 1$ for all t , $\hat{\tau}(w^s, \lambda) \rightarrow_{\mathbb{P}} \sum_{t \geq s} \mathbb{E}[\tau_{it} | t_i^* = s] / (T + 1 - s)$ which is a simple average of ATTs over periods $t \geq s$.

Remark 2 (Not-yet-treated) *This result, and all subsequent results for the pairwise matched DiD estimator, immediately generalizes to the case where units in cohort s are matched to units in some other eventually-treated cohort $s' \neq s$ instead of the never-treated, as long as cohort s' is untreated in all periods for which $\lambda_t = 1$. This corresponds to the case where the not-yet-treated are used as comparison units, as discussed in [Callaway and Sant'Anna \(2021\)](#) and [Sun and Abraham \(2021\)](#).*

These results demonstrate that the inconsistency of the full-sample matched 2WFE estimator can be avoided by comparing each treated cohort to the never treated, which consistently estimates an average of ATTs over time. When conducting inference based on the pairwise matched DiD estimators, the variance needs to account for the variability introduced by the matching step, and failing to account for this variability generally results in invalid standard error estimators. The next section studies the asymptotic distribution of the pairwise matched DiD estimators, proposes valid standard error estimators and characterizes the bias of the naive variance estimators that ignore the matching step.

4 Asymptotic Distribution and Inference

In this section, we study the asymptotic distribution and valid inference procedures for the pairwise estimators (3). In what follows, for a given $s \in \mathcal{S}$, let $\Delta \bar{Y}_i^s(\lambda) = \bar{Y}_i^{\text{post},s}(\lambda) - \bar{Y}_i^{\text{pre},s}(\lambda)$, $\mu_\lambda(s, x) = \mathbb{E}[\Delta \bar{Y}_i^s(\lambda) | t_i^* = s, X_i = x]$ and $\sigma_\lambda^2(s, x) = \mathbb{V}[\Delta \bar{Y}_i^s(\lambda) | t_i^* = s, X_i = x]$ where we use the notation $\mu_\lambda(0, x)$ and $\sigma_\lambda^2(0, x)$ when $s = \infty$. Also let:

$$\bar{\tau}_i^{\text{post},s}(\lambda) = \frac{\sum_{t \geq s} \lambda_t \tau_{it}}{\sum_{t \geq s} \lambda_t}$$

be the average unit-level treatment effect over the included post-treatment periods $t \geq s$ with $s < \infty$. The following result characterizes the asymptotic distribution of the pairwise matched DiD estimator.

Theorem 3 *Under Assumptions 1-5, for $s \in \mathcal{S} \setminus \{\infty\}$,*

$$\sqrt{n} \left(\hat{\tau}(w^s, \lambda) - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s] - B_\lambda(s, M) \right) \rightarrow_{\mathcal{D}} \mathcal{N}(0, V(s, \lambda))$$

where

$$\begin{aligned} V(s, \lambda) &= \frac{1}{p_s} \mathbb{E} \left[\left(\mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s, X_i] - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s] \right)^2 \middle| t_i^* = s \right] + \frac{1}{p_s} \mathbb{E}[\sigma_\lambda^2(s, X_i) | t_i^* = s] \\ &\quad + \frac{1}{p_s^2 M^2} \mathbb{E} \left[\left(M + \alpha(M, q) \frac{e_s(X_i)}{e_\infty(X_i)} \right) e_s(X_i) \sigma_\lambda^2(0, X_i) \right], \end{aligned}$$

$\alpha(M, q)$ is a distribution-free, nonrandom function defined in [Chen and Han \(2024\)](#) and

$$B_\lambda(s, M) = \frac{1}{N_s} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)}{M} \right) \mu_\lambda(0, X_i).$$

Theorem 3 shows that the pairwise matched DiD estimator is asymptotically normal after appropriate centering and scaling, which means that valid inference can be conducted using standard procedures. In general, the estimator needs to be bias-corrected, as initially pointed out by [Abadie and Imbens \(2006\)](#). We refer to this bias $B_\lambda(s, M)$ as the “matching bias”. The matching bias $B_\lambda(s, M)$ is a finite-sample bias due to the fact that, when covariates are continuous, it is not possible to find perfect matches for treated units in any particular sample. This type of finite sample bias is typical in nonparametric estimators. It can be shown that under the assumptions in the theorem, $B_\lambda(s, M) = O_{\mathbb{P}}(n^{-1/q})$ where q is the dimension of X_i . This implies that $B_\lambda(s, M) = o_{\mathbb{P}}(1)$ so the matching bias does not appear in the probability limit of the estimator, but it needs to be accounted for in the asymptotic distribution because it is multiplied by the convergence rate.

In some cases, the bias correction is not needed. For example, when matching on a single covariate, $\sqrt{n}B_\lambda(s, M) = o_{\mathbb{P}}(1)$. [Abadie and Imbens \(2006\)](#) also provide an alternative sampling scheme in which the bias is asymptotically negligible when the never-treated group is “sufficiently larger” than the treated group; see Section 7.2 for further discussion. Finally, when matching on discrete covariates, matches are exact, so the bias is zero, as discussed in Section 7.1.

For these reasons, the nature of the matching bias is very different from that of the pooling bias of the full-sample matched 2WFE estimator characterized in Theorem 1, B_{pool} . Specifically, the pooling bias is a large-sample bias that appears in the probability limit of the estimator, and does not decrease with the sample size.

Because the pairwise estimator is asymptotically normal, valid inference can be conducted

using standard methods after estimating the bias and the asymptotic variance that accounts for the matching step. Proposition 2 shows that the pairwise matched DiD estimator can be written as a cross-sectional estimator, and therefore the methods available in the literature for bias and variance estimation like the ones in Abadie and Imbens (2006, 2011) can be applied in our setting after transforming the data appropriately.

4.1 Efficiency

The analytic formula for the asymptotic variance we derive in Theorem 3 allows us to study the efficiency of the pairwise matched DiD estimator for a general dimension of the covariate vector. Under Assumptions 2, 3 and 4, Sant’Anna and Zhao (2020) derive the semiparametric efficiency bound for the ATT with panel data under the conditional parallel trends assumption. When comparing cohort s against the never-treated, the bound is given by:

$$V_{\text{SEB}}(\lambda) = \mathbb{E} \left[\frac{\sigma_\lambda^2(s, X_i) e_s(X_i)}{p_s^2} + \frac{(\mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s, X_i] - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s])^2 e_s(X_i)}{p_s^2} \right] + \mathbb{E} \left[\frac{e_s(X_i)^2 \sigma_\lambda^2(0, X_i)}{p_s^2 e_\infty(X_i)} \right]. \quad (4)$$

Then we have the following result.

Proposition 3

$$V(s, \lambda) - V_{\text{SEB}}(\lambda) = \frac{1}{M} \mathbb{E} \left[\frac{e_s(X_i) \sigma_\lambda^2(0, X_i)}{p_s^2} \right] + \left(\frac{\alpha(M, q)}{M^2} - 1 \right) \mathbb{E} \left[\frac{e_s(X_i)^2 \sigma_\lambda^2(0, X_i)}{e_\infty(X_i) p_s^2} \right].$$

Proposition 3 shows that the matched DiD estimator does not attain the semiparametric efficiency bound, and that the difference between the asymptotic variance and the bound depends on the number of nearest neighbors M and on the dimension of the covariates q through the function $\alpha(M, q)$. As an illustration, with a single covariate $q = 1$, $\alpha(M, 1) = M(2M + 1)/2$ (Chen and Han, 2024) and thus:

$$V(s, \lambda) - V_{\text{SEB}} = \frac{1}{M} \mathbb{E} \left[\frac{e_s(X_i) \sigma_\lambda^2(0, X_i)}{p_s^2} \right] + \frac{1}{2M} \mathbb{E} \left[\frac{e_s(X_i)^2 \sigma_\lambda^2(0, X_i)}{e_\infty(X_i) p_s^2} \right].$$

In this case it is clear that the difference between the variance and the bound vanishes as $M \rightarrow \infty$, in line with the results in Abadie and Imbens (2006) and Lin, Ding, and Han (2023) for the ATE with cross-sectional data under unconfoundedness. Because the closed form of $\alpha(M, q)$ has not been characterized for a general M and q , it is hard to extend this statement for the general case. However, the tabulations in Chen and Han (2024) show that for $M \in \{2, 3, \dots, 10\}$ and $d \in \{2, 3, \dots, 10\}$, $\alpha(M, q)/M^2 - 1 < 0.15$, and that $\alpha(M, q)/M^2 - 1$ is positive and decreasing in M for each of these values of M and q .

4.2 Inference Ignoring the Matching Step

Because the pairwise matched DiD estimator is constructed in two steps, inference needs to account for the first-step estimation in the second-step asymptotic variance, as done in Theorem 3. In practice, however, researchers often rely on the cluster-robust variance estimator from the second-stage 2WFE regression, failing to account for the variability introduced by the matching step, which results in inconsistent variance estimators. With cross-sectional data under unconfoundedness, Abadie and Imbens (2016) compare the asymptotic variance of the oracle propensity-score matching ATT estimator with the variance of the ATT estimator using the estimated propensity score, and show that the latter can be smaller or larger depending on the data generating process. On the other hand, Abadie and Spiess (2022) consider post-matching regression estimators using cross-sectional data and matching without replacement and show that standard errors that fail to account for the matching step are inconsistent when the regression of interest is misspecified. In this section, we derive the probability limit of the cluster-robust 2WFE variance estimator that ignores the matching step, which we refer to as the naive cluster-robust variance estimator.

By Proposition 2, the cluster-robust 2WFE estimator of the variance of $\hat{\tau}(w^s, \lambda)$ can be written as:

$$\hat{V}_{\text{ncr}} = \frac{1}{N_s^2} \sum_i \mathbb{1}(t_i^* = s) \hat{\varepsilon}_i^2 + \frac{1}{N_s^2} \sum_i \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)^2}{M^2} \hat{\varepsilon}_i^2$$

where:

$$\begin{aligned} \hat{\varepsilon}_i = & \Delta \bar{Y}_i^s(\lambda) - \left(\frac{1}{N_s} \sum_j \mathbb{1}(t_j^* = s) \Delta \bar{Y}_j^s(\lambda) \right) \mathbb{1}(t_i^* = s) \\ & - \left(\frac{1}{N_s} \sum_j \mathbb{1}(t_j^* = \infty) \frac{K_M(j, s)}{M} \Delta \bar{Y}_j^s(\lambda) \right) \mathbb{1}(t_i^* = \infty) \end{aligned}$$

is the regression residual and the dependence on λ is left implicit to reduce notation. The following result characterizes the probability limit of this estimator.

Proposition 4 *Under Assumptions 1-5,*

$$\begin{aligned} n \hat{V}_{\text{ncr}} & \rightarrow_{\mathbb{P}} V(s, \lambda) \\ & + \frac{1}{p_s} \mathbb{V}[\mathbb{E}[\Delta \bar{Y}_i^s(0, \lambda) | t_i^* = s, X_i] | t_i^* = s] \\ & + \frac{2}{p_s} \text{Cov}(\mathbb{E}[\bar{\tau}_i^{\text{post}, s}(\lambda) | t_i^* = s, X_i], \mathbb{E}[\Delta \bar{Y}_i^s(0, \lambda) | t_i^* = s, X_i] | t_i^* = s) \\ & + \frac{1}{p_s^2 M^2} \mathbb{E} \left[\left(M + \alpha(M, q) \frac{e_s(X_i)}{e_\infty(X_i)} \right) e_s(X_i) (\mathbb{E}[\Delta \bar{Y}_i^s(0, \lambda) | t_i^* = s, X_i] - \mathbb{E}[\Delta \bar{Y}_i^s(0, \lambda) | t_i^* = s])^2 \right]. \end{aligned}$$

where

$$\Delta\bar{Y}_i^s(0, \lambda) = \frac{\sum_{t \geq s} \lambda_t Y_{it}(0)}{\sum_{t \geq s} \lambda_t} - \frac{\sum_{t < s} \lambda_t Y_{it}(0)}{\sum_{t < s} \lambda_t}.$$

Proposition 4 shows that if $\mathbb{E}[\Delta\bar{Y}_i^s(0, \lambda)|t_i^* = s, X_i] = \mathbb{E}[\Delta\bar{Y}_i^s(0, \lambda)|t_i^* = s]$, so that covariates do not affect the average counterfactual trend, the naive cluster-robust variance estimator that ignores the matching step is consistent for the true variance. This case, however, only occurs when the matching step is unnecessary, because counterfactual trends are parallel unconditionally. More generally, the naive cluster-robust variance estimator is inconsistent for the true asymptotic variance, with an asymptotic bias given by:

$$\begin{aligned} & \frac{1}{p_s} \mathbb{V}[\mathbb{E}[\Delta\bar{Y}_i^s(0, \lambda)|t_i^* = s, X_i]|t_i^* = s] \\ & + \frac{2}{p_s} \text{Cov}(\mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda)|t_i^* = s, X_i], \mathbb{E}[\Delta\bar{Y}_i^s(0, \lambda)|t_i^* = s, X_i]|t_i^* = s) \\ & + \frac{1}{p_s^2 M^2} \mathbb{E} \left[\left(M + \alpha(M, q) \frac{e_s(X_i)}{e_\infty(X_i)} \right) e_s(X_i) (\mathbb{E}[\Delta\bar{Y}_i^s(0, \lambda)|t_i^* = s, X_i] - \mathbb{E}[\Delta\bar{Y}_i^s(0, \lambda)|t_i^* = s])^2 \right]. \end{aligned}$$

The direction and magnitude of this bias depend on the data generating process. The first and third terms are always positive, whereas the covariance can be positive or negative. For example, if the conditional ATT, $\mathbb{E}[\tau_{it}|t_i^* = s, X_i]$, is constant over X_i , the covariance is zero and thus the naive variance estimator is upward biased. In such cases, ignoring the matching step results in overly conservative standard errors. On the other hand, when the covariance between the conditional ATT and the conditional trend of the untreated outcome is negative and large enough, the bias may be negative, resulting in standard errors that underestimate the true variability of the estimator. We further illustrate these cases with simulations in the next section.

5 Simulations

We now present two simulation studies to illustrate the problems with post-matching 2WFE estimators analyzed in previous sections. In the first study, we consider a staggered treatment adoption setting to illustrate the asymptotic bias. In the second study, we show how failing to account for the matching step when conducting inference results in invalid standard errors, considering cases where the unadjusted standard errors are too large and too small.

5.1 Bias in Staggered Designs

Our first setup consists of panel data with a total of $n = 1,000$ units and four time periods, $t = 1, 2, 3, 4$. Units belong to one of three cohorts, $t_i^* \in \{2, 3, \infty\}$, so the “early adopters” enter treatment in period $t = 2$ and the “late adopters” enter treatment in period $t = 3$. Units are

assigned to cohorts based on a multinomial rule:

$$\mathbb{P}[t_i^* = \infty | X_i] = \frac{1}{1 + \exp(X_i) + \exp(2X_i)}, \quad \mathbb{P}[t_i^* = s | X_i] = \frac{\exp((s-1)X_i)}{1 + \exp(X_i) + \exp(2X_i)}, \quad s = 2, 3$$

where X_i is a scalar covariate with $X_i \sim \text{Uniform}(-1, 2)$. The untreated potential outcome is defined as:

$$Y_{it}(0) = \alpha_i + (t-1) + 5X_i(t-1) + u_{it}, \quad \alpha_i \sim \mathcal{N}(0, 1), \quad u_{it} \sim \mathcal{N}(0, 1)$$

with (α_i, u_{it}, X_i) mutually independent. This implies that $\mathbb{E}[Y_{it}(0) - Y_{it-1}(0) | t_i^* = s, X_i] = 1 + 5X_i$ for all s , so the conditional parallel trends assumption holds (the unconditional parallel trends assumption does not because $\mathbb{E}[X_i | t_i^* = s]$ differs across s). To focus on the bias due to the matching procedure and abstract from the other biases characterized in Theorem 1, we assume that the treatment effect is constant, $\tau = 5$. The model for the treated potential outcome is:

$$Y_{it}(1) = 5 + \alpha_i + (t-1) + 5X_i(t-1) + v_{it}, \quad v_{it} \sim \mathcal{N}(0, 1).$$

We estimate τ by matching each treated unit to a never-treated unit in the first step and running Equation (1) in the reweighted data. The results are displayed in the first column of Table 1. The results show that the bias of the full sample matched 2WFE is about 0.7 or 15% of the true ATT. In addition, the naive cluster-robust standard error severely overestimates the true variability.

5.2 Post-Matching Inference

Our second simulation exercise illustrates how failing to account for the matching step leads to invalid inference. We consider panel data with a total of $n = 1,000$ units and four time periods, $t = 1, 2, 3, 4$. Units belong to one of two cohorts, $t_i^* \in \{3, \infty\}$, so there is a treated cohort that enters treatment in period 3, and a never-treated cohort. Units are assigned to treatment based on a logistic model:

$$\mathbb{P}[t_i^* = 3 | X_i] = \frac{\exp(X_i)}{1 + \exp(X_i)}$$

where X_i is a scalar covariate with $X_i \sim \text{Uniform}[-1/2, 1/2]$. We consider two designs. The first design has a constant treatment effect equal to 5 and the potential outcomes are given by:

$$\begin{aligned} Y_{it}(0) &= \alpha_i + (t-1) + 5X_i(t-1) + u_{it} \\ Y_{it}(1) &= 5 + \alpha_i + (t-1) + 5X_i(t-1) + v_{it} \end{aligned}$$

where as before $u_{it} \sim \mathcal{N}(0, 1)$, $v_{it} \sim \mathcal{N}(0, 1)$ and (α_i, u_{it}, X_i) mutually independent. Because in this setup the treatment effect is constant, Proposition 4 shows that the naive cluster-robust

standard error is biased upwards.

In the second design, the potential outcomes are given by:

$$\begin{aligned}
 Y_{it}(0) &= \alpha_i + (t - 1) - 2X_i(t - 1) + u_{it} \\
 Y_{it}(1) &= 5 + \alpha_i + (t - 1) + 5X_i(t - 1) + v_{it}
 \end{aligned}$$

which implies that the conditional ATT is $\mathbb{E}[Y_{it}(1) - Y_{it}(0)|t_i^* = 3, X_i] = 7X_i(t - 1)$ which changes across X_i and over time. On the other hand, $\mathbb{E}[\Delta Y_{it}(0)|X_i, t_i^* = s] = 1 - 2X_i$ and $\text{Cov}(\mathbb{E}[Y_{it}(1) - Y_{it}(0)|t_i^* = 3, X_i], \mathbb{E}[\Delta Y_{it}(0)|X_i, t_i^* = s]|t_i^* = s) = -14(t - 1)\mathbb{V}[X_i]$ which in this case results in downward-biased standard errors.

The results from these simulations are shown Table 1. Column (2) reports the result for the constant treatment effect scenario. As shown in Proposition 4, when the treatment effect is constant, the naive variance estimator overestimates the true variability. In this case, the Monte Carlo (MC) standard error is 0.086, while the average of the naive standard error across simulations is 0.257, about three times larger. This upward bias results in overcoverage of confidence intervals and loss of power. On the other hand, the standard error estimator that accounts for the variability of the matching procedure is 0.085, almost identical to the MC standard error, and reaches correct coverage.

Column (3) reports the results for the case with a heterogeneous treatment effect and a negative covariance between the conditional ATT and the conditional outcome trend under no treatment, the third term in Proposition 4. In this case, the negative covariance term yields a downward-biased standard error of 0.209, about 13% smaller than the MC standard error (0.239) and the naive CI undercovers the true parameter. As before, the standard error that accounts for the matching procedure gives is very close to the MC standard error and reaches correct coverage.

Table 1: Simulations Results

	(1)	(2)	(3)
Bias	0.702	-0.000	-0.002
MC se	0.195	0.086	0.239
Naive CR se	1.274	0.257	0.209
Coverage (naive)	100.0	100.0	91.2
CR se	-	0.085	0.239
Coverage	-	94.3	95.0

Notes: “Bias” is the average difference between the estimates and the true parameter across simulations. “MC se” is the Monte Carlo standard standard deviation of the estimates. “Naive CR se” is the average naive cluster-robust standard error that ignores the matching step, and “Coverage (naive)” is the coverage of the corresponding 95% confidence interval. “CR se” is the cluster-robust standard error that accounts for the matching step, and “Coverage” is the coverage of the corresponding 95% confidence interval. Results from 5,000 simulations.

6 Empirical Application

We now illustrate our results using data from the National Supported Work (NSW) Demonstration. Using the NSW experimental data as a benchmark, [LaLonde \(1986\)](#) compared different non-experimental methods to determine whether they can replicate the experimental findings. This study sparked an extensive debate about the credibility of non-experimental methods (see [Imbens and Xu, 2025](#), for a recent review).

Following [LaLonde \(1986\)](#), we use data from the Current Population Survey (CPS) to construct a comparison group for the experimental treatment group using nearest-neighbor matching. The outcome of interest is earnings, which is measured in 1974, 1975 (before treatment) and 1978 (after treatment). Our list of covariates includes indicators of age at baseline, indicators of years of education and indicators of being married, Black and Hispanic.

We compare the following specifications using earnings data from 1975 and 1978, which constitutes a two-period panel: (1) a difference in mean earnings between treated and controls in the experimental sample (“Experimental”), (2) an unmatched 2WFE regression (“2WFE”), (3) a naive matched 2WFE regression where standard errors do not account for the matching step (“Naive matched 2WFE”), (4) a matched 2WFE regression calculating valid standard errors (“Matched 2WFE”), (5) a matched 2WFE regression with valid standard errors and bias correction using a linear specification for the covariates (“Matched 2WFE - BC”) and (6) a 2WFE regression on the matched sample, that is, on the sample excluding any untreated units that are not matched to any treated unit, but without accounting for the number of times each untreated is used as a match (“2WFE in matched sample”). All standard errors are clustered at the individual level.

The results are shown in Column (1) of [Table 2](#). The experimental estimate of the ATT is \$886.3 with a p-value of 0.069. We use this value as a benchmark. The unmatched 2WFE regression yields an estimate of about \$1,714, almost twice as big as the experimental benchmark, with a p-value close to zero. The weighted 2WFE that uses matching in the first step gives an estimate of around \$930 (p-value 0.117), much closer to the experimental benchmark. As shown in [Proposition 4](#), the standard error estimator in this specification is inconsistent because it does not account for the matching step. When adjusting for the matching first step, the standard error increases from \$593 to \$635. The bias-corrected matched 2WFE specification reports a slightly lower estimate of \$809.21 with a very similar standard error. Finally, the 2WFE regression dropping unmatched units reveals an estimate of about \$418, approximately half the experimental benchmark.

Column (2) of [Table 2](#) repeats the same analysis for a placebo outcome. In the experimental sample, this corresponds to a difference in mean earnings in 1975 (before the treatment) between treated and controls. For all other specifications, the placebo test uses data on earnings in 1974 and 1975, and can be interpreted as a pre-trends test. Because all the data in these specifications is pre-treatment, all estimated effects should be close to zero and insignificant. While all the estimates are insignificant at the usual levels, the one from the unweighted 2WFE and the one

from the regression that ignores repeated matches show the largest magnitudes and lowest p-values.

7 Extensions to Other Types of Matching

7.1 Matching on Discrete Covariates

When the vector X_i is a low-dimensional vector of discrete covariates, the matching can be done exactly, which eliminates the matching bias. Without loss of generality, suppose that X_i is a scalar taking values in a finite set \mathcal{X} . For such cases, a consistent and asymptotically normal estimator is given by:

$$\hat{\tau}_{\text{disc}}(\lambda) = \frac{1}{N_s} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{\hat{e}_s(X_i)}{\hat{e}_\infty(X_i)} \right) \Delta \bar{Y}_i^s(\lambda)$$

where

$$\hat{e}_s(X_i) = \hat{\mathbb{P}}[t_i^* = s | X_i = x] = \frac{\sum_i \mathbb{1}(t_i^* = s) \mathbb{1}(X_i = x)}{\sum_i \mathbb{1}(X_i = x)}.$$

This estimator that reweights never-treated units by the ratio of the estimated propensity scores is equivalent to the matching DiD estimator (3) that matches each treated unit to all never-treated units in the same covariate cell $X_i = x$. Because in this setting the number of units in each covariate cell increases as $n \rightarrow \infty$, this estimator can be seen as a matching estimator with a diverging number of matches, similar to the ATE estimator considered in a cross-sectional setting under unconfoundedness by [Lin, Ding, and Han \(2023\)](#). This estimator attains the semiparametric efficiency bound (4), as shown in the following result.

Proposition 5 *Under Assumptions 1-4, if $E[Y_{it}(\mathbf{d}_t)^4 | t_i^* = s, X_i = x]$ is bounded and $\mathbb{V}[Y_{it}(\mathbf{d}_t) | t_i^* = s, X_i = x] > 0$ for all \mathbf{d}_t, s and x ,*

$$\sqrt{n}(\hat{\tau}_{\text{disc}}(\lambda) - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s]) \rightarrow_{\mathcal{D}} \mathcal{N}(0, V_{\text{SEB}}(\lambda)).$$

7.2 Matching Without Replacement

Matching is sometimes carried out without replacement. In such cases, whenever an untreated unit is matched to a treated unit, that unit is removed from the pool of untreated. This implies that each treated unit is used at most once, $K_M(i) \in \{0, 1\}$, and the size of the matched sample is $(M+1)N_s$. We now discuss how to adapt our results to the case of matching without replacement, building on the results by [Abadie and Imbens \(2012\)](#) and [Abadie and Spiess \(2022\)](#). For simplicity we consider the case where the population consists of only two cohorts, $t_i^* \in \{s, \infty\}$. In this case,

Table 2: Estimated Effects of the NSW Job Training Program

	(1) ATT estimate	(2) Placebo test
Experiment		
Coefficient	886.30	39.42
Std. err.	488.14	379.01
p-value	0.069	0.917
95% CI	[-70.45 ; 1,843.06]	[-703.44 ; 782.27]
2WFE		
Coefficient	1,714.40	-138.90
Std. err.	485.93	179.79
p-value	<0.001	0.440
95% CI	[761.98 ; 2,666.81]	[-491.30 ; 213.49]
Naive matched 2WFE		
Coefficient	929.82	44.20
Std. err.	593.26	309.65
p-value	0.117	0.886
95% CI	[-232.97 ; 2,092.62]	[-562.71 ; 651.11]
Matched 2WFE		
Coefficient	929.82	44.20
Std. err.	635.38	339.08
p-value	0.143	0.896
95% CI	[-315.52 ; 2,175.17]	[-620.39 ; 708.79]
Matched 2WFE - BC		
Coefficient	809.21	219.10
Std. err.	636.85	342.10
p-value	0.204	0.522
95% CI	[-439.02 ; 2,057.43]	[-451.42 ; 889.61]
2WFE in matched sample		
Coefficient	418.10	-306.91
Std. err.	521.98	215.93
p-value	0.423	0.155
95% CI	[-604.99 ; 1,441.19]	[-730.13 ; 116.31]

Notes: the table show the estimates from (1) a difference in mean earnings between treated and controls in the experimental sample (“Experimental”), (2) an unmatched 2WFE regression (“2WFE”), (3) a naive matched 2WFE regression where standard errors do not account for the matching step (“Naive matched 2WFE”), (4) a matched 2WFE regression calculating valid standard errors (“Matched 2WFE”), (5) a matched 2WFE regression with valid standard errors and bias correction using a linear specification for the covariates (“Matched 2WFE - BC”) and (6) a 2WFE regression on the matched sample, that is, on the sample excluding any untreated units that are not matched to any treated unit, but without accounting for the number of times each untreated is used as a match (“2WFE in matched sample”). All standard errors are clustered at the individual level.

the post-matching 2WFE estimator using matching without replacement is:

$$\hat{\tau}_{\text{nr}}(\lambda) = \frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = s) \left(\Delta Y_i^s(\lambda) - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i, \infty)} \Delta Y_j^s(\lambda) \right).$$

The following proposition summarizes the behavior of the ATT estimator under matching without replacement.

Proposition 6 *Suppose that the sample is obtained by drawing N_s treated observations and N_0 untreated observations independently from the conditional distribution $Y_1, \dots, Y_T, X | t^* = s$ and $Y_1, \dots, Y_T, X | t^* = \infty$, respectively, and $N_s = O(N_0^{1/r})$ for some $r > q/2$. Then, under Assumptions 1, 2, 3 and 5,*

$$\sqrt{(M+1)N_s}(\hat{\tau}_{\text{nr}}(\lambda) - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s]) \rightarrow_{\mathcal{D}} \mathcal{N}(0, V_{\text{nr}}(\lambda))$$

where:

$$\begin{aligned} V_{\text{nr}}(\lambda) &= (1+M)\mathbb{E} \left[\left(\mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s, X_i] - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s] \right)^2 \middle| t_i^* = s \right] \\ &+ (1+M)\mathbb{E} \left[\sigma_\lambda^2(s, X_i) + \frac{\sigma_\lambda^2(0, X_i)}{M} \middle| t_i^* = s \right]. \end{aligned}$$

The conditional sampling scheme in Proposition 6 and the requirement that $N_s = O(N_0^{1/r})$, which implies that the untreated group is “sufficiently larger” than the treated group, ensures that the bias of the matching estimator is $o_{\mathbb{P}}(n^{-1/2})$ (Abadie and Imbens, 2006) and thus bias correction is not required for asymptotic normality. This is not specific to matching without replacement, and can also be applied to the results in the previous sections of the paper.

7.3 Propensity Score Matching

When the number of (continuous) covariates is large, matching directly on the vector of covariates can be computationally cumbersome. In such cases, researchers often resort to matching on the propensity score, which reduces the problem to matching on a single covariate. Rosenbaum and Rubin (1983) showed that the propensity score is a balancing score, and thus matching on the propensity score instead of the full vector of covariates is sufficient for eliminating confounding bias under a selection on observables assumption. In practice, the propensity score is unknown and needs to be estimated, which introduces additional variability that needs to be accounted for when conducting inference. Abadie and Imbens (2016) derive the asymptotic distribution of the cross-sectional propensity score matching estimator. For the sake of brevity, we do not elaborate on the technical aspects of this case, but we note that the results in Abadie and Imbens (2016)

can be adapted to our setting to derive the asymptotic distribution of our pairwise matched DiD estimator when matching on the estimated propensity score.

8 Conclusion

We analyze the commonly employed empirical strategy of estimating 2WFE regressions in a reweighted sample where treated units are matched to never-treated units in a first step. This approach is often used when the parallel trends assumption is believed to hold conditionally on covariates. By formally characterizing the post-matching 2WFE estimators and their asymptotic distribution, we highlight two problems with this strategy that generally invalidates its findings. First, when treated units exhibit variation in treatment timing, the post-matching 2WFE estimator that pools all treated units contains an asymptotic bias that remains even when the treatment effect is constant. The reason for this bias is that the 2WFE estimator implicitly involves comparisons between different treatment cohorts, using one of them as a comparison group. Because different treated cohorts are not matched to each other, they generally have different covariate distributions, and thus comparing their outcome evolutions over time is invalid when the parallel trends assumption holds conditionally, even when the comparison group is not yet treated.

Second, ignoring the matching step when conducting inference based on the post-matching 2WFE estimators results in inconsistent variance estimators, which can be biased upwards or downwards depending on the data generating process. Specifically, the naive variance estimator is upward-biased (conservative) when the treatment effect is constant, but can be downward-biased when there is sufficient treatment effect heterogeneity.

Then, we provide conditions under which the post-matching 2WFE estimator that compares each treated cohort to the never treated separately is consistent for an average of cohort-specific ATTs over time, and derive a closed-form representation of its asymptotic variance which can be used to conduct valid inference post-matching. We also show that this estimator does not generally attain the semiparametric efficiency bound, but it approaches it as the number of neighbors grows. We also illustrate our results using simulations and a reanalysis of the data from the NSW job training program.

While different methods are available to incorporate covariates when the parallel trends assumption holds conditionally, matching-based methods have the advantage of being intuitively appealing, easy to implement, computationally stable and fully nonparametric. Our findings enable empirical researchers to exploit these methods and conduct valid inference on the parameters of interest. Moreover, because we show that post-matching 2WFE estimators can be recast as cross-sectional matching estimators, existing statistical software for matching estimators such as the `teffects nnmatch` package in Stata can be used after appropriately transforming the data.

References

- ABADIE, A. (2005): “Semiparametric difference-in-differences estimators,” *The Review of Economic Studies*, 72(1), 1–19.
- ABADIE, A., AND G. W. IMBENS (2002): “Simple and Bias-Corrected Matching Estimators for Average Treatment Effects,” *NBER working paper*, (t0283).
- (2006): “Large sample properties of matching estimators for average treatment effects,” *Econometrica*, 74(1), 235–267.
- (2008): “On the failure of the bootstrap for matching estimators,” *Econometrica*, 76(6), 1537–1557.
- (2011): “Bias-corrected matching estimators for average treatment effects,” *Journal of Business & Economic Statistics*, 29(1), 1–11.
- (2012): “A martingale representation for matching estimators,” *Journal of the American Statistical Association*, 107(498), 833–843.
- (2016): “Matching on the estimated propensity score,” *Econometrica*, 84(2), 781–807.
- ABADIE, A., AND J. SPIESS (2022): “Robust post-matching inference,” *Journal of the American Statistical Association*, 117(538), 983–995.
- ATHEY, S., AND G. W. IMBENS (2022): “Design-based analysis in Difference-In-Differences settings with staggered adoption,” *Journal of Econometrics*, 226(1), 62–79.
- BORUSYAK, K., X. JARAVEL, AND J. SPIESS (2024): “Revisiting Event-Study Designs: Robust and Efficient Estimation,” *The Review of Economics Studies*, 91(6), 3253–3285.
- CALLAWAY, B., AND P. H. SANT’ ANNA (2021): “Difference-in-differences with multiple time periods,” *Journal of Econometrics*, 225(2), 200–230.
- CHEN, S., AND F. HAN (2024): “On the limiting variance of matching estimators,” *working paper*.
- DE CHAISEMARTIN, C., AND X. D’HAULTFOEUILLE (2020): “Two-way fixed effects estimators with heterogeneous treatment effects,” *American Economic Review*, 110(9), 2964–2996.
- DE CHAISEMARTIN, C., AND X. D’HAULTFOEUILLE (2022): “Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey,” *The Econometrics Journal*.
- GOODMAN-BACON, A. (2021): “Difference-in-differences with variation in treatment timing,” *Journal of Econometrics*, 225(2), 254–277.

- HECKMAN, J. J., H. ICHIMURA, AND P. TODD (1998): “Matching as an econometric evaluation estimator,” *The Review of Economic Studies*, 72(1), 1–19.
- HECKMAN, J. J., H. ICHIMURA, AND P. E. TODD (1997): “Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme,” *The Review of Economic Studies*, 64(4), 605–654.
- IMAI, K., AND I. S. KIM (2021): “On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data,” *Political Analysis*, 29(3), 405–415.
- IMAI, K., I. S. KIM, AND E. H. WANG (2023): “Matching methods for causal inference with time-series cross-sectional data,” *American Journal of Political Science*, 67(3), 587–605.
- IMBENS, G. W., AND Y. XU (2025): “Comparing Experimental and Nonexperimental Methods: What Lessons Have We Learned Four Decades after LaLonde (1986)?,” *Journal of Economic Perspectives*, 39(4), 173–202.
- LALONDE, R. J. (1986): “Evaluating the Econometric Evaluations of Training Programs with Experimental Data,” *The American Economic Review*, 76(4), 604–620.
- LIN, Z., P. DING, AND F. HAN (2023): “Estimation Based on Nearest Neighbor Matching: From Density Ratio to Average Treatment Effect,” *Econometrica*, 91(6), 2187–2217.
- ROSENBAUM, P. R., AND D. B. RUBIN (1983): “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, 70(1), 41–55.
- ROTH, J., P. H. SANT’ANNA, A. BILINSKI, AND J. POE (2023): “What’s trending in difference-in-differences? A synthesis of the recent econometrics literature,” *Journal of Econometrics*, 235(2), 2218–2244.
- SANT’ANNA, P. H., AND J. ZHAO (2020): “Doubly robust difference-in-differences estimators,” *Journal of Econometrics*, 219(1), 101–122.
- STEIGERWALD, D. G., G. VAZQUEZ-BARE, AND J. MAIER (2021): “Measuring heterogeneous effects of environmental policies using panel data,” *Journal of the Association of Environmental and Resource Economists*, 8(2), 277–313.
- SUN, L., AND S. ABRAHAM (2021): “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects,” *Journal of Econometrics*, 225(2), 175–199.

Appendix

A Survey of Empirical Literature

Table A1 lists 22 papers published between 2018 and 2025 in top Economics and Political Science journals using some version of the post-matching 2WFE estimators analyzed in this paper. The journals we consider are, in alphabetical order: American Economic Journal: Applied Economics (AEJEP), American Economic Journal: Economic Policy (AEJEP), American Economic Review (AER), American Political Science Review (APSR) and Quarterly Journal of Economics (QJE).

B Auxiliary results

Lemma B1 *Let*

$$\hat{\tau}(w, \lambda) = \frac{\sum_{i=1}^n \sum_{t=1}^T w_i Y_{it} (D_{it} - \bar{D}_i - \tilde{D}_t + \bar{D})}{\sum_{i=1}^n \sum_{t=1}^T w_i D_{it} (D_{it} - \bar{D}_i - \tilde{D}_t + \bar{D})}$$

where $D_{it} = \mathbb{1}(t \geq t_i^*)$ and

$$\bar{D}_i = \frac{\sum_t \lambda_t D_{it}}{\sum_t \lambda_t}, \quad \tilde{D}_t = \frac{\sum_i w_i D_{it}}{\sum_i w_i}, \quad \bar{D} = \frac{\sum_i \sum_t w_i \lambda_t D_{it}}{\sum_i w_i \sum_t \lambda_t}.$$

Then $\hat{\tau}(w, \lambda) = \hat{\tau}_{\text{num}}(w, \lambda) / \hat{\tau}_{\text{den}}(w, \lambda)$ where:

$$\begin{aligned} \hat{\tau}_{\text{num}}(w, \lambda) &= \sum_{s < \infty} \sum_{s' \neq s} \sum_{t \geq s} \sum_{t' < s} \lambda_t \lambda_{t'} \hat{p}_s^w \hat{p}_{s'}^w \hat{\tau}_{s'}^s(t, t') \\ \hat{\tau}_{\text{den}}(w, \lambda) &= \left(\sum_t \lambda_t \right)^2 \sum_{s < \infty} \hat{p}_s^w \left\{ \hat{p}_\infty^w \Lambda_s (1 - \Lambda_s) + \sum_{s' < s} \hat{p}_{s'}^w \Lambda_s (\Lambda_{s'} - \Lambda_s) + \sum_{s < s' < \infty} \hat{p}_{s'}^w (1 - \Lambda_s) (\Lambda_s - \Lambda_{s'}) \right\} \\ \hat{\tau}_{s'}^s(t, t') &= \frac{\sum_i w_i \mathbb{1}(t_i^* = s) (Y_{it} - Y_{it'})}{\sum_i w_i \mathbb{1}(t_i^* = s)} - \frac{\sum_i w_i \mathbb{1}(t_i^* = s') (Y_{it} - Y_{it'})}{\sum_i w_i \mathbb{1}(t_i^* = s')} \\ \hat{p}_s^w &= \frac{\sum_i w_i \mathbb{1}(t_i^* = s)}{\sum_i w_i}. \end{aligned}$$

Lemma B2 *Consider a subset of the data with two cohorts s and s' , where each unit in cohort s is matched to M nearest neighbors from cohort s' . Let $(Z_i)_{i=1}^n$ be a sequence of iid random variables with $\mathbb{E}[Z_i | t_i^* = s, X_i = x] = \mu_Z(s, x)$ and $\mathbb{V}[Z_i | t_i^* = s, X_i = x] = \sigma_Z^2(s, x)$ with both functions being Lipschitz-continuous and with $\mathbb{E}[Z_i^4 | t_i^* = s, X_i = x]$ uniformly bounded over x for s and s' . Under Assumption 3 and 5(1),*

$$\frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = s') \frac{K_M(i, s)}{M} Z_i \rightarrow_{\mathbb{P}} \mathbb{E}[\mu_Z(s', X_i) | t_i^* = s]$$

Table A1: Survey of Empirical Literature

Author(s)	Title	Year	Journal	Matching type
1 Powell, Seabury	Medical Care Spending and Labor Market Outcomes: Evidence from Workers' Compensation Reforms	2018	AER	nearest neighbor matching
2 Alfaro-arena, Manelici, Vasquez	The Effects of Joining Multinational Supply Chains: New Evidence from Firm to Firm Linkages	2022	QJE	nearest neighbor matching, propensity score matching
3 Aneja and Xu	The Costs of Employment Segregation: Evidence from the Federal Government under Woodrow Wilson	2022	QJE	coarsened Exact Matching
4 Adams, Huttunen, Nix, Zhang	Violence Against Women at Work	2024	QJE	nearest neighbour matching
5 Adams, Huttunen, Nix, Zhang	The Dynamics of Abusive Relationships	2024	QJE	nearest neighbour matching, propensity score matching
6 Goldschmidt and Schmieder	The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure	2017	QJE	nearest neighbour matching, propensity score matching
7 Cooper, Craig, Gaynor, Reenen	The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured	2019	QJE	nearest neighbour matching, mahalnobis matching
8 Aghamolla, Karaca-Mandic, Li, Thakor	Merchants of Death: The Effect of Credit Supply Shocks on Hospital Outcomes	2024	AER	propensity score matching
9 Aneja, Luca, Reshef	The Benefits of Revealing Race: Evidence from Minority-Owned Local Businesses	2025	AER	coarsened exact matching

Table A1: Survey of Empirical Literature (continued)

Author(s)	Title	Year	Journal	Matching type
10 Fenizia and Saggio	Organized Crime and Economic Growth: Evidence from Municipalities Infiltrated by the Mafia	2024	AER	propensity score matching
11 Sviatschi	Spreading Gangs: Exporting US Criminal Capital to El Salvador	2022	AER	propensity score matching
12 Schmieder, Heining	Wachter, The Costs of Job Displacement over the Business Cycle and Its Sources: Evidence from Germany	2023	AER	propensity score matching
13 Fetzer	Did Austerity Cause Brexit?	2019	AER	propensity score matching
14 Muehlenbachs, Timmins	Spiller, The Housing Market Impacts of Shale Gas Development	2015	AER	nearest neighbour matching
15 Fadlon, Nielsen	Family Health Behaviors	2019	AER	propensity score matching
16 Hvide, Jones	University Innovation and the Professor's Privilege	2018	AER	propensity score matching
17 Sigurdsson	Labor Supply Responses and Adjustment Frictions: A Tax-Free Year in Iceland	2018	AEJEP	coarsened exact matching
18 Deryugina, Reif	MacKay, The Long-Run Dynamics of Electricity Demand: Evidence from Municipal Aggregation	2020	AEJAE	nearest neighbour matching
19 Matsa, Miller	A Female Style in Corporate Leadership? Evidence from Quotas	2013	AEJAE	nearest neighbour matching
20 Ben-Manachem, Morris	Turnout: The Participatory Consequences of Low-Level Police Contact	2023	APSR	nearest neighbour matching
21 Incerti	How Firms, Bureaucrats, and Ministries Benefit from the Revolving Door: Evidence from Japan	2025	APSR	nearest neighbour matching
22 Rosenstiel	The Distributive Politics of Grants-in-Aid	2025	APSR	nearest neighbour matching

and

$$\frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = s') \frac{K_M(i, s)^2}{M^2} Z_i^2 \rightarrow_{\mathbb{P}} \frac{1}{M} \mathbb{E} \left[e_s(X_i) \frac{\mathbb{E}[Z_i^2 | t_i^* = s', X_i]}{p_s} \right] + \frac{\alpha(M, q)}{M^2} \mathbb{E} \left[\frac{e_s(X_i)^2 \mathbb{E}[Z_i^2 | t_i^* = s', X_i]}{e_{s'}(X_i) p_s} \right].$$

C Proofs of Main Results

Proof of Lemma 1

$$\begin{aligned} \mathbb{E}[Y_{it} - Y_{it'} | t_i^* = s] &= \mathbb{E}[Y_{it}(t_i^*) - Y_{it'}(0) | t_i^* = s] \\ &= \mathbb{E}[\tau_{it} | t_i^* = s] + \mathbb{E}[Y_{it}(0) - Y_{it'}(0) | t_i^* = s] \\ &= \mathbb{E}[\tau_{it} | t_i^* = s] + \mathbb{E} \{ \mathbb{E}[Y_{it}(0) - Y_{it'}(0) | t_i^* = s, X_i] | t_i^* = s \} \\ &= \mathbb{E}[\tau_{it} | t_i^* = s] + \mathbb{E} \{ \mathbb{E}[Y_{it}(0) - Y_{it'}(0) | t_i^* = s', X_i] | t_i^* = s \} \\ &= \mathbb{E}[\tau_{it} | t_i^* = s] + \mathbb{E} \{ \mathbb{E}[Y_{it} - Y_{it'} | t_i^* = s', X_i] | t_i^* = s \} \end{aligned}$$

as required. \square

Proof of Proposition 1

When $w_i^* = \sum_{s \neq \infty} \mathbb{1}(t_i^* = s) + \mathbb{1}(t_i^* = \infty) K_M(i) / M$,

$$\sum_i w_i^* \mathbb{1}(t_i^* = s) = N_s, \quad \sum_i w_i^* \mathbb{1}(t_i^* = \infty) = \frac{1}{M} \sum_i \mathbb{1}(t_i^* = \infty) K_M(i) = \sum_{s \neq \infty} N_s = n - N_0$$

which implies that $\sum_i w_i^* = 2(n - N_0)$, $\hat{p}_\infty^w = 1/2$ and $\hat{p}_s^w = N_s / (2(n - N_0)) = \hat{p}_s / (2(1 - \hat{p}_\infty))$ for $s \neq \infty$, and the result follows by Lemma B1. \square

Proof of Theorem 1

The numerator of the estimator can be decomposed as:

$$\begin{aligned}
\hat{\tau}_{\text{num}}^{\text{pool}}(\lambda) &= \sum_{s < \infty} \sum_{t \geq s} \sum_{t' < s} \lambda_t \lambda_{t'} \hat{p}_s \hat{\tau}_{\infty}^s(t, t') \\
&+ \sum_{s < \infty} \sum_{s < s' < \infty} \sum_{s \leq t < s'} \sum_{t' < s} \lambda_t \lambda_{t'} \frac{\hat{p}_s \hat{p}_{s'}}{1 - \hat{p}_{\infty}} \hat{\tau}_{s'}^s(t, t') \\
&+ \sum_{s < \infty} \sum_{s' < s} \sum_{t \geq s} \sum_{t' < s'} \lambda_t \lambda_{t'} \frac{\hat{p}_s \hat{p}_{s'}}{1 - \hat{p}_{\infty}} \hat{\tau}_{s'}^s(t, t') \\
&+ \sum_{s < \infty} \sum_{s' < s} \sum_{t \geq s} \sum_{s' \leq t' < s} \lambda_t \lambda_{t'} \frac{\hat{p}_s \hat{p}_{s'}}{1 - \hat{p}_{\infty}} \hat{\tau}_{s'}^s(t, t') \\
&+ \sum_{s < \infty} \sum_{s < s' < \infty} \sum_{t \geq s'} \sum_{t' < s} \lambda_t \lambda_{t'} \frac{\hat{p}_s \hat{p}_{s'}}{1 - \hat{p}_{\infty}} \hat{\tau}_{s'}^s(t, t').
\end{aligned}$$

Of these five set of comparisons, only the first two use comparison units are untreated in both periods. This information is summarized in the table below.

Comparison	Cohort	Status in t'	Status in t
$t' < s \leq t, s' = \infty$	Never	Untreated	Untreated
$t' < s \leq t < s' < \infty$	Later	Untreated	Untreated
$t' < s < s' \leq t$	Later	Untreated	Treated
$t' < s' < s \leq t$	Earlier	Untreated	Treated
$s' \leq t' < s \leq t$	Earlier	Treated	Treated

Let

$$B_{s'}^s(t, t') = \mathbb{E}[Y_{it}(0) - Y_{it'}(0) | t_i^* = s] - \mathbb{E}[Y_{it}(0) - Y_{it'}(0) | t_i^* = s'],$$

By Lemma B2 and Lemma 1,

$$\begin{aligned}
\hat{\tau}_{s'}^s(t, t') &\rightarrow_{\mathbb{P}} \mathbb{E}[\tau_{it} | t_i^* = s] \mathbb{1}(s' > t) + (\mathbb{E}[\tau_{it} | t_i^* = s] - \mathbb{E}[\tau_{it} - \tau_{it'} | t_i^* = s']) \mathbb{1}(s' \leq t') \\
&+ (\mathbb{E}[\tau_{it} | t_i^* = s] - \mathbb{E}[\tau_{it} | t_i^* = s']) \mathbb{1}(t' < s' \leq t) \\
&+ B_{s'}^s(t, t') \mathbb{1}(s' \neq \infty)
\end{aligned}$$

and by the law of large numbers $\hat{p}_s \rightarrow_{\mathbb{P}} p_s$. Then,

$$\begin{aligned}
\hat{\tau}_{\text{num}}^{\text{pool}}(\lambda) &\rightarrow_{\mathbb{P}} \sum_{s < \infty} p_s \sum_{t \geq s} \lambda_t \mathbb{E}[\tau_{it} | t_i^* = s] \left(\sum_{t' < s} \lambda_{t'} \right) \\
&+ \sum_{s < \infty} \sum_{s < s' < \infty} \sum_{s \leq t < s'} \sum_{t' < s} \lambda_t \lambda_{t'} \frac{p_s p_{s'}}{1 - p_\infty} (\mathbb{E}[\tau_{it} | t_i^* = s] + B_{s'}^s(t, t')) \\
&+ \sum_{s < \infty} \sum_{s' < s} \sum_{t \geq s} \sum_{t' < s'} \lambda_t \lambda_{t'} \frac{p_s p_{s'}}{1 - p_\infty} (\mathbb{E}[\tau_{it} | t_i^* = s] - \mathbb{E}[\tau_{it} | t_i^* = s'] + B_{s'}^s(t, t')) \\
&+ \sum_{s < \infty} \sum_{s' < s} \sum_{t \geq s} \sum_{s' \leq t' < s} \lambda_t \lambda_{t'} \frac{p_s p_{s'}}{1 - p_\infty} (\mathbb{E}[\tau_{it} | t_i^* = s] - \mathbb{E}[\tau_{it} - \tau_{it'} | t_i^* = s'] + B_{s'}^s(t, t')) \\
&+ \sum_{s < \infty} \sum_{s < s' < \infty} \sum_{t \geq s'} \sum_{t' < s} \lambda_t \lambda_{t'} \frac{p_s p_{s'}}{1 - p_\infty} (\mathbb{E}[\tau_{it} | t_i^* = s] - \mathbb{E}[\tau_{it} | t_i^* = s'] + B_{s'}^s(t, t')).
\end{aligned}$$

Rearranging,

$$\begin{aligned}
\hat{\tau}_{\text{num}}^{\text{pool}}(\lambda) &\rightarrow_{\mathbb{P}} \sum_{s < \infty} p_s \Lambda_s (1 - \Lambda_s) \frac{\sum_{t \geq s} \lambda_t \mathbb{E}[\tau_{it} | t_i^* = s]}{\sum_{t \geq s} \lambda_t} \left(\sum_t \lambda_t \right)^2 \\
&+ \sum_{s < \infty} \sum_{s' < s} \frac{p_s p_{s'}}{1 - p_\infty} \Lambda_s (\Lambda_{s'} - \Lambda_s) \frac{\sum_{t \geq s} \lambda_t \mathbb{E}[\tau_{it} | t_i^* = s]}{\sum_{t \geq s} \lambda_t} \left(\sum_t \lambda_t \right)^2 \\
&+ \sum_{s < \infty} \sum_{s < s' < \infty} \frac{p_s p_{s'}}{1 - p_\infty} (1 - \Lambda_s) (\Lambda_s - \Lambda_{s'}) \frac{\sum_{s < t < s'} \lambda_t \mathbb{E}[\tau_{it} | t_i^* = s]}{\sum_{s < t < s'} \lambda_t} \left(\sum_t \lambda_t \right)^2 \\
&+ \sum_{s < \infty} \sum_{s' \neq s, \infty} \frac{p_s p_{s'}}{1 - p_\infty} \Lambda_{\max\{s, s'\}} (1 - \Lambda_{\min\{s, s'\}}) \frac{\sum_{t \geq \max\{s, s'\}} \lambda_t (\mathbb{E}[\tau_{it} | t_i^* = s] - \mathbb{E}[\tau_{it} | t_i^* = s'])}{\sum_{t \geq \max\{s, s'\}} \lambda_t} \left(\sum_t \lambda_t \right)^2 \\
&- \sum_{s < \infty} \sum_{s' < s} \frac{p_s p_{s'}}{1 - p_\infty} \Lambda_s (\Lambda_{s'} - \Lambda_s) \frac{\sum_{t \geq s} \sum_{s' \leq t' < s} \lambda_t \lambda_{t'} \mathbb{E}[\tau_{it} - \tau_{it'} | t_i^* = s']}{\sum_{t \geq s} \sum_{s' \leq t' < s} \lambda_t \lambda_{t'}} \left(\sum_t \lambda_t \right)^2 \\
&+ \sum_{s < \infty} \sum_{s' \neq s, \infty} \frac{p_s p_{s'}}{1 - p_\infty} \Lambda_s (1 - \Lambda_s) \frac{\sum_{t \geq s} \sum_{t' < s} \lambda_t \lambda_{t'} B_{s'}^s(t, t')}{\sum_{t \geq s} \sum_{t' < s} \lambda_t \lambda_{t'}} \left(\sum_t \lambda_t \right)^2.
\end{aligned}$$

On the other hand, for the denominator,

$$\begin{aligned}
\hat{\tau}_{\text{den}}^{\text{pool}}(\lambda) &\rightarrow \sum_{s < \infty} p_s \left\{ \Lambda_s (1 - \Lambda_s) + \sum_{s' < s} \frac{p_{s'}}{1 - p_\infty} \Lambda_s (\Lambda_{s'} - \Lambda_s) + \sum_{s < s' < \infty} \frac{p_{s'}}{1 - p_\infty} (1 - \Lambda_s) (\Lambda_s - \Lambda_{s'}) \right\} \\
&\times \left(\sum_t \lambda_t \right)^2
\end{aligned}$$

and the result follows after simplifying the term $(\sum_t \lambda_t)^2$. \square

Proof of Proposition 2

When $w_i^* = \mathbb{1}(t_i^* = s) + \mathbb{1}(t_i^* = \infty)K_M(i, s)/M$,

$$\sum_i w_i^* \mathbb{1}(t_i^* = s) = N_s, \quad \sum_i w_i^* \mathbb{1}(t_i^* = \infty) = \frac{1}{M} \sum_i \mathbb{1}(t_i^* = \infty)K_M(i, s) = N_s$$

using that $\sum_i \mathbb{1}(t_i^* = \infty)K_M(i, s) = N_s M$ because the number of times each untreated unit is used as a match equals the total number of matched units for the treated. This implies that $\sum_i w_i^* = 2N_s$ and $\hat{p}_\infty^w = \hat{p}_s^w = 1/2$ and the result follows by Lemma B1. \square

Proof of Theorem 2

This result follows from Lemma 1 and Lemma B2 under conditional parallel trends. \square

Proof of Theorem 3

Define $\Delta\bar{Y}_i^s(\lambda) = \bar{Y}_i^{\text{post},s} - \bar{Y}_i^{\text{pre},s}$, $\mu_\lambda(s, x) = \mathbb{E}[\Delta\bar{Y}_i^s(\lambda)|t_i^* = s, X_i = x]$ and $\sigma_\lambda^2(s, x) = \mathbb{V}[\Delta\bar{Y}_i^s(\lambda)|t_i^* = s, X_i = x]$ where we use the notation $\mu_\lambda(0, x)$ and $\sigma_\lambda^2(0, x)$ when $s = \infty$. Also let

$$\bar{\tau}_i^{\text{post},s}(\lambda) = \frac{\sum_{t \geq s} \lambda_t \tau_{it}}{\sum_{t \geq s} \lambda_t}$$

By Assumption 2,

$$\mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda)|t_i^* = s, X_i] = \mu_\lambda(s, X_i) - \mu_\lambda(0, X_i).$$

By Proposition 2, the weighted estimator is:

$$\hat{\tau}(w^s, \lambda) = \frac{1}{N_s} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)}{M} \right) \Delta\bar{Y}_i^s(\lambda)$$

This estimator can be rewritten as:

$$\begin{aligned} \hat{\tau}(w^s, \lambda) &= \frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = s) (\mu_\lambda(s, X_i) - \mu_\lambda(0, X_i)) \\ &\quad + \frac{1}{N_s} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)}{M} \right) (\Delta\bar{Y}_i^s(\lambda) - \mu_\lambda(t_i^*, X_i)) \\ &\quad + B_\lambda(s, M) \end{aligned}$$

where

$$\begin{aligned} B_\lambda(s, M) &= \frac{1}{N_s} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)}{M} \right) \mu_\lambda(0, X_i) \\ &= \frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = s) \frac{1}{M} \sum_{m=1}^M (\mu_\lambda(0, X_i) - \mu_\lambda(0, X_{j_m(i, \infty)})). \end{aligned}$$

Then,

$$\begin{aligned} \sqrt{n} (\hat{\tau}(w^s, \lambda) - \mathbb{E}[\bar{\tau}_i^{\text{post}, s}(\lambda) | t_i^* = s] - B_\lambda(s, M)) &= \\ \frac{n}{N_s} \left\{ \frac{1}{\sqrt{n}} \sum_i \mathbb{1}(t_i^* = s) (\mu_\lambda(s, X_i) - \mu_\lambda(0, X_i) - \mathbb{E}[\bar{\tau}_i^{\text{post}, s}(\lambda) | t_i^* = s]) \right\} \\ + \frac{n}{N_s} \left\{ \frac{1}{\sqrt{n}} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)}{M} \right) \varepsilon_i(t_i^*, X_i) \right\} \end{aligned}$$

where $\varepsilon_i(t_i^*, X_i) = \Delta \bar{Y}_i^s(\lambda) - \mu_\lambda(t_i^*, X_i)$. Define:

$$Z_{n,i} = \begin{cases} \frac{1}{\sqrt{n}} \mathbb{1}(t_i^* = s) (\mu_\lambda(s, X_i) - \mu_\lambda(0, X_i) - \mathbb{E}[\bar{\tau}_i^{\text{post}, s}(\lambda) | t_i^* = s]) & \text{if } 1 \leq i \leq n \\ \frac{1}{\sqrt{n}} \left(\mathbb{1}(t_{i-n}^* = s) - \mathbb{1}(t_{i-n}^* = \infty) \frac{K_M(i-n, s)}{M} \right) \varepsilon_{i-n}(t_{i-n}^*, X_{i-n}) & \text{if } n+1 \leq i \leq 2n \end{cases}$$

so that

$$\sqrt{n} (\hat{\tau}(w^s, \lambda) - \mathbb{E}[\bar{\tau}_i^{\text{post}, s}(\lambda) | t_i^* = s] - B_\lambda(s, M)) = \left(\frac{1}{p_s} + o_{\mathbb{P}}(1) \right) \sum_{i=1}^{2n} Z_{n,i}.$$

Define the σ -fields:

$$\mathcal{F}_{n,i} = \begin{cases} \sigma(\mathbf{t}^*, X_1, \dots, X_i) & \text{if } 1 \leq i \leq n \\ \sigma(\mathbf{t}^*, \mathbf{X}, \Delta \bar{Y}_1^s(\lambda), \dots, \Delta \bar{Y}_{i-n}^s(\lambda)) & \text{if } n+1 \leq i \leq 2n. \end{cases}$$

Then, the sequence

$$\left\{ \left(\sum_{j=1}^i Z_{n,j}, \mathcal{F}_{n,i} \right), 1 \leq i \leq 2n \right\}$$

is a martingale. Now, for $1 \leq i \leq n$,

$$\begin{aligned} \mathbb{E}[Z_{n,i}^2 | \mathcal{F}_{n,i}] &= \mathbb{E} \left[\left(\frac{1}{\sqrt{n}} \mathbb{1}(t_i^* = s) (\mu_\lambda(s, X_i) - \mu_\lambda(0, X_i) - \mathbb{E}[\bar{\tau}_i^{\text{post}, s}(\lambda) | t_i^* = s]) \right)^2 \middle| \mathbf{t}^*, X_1, \dots, X_{i-1} \right] \\ &= \frac{\mathbb{1}(t_i^* = s)}{n} \mathbb{E} \left[(\mu_\lambda(s, X_i) - \mu_\lambda(0, X_i) - \mathbb{E}[\bar{\tau}_i^{\text{post}, s}(\lambda) | t_i^* = s])^2 \middle| t_i^* = s \right] \end{aligned}$$

and for $n + 1 \leq i \leq 2n$,

$$\begin{aligned}
\mathbb{E}[Z_{n,i}^2 | \mathcal{F}_{n,i}] &= \mathbb{E} \left[\left(\frac{1}{\sqrt{n}} \left(\mathbb{1}(t_{i-n}^* = s) - \mathbb{1}(t_{i-n}^* = s') \frac{K_M(i-n, s)}{M} \right) \varepsilon_{i-n}(t_{i-n}^*, X_{i-n}) \right)^2 \middle| \mathcal{F}_{n,i-1} \right] \\
&= \frac{1}{n} \left(\mathbb{1}(t_{i-n}^* = s) - \mathbb{1}(t_{i-n}^* = s') \frac{K_M(i-n, s)}{M} \right)^2 \mathbb{E} [\varepsilon_{i-n}^2(t_{i-n}^*, X_{i-n}) | t_{i-n}^*, X_{i-n}] \\
&= \frac{\mathbb{1}(t_{i-n}^* = s)}{n} \sigma_\lambda^2(s, X_{i-n}) + \frac{\mathbb{1}(t_{i-n}^* = \infty)}{n} \cdot \frac{K_M(i-n, s)^2}{M^2} \sigma_\lambda^2(0, X_{i-n}).
\end{aligned}$$

It follows that:

$$\begin{aligned}
\sum_{i=1}^{2n} \mathbb{E}[Z_{n,i}^2 | \mathcal{F}_{n,i}] &= \frac{N_s}{n} \mathbb{E} \left[\left(\mu_\lambda(s, X_i) - \mu_\lambda(0, X_i) - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s] \right)^2 \middle| t_i^* = s \right] \\
&\quad + \frac{1}{n} \sum_{i=n+1}^{2n} \mathbb{1}(t_{i-n}^* = s) \sigma_\lambda^2(s, X_{i-n}) + \frac{1}{n} \sum_{i=n+1}^{2n} \mathbb{1}(t_{i-n}^* = \infty) \frac{K_M(i-n, s)^2}{M^2} \sigma_\lambda^2(0, X_{i-n}) \\
&\rightarrow_{\mathbb{P}} p_s \mathbb{E} \left[\left(\mu_\lambda(s, X_i) - \mu_\lambda(0, X_i) - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s] \right)^2 \middle| t_i^* = s \right] \\
&\quad + p_s \mathbb{E}[\sigma_\lambda^2(s, X_i) | t_i^* = s] + \frac{1}{M} \mathbb{E} [e_s(X_i) \sigma_\lambda^2(0, X_i)] + \frac{\alpha(M, q)}{M^2} \mathbb{E} \left[\frac{e_s(X_i)^2}{e_\infty(X_i)} \sigma_\lambda^2(0, X_i) \right]
\end{aligned}$$

by Lemma B2. On the other hand, under finite moments, for $1 \leq i \leq n$,

$$\begin{aligned}
\mathbb{E}[Z_{n,i}^4 | \mathcal{F}_{n,i}] &= \mathbb{E} \left[\left(\frac{1}{\sqrt{n}} \mathbb{1}(t_i^* = s) \left(\mu_\lambda(s, X_i) - \mu_\lambda(0, X_i) - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s] \right) \right)^4 \middle| \mathbf{t}^*, X_1, \dots, X_{i-1} \right] \\
&= \frac{\mathbb{1}(t_i^* = s)}{n^2} \mathbb{E} \left[\left(\mu_\lambda(s, X_i) - \mu_\lambda(s', X_i) - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s] \right)^4 \middle| t_i^* = s \right] \\
&\leq \frac{\mathbb{1}(t_i^* = s)}{n^2} C
\end{aligned}$$

for some finite constant C and for $n + 1 \leq i \leq 2n$,

$$\begin{aligned}
\mathbb{E}[Z_{n,i}^4 | \mathcal{F}_{n,i}] &= \mathbb{E} \left[\left(\frac{1}{\sqrt{n}} \left(\mathbb{1}(t_{i-n}^* = s) - \mathbb{1}(t_{i-n}^* = \infty) \frac{K_M(i-n, s)}{M} \right) \varepsilon_{i-n}(t_{i-n}^*, X_{i-n}) \right)^4 \middle| \mathcal{F}_{n,i-1} \right] \\
&= \frac{1}{n^2} \left(\mathbb{1}(t_{i-n}^* = s) - \mathbb{1}(t_{i-n}^* = s') \frac{K_M(i-n, s)}{M} \right)^4 \mathbb{E} [\varepsilon_{i-n}^4(t_{i-n}^*, X_{i-n}) | t_{i-n}^*, X_{i-n}] \\
&= \frac{\mathbb{1}(t_{i-n}^* = s)}{n^2} \mathbb{E} [\varepsilon_{i-n}^4(t_{i-n}^*, X_{i-n}) | t_{i-n}^* = s, X_{i-n}] \\
&\quad + \frac{\mathbb{1}(t_{i-n}^* = \infty)}{n^2} \cdot \frac{K_M(i-n, s)^4}{M^2} \mathbb{E} [\varepsilon_{i-n}^4(t_{i-n}^*, X_{i-n}) | t_{i-n}^* = \infty, X_{i-n}] \\
&\leq C' \left(\frac{\mathbb{1}(t_{i-n}^* = s)}{n^2} + \frac{\mathbb{1}(t_{i-n}^* = \infty)}{n^2} \cdot \frac{K_M(i-n, s)^4}{M^2} \right)
\end{aligned}$$

for some finite constant C' . Thus,

$$\begin{aligned} \sum_{i=1}^{2n} \mathbb{E}[Z_{n,i}^4 | \mathcal{F}_{n,i}] &\leq \frac{C}{n} \cdot \frac{1}{n} \sum_{i=1}^n \mathbb{1}(t_i^* = s) + \frac{C'}{n} \left(\frac{1}{n} \sum_{i=n+1}^{2n} \mathbb{1}(t_{i-n}^* = s) + \frac{1}{n} \sum_{i=n+1}^{2n} \frac{\mathbb{1}(t_{i-n}^* = \infty) K_M(i-n, s)^4}{M^2} \right) \\ &\rightarrow_{\mathbb{P}} 0 \end{aligned}$$

because by Lemma 3 in [Abadie and Imbens \(2006\)](#) $\mathbb{E}[K_M(i, s)^4]$ is uniformly bounded. Thus, by the martingale central limit theorem and the Slutsky theorem,

$$\sqrt{n} (\hat{\tau}(w^s, \lambda) - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s] - B_\lambda(s, M)) \rightarrow_{\mathcal{D}} \mathcal{N}(0, V(s, \lambda))$$

where

$$\begin{aligned} V(s, \lambda) &= \frac{1}{p_s} \mathbb{E} \left[(\mu_\lambda(s, X_i) - \mu_\lambda(0, X_i) - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s])^2 \middle| t_i^* = s \right] + \frac{1}{p_s} \mathbb{E}[\sigma_\lambda^2(s, X_i) | t_i^* = s] \\ &\quad + \frac{1}{p_s^2 M^2} \mathbb{E} \left[\left(M + \alpha(M, q) \frac{e_s(X_i)}{e_\infty(X_i)} \right) e_s(X_i) \sigma_\lambda^2(0, X_i) \right] \end{aligned}$$

which completes the proof. \square

Proof of Proposition 3

From Theorem 3,

$$\begin{aligned} V(s, \lambda) &= \frac{1}{p_s} \mathbb{E} \left[(\mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s, X_i] - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s])^2 \middle| t_i^* = s \right] + \frac{1}{p_s} \mathbb{E}[\sigma_\lambda^2(s, X_i) | t_i^* = s] \\ &\quad + \frac{1}{p_s^2 M^2} \mathbb{E} \left[\left(M + \alpha(M, q) \frac{e_s(X_i)}{e_\infty(X_i)} \right) e_s(X_i) \sigma_\lambda^2(0, X_i) \right] \end{aligned}$$

but

$$\begin{aligned} \frac{1}{p_s} \mathbb{E} \left[(\mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s, X_i] - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s])^2 \middle| t_i^* = s \right] &= \\ \frac{1}{p_s^2} \mathbb{E} \left[(\mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s, X_i] - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s])^2 e_s(X_i) \right] & \end{aligned}$$

by the law of iterated expectations. Similarly,

$$\frac{1}{p_s} \mathbb{E} [\sigma_\lambda^2(s, X_i) | t_i^* = s] = \frac{1}{p_s^2} \mathbb{E} [\sigma_\lambda^2(s, X_i) e_s(X_i)]$$

and the result follows from straightforward algebra. \square

Proof of Proposition 4

The regression residual is:

$$\begin{aligned}\hat{\varepsilon}_i &= \Delta\bar{Y}_i^s(\lambda) - \left(\frac{1}{N_s} \sum_j \mathbb{1}(t_j^* = s) \Delta\bar{Y}_j^s(\lambda) \right) \mathbb{1}(t_i^* = s) \\ &\quad - \left(\frac{1}{N_s} \sum_j \mathbb{1}(t_j^* = \infty) \frac{K_M(j, s)}{M} \Delta\bar{Y}_j^s(\lambda) \right) \mathbb{1}(t_i^* = \infty) \\ &= \Delta\bar{Y}_i^s(\lambda) - W_i' \hat{\mu}\end{aligned}$$

where

$$W_i = (\mathbb{1}(t_i^* = s), \mathbb{1}(t_i^* = \infty))', \quad \hat{\mu} = \left(\frac{1}{N_s} \sum_j \mathbb{1}(t_j^* = s) \Delta\bar{Y}_j^s(\lambda), \frac{1}{N_s} \sum_j \mathbb{1}(t_j^* = \infty) \frac{K_M(j, s)}{M} \Delta\bar{Y}_j^s(\lambda) \right)'.$$

Start by noting that by the law of large numbers and Lemma B2,

$$\hat{\mu} \rightarrow_{\mathbb{P}} \mu := (\mathbb{E}[\Delta\bar{Y}_i^s(\lambda)|t_i^* = s], \mathbb{E}[\mathbb{E}[\Delta\bar{Y}_i^s(\lambda)|t_i^* = \infty, X_i]|t_i^* = s])'.$$

Define $\varepsilon_i = \Delta\bar{Y}_i^s(\lambda) - W_i' \mu$ and note that $\hat{\varepsilon}_i = \varepsilon_i - W_i'(\hat{\mu} - \mu)$. By the triangle and Schwarz inequalities,

$$|\hat{\varepsilon}_i^2 - \varepsilon_i^2| \leq \|W_i\| \|\hat{\mu} - \mu\|^2 + 2|\varepsilon_i| \|W_i\| \|\hat{\mu} - \mu\| \leq \|\hat{\mu} - \mu\|^2 + 2|\varepsilon_i| \|\hat{\mu} - \mu\|$$

using that $\|W_i\| = \mathbb{1}(t_i^* = s) + \mathbb{1}(t_i^* = \infty) \leq 1$. Then we have that:

$$\begin{aligned}n\hat{V}_{\text{cr}} &= \frac{n}{N_s^2} \sum_i \mathbb{1}(t_i^* = s) \hat{\varepsilon}_i^2 + \frac{n}{N_s^2} \sum_i \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)^2}{M^2} \hat{\varepsilon}_i^2 \\ &= \frac{n^2}{N_s^2} \left\{ \frac{1}{n} \sum_i \mathbb{1}(t_i^* = s) \varepsilon_i^2 + \frac{1}{n} \sum_i \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)^2}{M^2} \varepsilon_i^2 \right\} \\ &\quad + \frac{n^2}{N_s^2} \cdot \frac{1}{n} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)}{M} \right)^2 (\hat{\varepsilon}_i^2 - \varepsilon_i^2).\end{aligned}$$

Define

$$R_n = \frac{1}{n} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)}{M} \right)^2 (\hat{\varepsilon}_i^2 - \varepsilon_i^2).$$

We have that:

$$\begin{aligned}
|R_n| &\leq \frac{1}{n} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)}{M} \right)^2 \|\hat{\mu} - \mu\|^2 \\
&\quad + \frac{2}{n} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)}{M} \right)^2 |\varepsilon_i| \|\hat{\mu} - \mu\| \\
&:= R_{1,n} \|\hat{\mu} - \mu\|^2 + 2R_{2,n} \|\hat{\mu} - \mu\|.
\end{aligned}$$

Note that

$$\mathbb{E}[R_{1,n}] \leq 1 + \frac{1}{M^2} \mathbb{E} [K_M(i, s)^2] < C$$

by Lemma 3 in [Abadie and Imbens \(2006\)](#) and thus by Markov's inequality $R_{1,n} = O_{\mathbb{P}}(1)$ and $R_{1,n} \|\hat{\mu} - \mu\|^2 = o_{\mathbb{P}}(1)$. Similarly, for the second term,

$$\begin{aligned}
\mathbb{E}[R_{2,n}] &= \mathbb{E} \left[\left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)}{M} \right)^2 |\varepsilon_i| \right] \\
&\leq \left(1 + \frac{\mathbb{E} [K_M(i, s)^4]}{M^4} \right)^{1/2} \mathbb{E}[\varepsilon_i^2]^{1/2}
\end{aligned}$$

which again is uniformly bounded and thus $2R_{2,n} \|\hat{\mu} - \mu\| = o_{\mathbb{P}}(1)$. Then,

$$\begin{aligned}
n\hat{V}_{\text{cr}} &= \frac{n^2}{N_s^2} \left\{ \frac{1}{n} \sum_i \mathbb{1}(t_i^* = s) \varepsilon_i^2 + \frac{1}{n} \sum_i \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s)^2}{M^2} \varepsilon_i^2 \right\} + o_{\mathbb{P}}(1) \\
&:= \frac{n^2}{N_s^2} \{T_{1,n} + T_{2,n}\} + o_{\mathbb{P}}(1)
\end{aligned}$$

By the law of large numbers, $T_{1,n} \rightarrow_{\mathbb{P}} \mathbb{E}[\varepsilon_i^2 | t_i^* = s] p_s$. On the other hand, by Lemma [B2](#),

$$T_{2,n} \rightarrow_{\mathbb{P}} \frac{1}{M} \mathbb{E} [e_s(X_i) \mathbb{E}[\varepsilon_i^2 | t_i^* = \infty, X_i]] + \frac{\alpha(M, q)}{M^2} \mathbb{E} \left[\frac{e_s(X_i)^2}{e_{\infty}(X_i)} \mathbb{E}[\varepsilon_i^2 | t_i^* = \infty, X_i] \right].$$

Collecting the results,

$$\begin{aligned}
n\hat{V}_{\text{cr}} &\rightarrow_{\mathbb{P}} \frac{\mathbb{E}[\varepsilon_i^2 | t_i^* = s]}{p_s} + \frac{1}{Mp_s^2} \mathbb{E} [e_s(X_i) \mathbb{E}[\varepsilon_i^2 | t_i^* = \infty, X_i]] + \frac{\alpha(M, q)}{M^2 p_s^2} \mathbb{E} \left[\frac{e_s(X_i)^2}{e_{\infty}(X_i)} \mathbb{E}[\varepsilon_i^2 | t_i^* = \infty, X_i] \right] \\
&= \frac{\mathbb{E}[\varepsilon_i^2 | t_i^* = s]}{p_s} + \frac{1}{p_s^2 M^2} \mathbb{E} \left[\left(M + \alpha(M, q) \frac{e_s(X_i)}{e_{\infty}(X_i)} \right) e_s(X_i) \mathbb{E}[\varepsilon_i^2 | t_i^* = \infty, X_i] \right].
\end{aligned}$$

Now,

$$\begin{aligned}
\mathbb{E}[\varepsilon_i^2 | t_i^* = s] &= \mathbb{E}[(\Delta \bar{Y}_i^s(\lambda) - \mathbb{E}[\Delta \bar{Y}_i^s(\lambda) | t_i^* = s])^2 | t_i^* = s] = \mathbb{V}[\Delta \bar{Y}_i^s(\lambda) | t_i^* = s] \\
&= \mathbb{V}[\mu_\lambda(s, X_i) | t_i^* = s] + \mathbb{E}[\sigma_\lambda^2(s, X_i) | t_i^* = s] \\
&= \mathbb{V}[\mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s, X_i] | t_i^* = s] + \mathbb{V}[\mathbb{E}[\Delta \bar{Y}_i^s(0, \lambda) | t_i^* = s, X_i] | t_i^* = s] \\
&\quad + 2\text{Cov}(\mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s, X_i], \mathbb{E}[\Delta \bar{Y}_i^s(0, \lambda) | t_i^* = s, X_i] | t_i^* = s) + \mathbb{E}[\sigma_\lambda^2(s, X_i) | t_i^* = s]
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[\varepsilon_i^2 | t_i^* = \infty, X_i] &= \mathbb{E}[(\Delta \bar{Y}_i^s(\lambda) - \mathbb{E}[\mu(0, X_i) | t_i^* = s])^2 | t_i^* = \infty, X_i] \\
&= \mathbb{E}[(\Delta \bar{Y}_i^s(\lambda) - \mu_\lambda(0, X_i))^2 | t_i^* = \infty, X_i] + \mathbb{E}[(\mu_\lambda(0, X_i) - \mathbb{E}[\mu_\lambda(0, X_i) | t_i^* = s])^2 | t_i^* = \infty, X_i] \\
&= \sigma_\lambda^2(0, X_i) + (\mu_\lambda(0, X_i) - \mathbb{E}[\mu(0, X_i) | t_i^* = s])^2
\end{aligned}$$

Next, use that under conditional parallel trends,

$$\begin{aligned}
\mu_\lambda(0, X_i) &= \mathbb{E}[\Delta \bar{Y}_i^s(0, \lambda) | t_i^* = s, X_i] \\
\mathbb{E}[\mu_\lambda(0, X_i) | t_i^* = s] &= \mathbb{E}[\Delta \bar{Y}_i^s(0, \lambda) | t_i^* = s]
\end{aligned}$$

Collecting the results,

$$\begin{aligned}
n\hat{V}_{\text{cr}} &\rightarrow_{\mathbb{P}} V(s, \lambda) \\
&\quad + \frac{1}{p_s} \mathbb{V}[\mathbb{E}[\Delta \bar{Y}_i^s(0, \lambda) | t_i^* = s, X_i] | t_i^* = s] \\
&\quad + \frac{2}{p_s} \text{Cov}(\mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s, X_i], \mathbb{E}[\Delta \bar{Y}_i^s(0, \lambda) | t_i^* = s, X_i] | t_i^* = s) \\
&\quad + \frac{1}{p_s^2 M^2} \mathbb{E} \left[\left(M + \alpha(M, q) \frac{e_s(X_i)}{e_\infty(X_i)} \right) e_s(X_i) (\mathbb{E}[\Delta \bar{Y}_i^s(0, \lambda) | t_i^* = s, X_i] - \mathbb{E}[\Delta \bar{Y}_i^s(0, \lambda) | t_i^* = s])^2 \right].
\end{aligned}$$

as required. \square

Proof of Proposition 5

Suppose the vector X_i includes only discrete variables. By Lemma 1,

$$\begin{aligned}
\mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s] &= \mathbb{E}[\Delta \bar{Y}_i^s(\lambda) | t_i^* = s] - \sum_{x \in \mathcal{X}} \mu_\lambda(0, x) \mathbb{P}[X_i = x | t_i^* = s] \\
&= \mathbb{E}[\Delta \bar{Y}_i^s(\lambda) | t_i^* = s] - \sum_{x \in \mathcal{X}} \mu_\lambda(0, x) \mathbb{P}[t_i^* = s | X_i = x] \frac{p(x)}{p_s} \\
&= \mathbb{E}[\Delta \bar{Y}_i^s(\lambda) | t_i^* = s] - \sum_{x \in \mathcal{X}} \mu_\lambda(0, x) e_s(x) \frac{p(x)}{p_s}
\end{aligned}$$

where $e_s(x) = \mathbb{P}[t_i^* = s | X_i = x]$, $p(x) = \mathbb{P}[X_i = x]$ and $p_s = \mathbb{P}[t_i^* = s]$. This term can be estimated as:

$$\begin{aligned}
\hat{\tau}(s, \lambda) &= \frac{1}{N_s} \sum_i \Delta \bar{Y}_i^s(\lambda) \mathbb{1}(t_i^* = s) - \sum_{x \in \mathcal{X}} \frac{\sum_i \Delta \bar{Y}_i^s(\lambda) \mathbb{1}(t_i^* = \infty) \mathbb{1}(X_i = x)}{\sum_i \mathbb{1}(t_i^* = \infty) \mathbb{1}(X_i = x)} \cdot \frac{\sum_i \mathbb{1}(t_i^* = s) \mathbb{1}(X_i = x)}{\sum_i \mathbb{1}(t_i^* = s)} \\
&= \frac{1}{N_s} \sum_i \Delta \bar{Y}_i^s(\lambda) \mathbb{1}(t_i^* = s) - \sum_{x \in \mathcal{X}} \frac{\sum_i \Delta \bar{Y}_i^s(\lambda) \mathbb{1}(t_i^* = \infty) \mathbb{1}(X_i = x)}{N_s} \cdot \frac{\hat{e}_s(x)}{\hat{e}_\infty(x)} \\
&= \frac{1}{N_s} \sum_i \Delta \bar{Y}_i^s(\lambda) \mathbb{1}(t_i^* = s) - \frac{1}{N_s} \sum_i \Delta \bar{Y}_i^s(\lambda) \mathbb{1}(t_i^* = \infty) \sum_{x \in \mathcal{X}} \mathbb{1}(X_i = x) \frac{\hat{e}_s(x)}{\hat{e}_\infty(x)} \\
&= \frac{1}{N_s} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{\hat{e}_s(X_i)}{\hat{e}_\infty(X_i)} \right) \Delta \bar{Y}_i^s(\lambda).
\end{aligned}$$

Following the proof of Theorem 3, the estimator can be rewritten as:

$$\begin{aligned}
\hat{\tau}(s, \lambda) &= \frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = s) (\mu(s, X_i) - \mu(0, X_i)) \\
&\quad + \frac{1}{N_s} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{\hat{e}_s(X_i)}{\hat{e}_\infty(X_i)} \right) (\Delta \bar{Y}_i^s(\lambda) - \mu_\lambda(t_i^*, X_i)) \\
&\quad + B_\lambda(s)
\end{aligned}$$

where in this case:

$$B_\lambda(s) = \frac{1}{N_s} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{\hat{e}_s(X_i)}{\hat{e}_\infty(X_i)} \right) \mu_\lambda(0, X_i).$$

When covariates are discrete, the matching can be done exactly, and thus the bias is zero. To see this,

$$\begin{aligned}
B_\lambda(s) &= \frac{1}{N_s} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{\hat{e}_s(X_i)}{\hat{e}_\infty(X_i)} \right) \mu_\lambda(0, X_i) \\
&= \sum_{x \in \mathcal{X}} \mu_\lambda(0, x) \frac{1}{N_s} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{\hat{e}_s(x)}{\hat{e}_\infty(x)} \right) \mathbb{1}(X_i = x) \\
&= \sum_{x \in \mathcal{X}} \mu_\lambda(0, x) \left(\frac{\sum_i \mathbb{1}(t_i^* = s) \mathbb{1}(X_i = x)}{\sum_i \mathbb{1}(t_i^* = s)} - \frac{\sum_i \mathbb{1}(t_i^* = s) \mathbb{1}(X_i = x)}{\sum_i \mathbb{1}(t_i^* = \infty) \mathbb{1}(X_i = x)} \cdot \frac{\sum_i \mathbb{1}(t_i^* = \infty) \mathbb{1}(X_i = x)}{\sum_i \mathbb{1}(t_i^* = s)} \right) \\
&= 0.
\end{aligned}$$

Thus,

$$\begin{aligned}\hat{\tau}(s, \lambda) - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda)|t_i^* = s] &= \frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = s)(\mu_\lambda(s, X_i) - \mu_\lambda(0, X_i) - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda)|t_i^* = s]) \\ &+ \frac{1}{N_s} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{e_s(X_i)}{e_\infty(X_i)} \right) (\Delta \bar{Y}_i^s(\lambda) - \mu_\lambda(t_i^*, X_i)) \\ &- \sum_{x \in \mathcal{X}} \left(\frac{\hat{e}_s(x)}{\hat{e}_\infty(x)} - \frac{e_s(x)}{e_\infty(x)} \right) \frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = \infty) \mathbb{1}(X_i = x) (\Delta \bar{Y}_i^s(\lambda) - \mu_\lambda(0, x)).\end{aligned}$$

Now,

$$\begin{aligned}\frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = \infty) \mathbb{1}(X_i = x) (\Delta \bar{Y}_i^s(\lambda) - \mu_\lambda(0, x)) &\rightarrow_{\mathbb{P}} \frac{e_\infty(x) p_x}{p_s} \mathbb{E} [\Delta \bar{Y}_i^s(\lambda) - \mu_\lambda(0, x) | t_i^* = \infty, X_i = x] \\ &= 0\end{aligned}$$

under finite conditional moments of the outcome whereas

$$\begin{aligned}\frac{\hat{e}_s(x)}{\hat{e}_\infty(x)} - \frac{e_s(x)}{e_\infty(x)} &= \frac{\hat{e}_s(x)e_\infty(x) - \hat{e}_\infty(x)e_s(x)}{\hat{e}_\infty(x)e_\infty(x)} \\ &= \frac{(\hat{e}_s(x) - e_s(x))e_\infty(x) - (\hat{e}_\infty(x) - e_\infty(x))e_s(x)}{\hat{e}_\infty(x)e_\infty(x)} \\ &= \frac{1}{\hat{e}_\infty(x)} (\hat{e}_s(x) - e_s(x)) - \frac{1}{\hat{e}_\infty(x)} \frac{e_s(x)}{e_\infty(x)} (\hat{e}_\infty(x) - e_\infty(x))\end{aligned}$$

and

$$\hat{e}_s(x) - e_s(x) = \frac{\sum_i (\mathbb{1}(t_i^* = s) - e_s(x)) \mathbb{1}(X_i = x)}{\sum_i \mathbb{1}(X_i = x)} = \frac{1}{n} \sum_i \frac{(\mathbb{1}(t_i^* = s) - e_s(x)) \mathbb{1}(X_i = x)}{\hat{p}(x)}$$

from which

$$\begin{aligned}\frac{\hat{e}_s(x)}{\hat{e}_\infty(x)} - \frac{e_s(x)}{e_\infty(x)} &= \frac{1}{\hat{e}_\infty(x) \hat{p}(x)} \cdot \frac{1}{n} \sum_i (\mathbb{1}(t_i^* = s) - e_s(x)) \mathbb{1}(X_i = x) \\ &- \frac{1}{\hat{e}_\infty(x) \hat{p}(x)} \frac{e_s(x)}{e_\infty(x)} \cdot \frac{1}{n} \sum_i (\mathbb{1}(t_i^* = \infty) - e_\infty(x)) \mathbb{1}(X_i = x) \\ &= O_{\mathbb{P}}(n^{-1/2})\end{aligned}$$

as long as $\hat{e}_s(x) > 0$ for all s, x . Therefore,

$$\sum_{x \in \mathcal{X}} \left(\frac{\hat{e}_s(x)}{\hat{e}_\infty(x)} - \frac{e_s(x)}{e_\infty(x)} \right) \frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = \infty) \mathbb{1}(X_i = x) (\Delta \bar{Y}_i^s(\lambda) - \mu_\lambda(0, x)) = o_{\mathbb{P}}(n^{-1/2})$$

and

$$\begin{aligned}\sqrt{n}(\hat{\tau}(s, \lambda) - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda)|t_i^* = s]) &= \frac{1}{\sqrt{n}} \sum_i \psi_i(s, \lambda) + o_{\mathbb{P}}(1) \\ &\rightarrow_{\mathcal{D}} \mathcal{N}(0, \mathbb{V}[\psi_i(s, \lambda)])\end{aligned}$$

where

$$\begin{aligned}\psi_i(s, \lambda) &= \frac{\mathbb{1}(t_i^* = s)}{p_s} (\mu_\lambda(s, X_i) - \mu_\lambda(0, X_i) - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda)|t_i^* = s]) \\ &\quad + \frac{1}{p_s} \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{e_s(X_i)}{e_\infty(X_i)} \right) (\Delta \bar{Y}_i^s(\lambda) - \mu_\lambda(t_i^*, X_i)).\end{aligned}$$

The variance can be expressed as:

$$\begin{aligned}\mathbb{V}[\psi_i(s, \lambda)] &= \frac{1}{p_s} \mathbb{E} \left[(\mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda)|t_i^* = s, X_i] - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda)|t_i^* = s])^2 | t_i^* = s \right] + \frac{1}{p_s} \mathbb{E}[\sigma_\lambda^2(s, X_i)|t_i^* = s] \\ &\quad + \frac{1}{p_s^2} \mathbb{E} \left[\frac{e_s(X_i)^2}{e_\infty(X_i)} \sigma^2(0, X_i) \right]\end{aligned}$$

as required. \square

Proof of Proposition 6

By Proposition 2, the estimator of interest can be written as the estimator from a regression of $\Delta Y_i^s(\lambda)$ on an intercept and a treated cohort indicator $\mathbb{1}(t_i^* = s)$. This corresponds to the case analyzed in [Abadie and Spiess \(2022\)](#) with $Z_i = (1, \mathbb{1}(t_i^* = s))'$. Under the conditions in the proposition's statement, Assumptions 1-5 in [Abadie and Spiess \(2022\)](#) hold, and thus their Proposition 2 apply to $\hat{\beta} = (\hat{\alpha}, \hat{\tau}_{\text{nr}}(\lambda))'$ where $\hat{\alpha}$ is the intercept from the regression. Thus,

$$\sqrt{(M+1)N_s}(\hat{\beta} - \beta) \rightarrow_{\mathcal{D}} \mathcal{N}(0, H^{-1} J H^{-1})$$

where in this case $\beta = (\mathbb{E}[\mu_\lambda(0, X_i)|t_i^* = s], \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda)|t_i^* = s])'$,

$$H = \frac{1}{1+M} \begin{bmatrix} 1+M & 1 \\ 1 & 1 \end{bmatrix}, \quad H^{-1} = \frac{1+M}{M} \begin{bmatrix} 1 & -1 \\ -1 & 1+M \end{bmatrix}$$

and $J = J_1 + J_2$ where

$$J_1 = \frac{1}{1+M} \mathbb{V} \left[\begin{array}{c} \mu_\lambda(s, X_i) + M\mu_\lambda(0, X_i) \\ \mu_\lambda(s, X_i) \end{array} \middle| t_i^* = s \right],$$

$$J_2 = \frac{1}{1+M} \mathbb{E} \left[\begin{array}{cc} \sigma_\lambda^2(s, X_i) + M\sigma_\lambda^2(0, X_i) & \sigma_\lambda^2(s, X_i) \\ \sigma_\lambda^2(s, X_i) & \sigma_\lambda^2(s, X_i) \end{array} \middle| t_i^* = s \right].$$

Then, by straightforward algebra,

$$V_{\text{nr}}(\lambda) = H^{-1} J H_{(2,2)}^{-1} = (1+M) \mathbb{E} \left[\left(\mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s, X_i] - \mathbb{E}[\bar{\tau}_i^{\text{post},s}(\lambda) | t_i^* = s] \right)^2 \middle| t_i^* = s \right]$$

$$+ (1+M) \mathbb{E} \left[\sigma_\lambda^2(s, X_i) + \frac{\sigma_\lambda^2(0, X_i)}{M} \middle| t_i^* = s \right]$$

as required. \square

D Proofs of Auxiliary Results

Proof of Lemma B1

Let $\ddot{D}_{it} = D_{it} - \bar{D}_i - \tilde{D}_t + \bar{D}$ and note that since $D_{it} = \mathbb{1}(t \geq t_i^*) = \sum_{s < \infty} \mathbb{1}(t_i^* = s) \mathbb{1}(t \geq s)$,

$$\bar{D}_i = \sum_{s < \infty} \mathbb{1}(t_i^* = s) \frac{\sum_t \lambda_t \mathbb{1}(t \geq s)}{\sum_t \lambda_t} = \sum_{s < \infty} \mathbb{1}(t_i^* = s) \Lambda_s$$

$$\tilde{D}_t = \sum_{s < \infty} \mathbb{1}(t \geq s) \frac{\sum_i w_i \mathbb{1}(t_i^* = s)}{\sum_i w_i} = \sum_{s < \infty} \mathbb{1}(t \geq s) \hat{p}_s^w$$

$$\bar{D} = \sum_{s < \infty} \hat{p}_s^w \Lambda_s$$

where $\Lambda_s = \sum_t \lambda_t \mathbb{1}(t \geq s) / \sum_t \lambda_t$ is the proportion of included post-periods $t \geq s$ and $\hat{p}_s^w = \sum_i w_i \mathbb{1}(t_i^* = s) / \sum_i w_i$ is the weighted proportion of cohort s in the sample. This implies that

$$\ddot{D}_{it} = \sum_{s < \infty} (\mathbb{1}(t_i^* = s) - \hat{p}_s^w) (\mathbb{1}(t \geq s) - \Lambda_s)$$

and

$$\begin{aligned}
D_{it}\ddot{D}_{it} &= \sum_{s<\infty} (\mathbb{1}(t_i^* = s) - \hat{p}_s^w)(\mathbb{1}(t \geq s) - \Lambda_s) \sum_{s'<\infty} \mathbb{1}(t_i^* = s')\mathbb{1}(t \geq s') \\
&= \sum_{s<\infty} (\mathbb{1}(t_i^* = s) - \hat{p}_s^w)(\mathbb{1}(t \geq s) - \Lambda_s)\mathbb{1}(t_i^* = s)\mathbb{1}(t \geq s) \\
&\quad + \sum_{s<\infty} \sum_{s' \neq s, \infty} (\mathbb{1}(t_i^* = s) - \hat{p}_s^w)(\mathbb{1}(t \geq s) - \Lambda_s)\mathbb{1}(t_i^* = s')\mathbb{1}(t \geq s') \\
&= \sum_{s<\infty} (1 - \hat{p}_s^w)(1 - \Lambda_s)\mathbb{1}(t_i^* = s)\mathbb{1}(t \geq s) \\
&\quad - \sum_{s<\infty} \sum_{s' \neq s, \infty} \hat{p}_s^w(\mathbb{1}(t \geq s) - \Lambda_s)\mathbb{1}(t_i^* = s')\mathbb{1}(t \geq s')
\end{aligned}$$

from which:

$$\begin{aligned}
\sum_i \sum_t w_i \lambda_t D_{it} \ddot{D}_{it} &= \sum_i \sum_t \sum_{s<\infty} w_i \lambda_t (1 - \hat{p}_s^w)(1 - \Lambda_s) \mathbb{1}(t_i^* = s) \mathbb{1}(t \geq s) \\
&\quad - \sum_i \sum_t \sum_{s<\infty} \sum_{s' \neq s, \infty} w_i \lambda_t \hat{p}_s^w (\mathbb{1}(t \geq s) - \Lambda_s) \mathbb{1}(t_i^* = s') \mathbb{1}(t \geq s') \\
&= \left(\sum_i \sum_t w_i \lambda_t \right) \left(\sum_{s<\infty} \hat{p}_s^w (1 - \hat{p}_s^w) \Lambda_s (1 - \Lambda_s) - \sum_{s<\infty} \sum_{s' \neq s, \infty} \hat{p}_s^w \hat{p}_{s'}^w (\Lambda_{s \vee s'} - \Lambda_s \Lambda_{s'}) \right)
\end{aligned}$$

where $s \vee s' = \max\{s, s'\}$. Using that $1 - \hat{p}_s^w = \sum_{s' \neq s} \hat{p}_{s'}^w = \hat{p}_\infty^w + \sum_{s' \neq s, \infty} \hat{p}_{s'}^w$, this term can be rewritten as:

$$\begin{aligned}
\sum_i \sum_t w_i \lambda_t D_{it} \ddot{D}_{it} &= \sum_{s<\infty} \hat{p}_s^w \left\{ \hat{p}_\infty^w \Lambda_s (1 - \Lambda_s) + \sum_{s' < s} \hat{p}_{s'}^w \Lambda_s (\Lambda_{s'} - \Lambda_s) + \sum_{s < s' < \infty} \hat{p}_{s'}^w (1 - \Lambda_s) (\Lambda_s - \Lambda_{s'}) \right\} \\
&\quad \times \left(\sum_i w_i \sum_t \lambda_t \right).
\end{aligned}$$

Next, let

$$\bar{Y}_i^{\text{post},s}(\lambda) = \frac{\sum_t \lambda_t \mathbb{1}(t \geq s) Y_{it}}{\sum_t \lambda_t \mathbb{1}(t \geq s)}, \quad \bar{Y}_i^{\text{pre},s}(\lambda) = \frac{\sum_t \lambda_t \mathbb{1}(t < s) Y_{it}}{\sum_t \lambda_t \mathbb{1}(t < s)}.$$

We have that:

$$\begin{aligned}
\sum_i \sum_t w_i \lambda_t Y_{it} \ddot{D}_{it} &= \sum_i \sum_t \sum_{s < \infty} w_i \lambda_t Y_{it} (\mathbb{1}(t_i^* = s) - \hat{p}_s^w) (\mathbb{1}(t \geq s) - \Lambda_s) \\
&= \sum_i \sum_{s < \infty} w_i (\mathbb{1}(t_i^* = s) - \hat{p}_s^w) \sum_t \lambda_t Y_{it} (\mathbb{1}(t \geq s) - \Lambda_s) \\
&= \sum_{s < \infty} \sum_i w_i (\mathbb{1}(t_i^* = s) - \hat{p}_s^w) \Lambda_s (1 - \Lambda_s) (\bar{Y}_i^{\text{post},s}(\lambda) - \bar{Y}_i^{\text{pre},s}(\lambda)) \left(\sum_t \lambda_t \right) \\
&= \sum_{s < \infty} (1 - \hat{p}_s^w) \Lambda_s (1 - \Lambda_s) \sum_i w_i \mathbb{1}(t_i^* = s) (\bar{Y}_i^{\text{post},s}(\lambda) - \bar{Y}_i^{\text{pre},s}(\lambda)) \left(\sum_t \lambda_t \right) \\
&\quad - \sum_{s < \infty} \hat{p}_s^w \Lambda_s (1 - \Lambda_s) \sum_i w_i \mathbb{1}(t_i^* \neq s) (\bar{Y}_i^{\text{post},s}(\lambda) - \bar{Y}_i^{\text{pre},s}(\lambda)) \left(\sum_t \lambda_t \right) \\
&= \sum_{s < \infty} \hat{p}_s^w (1 - \hat{p}_s^w) \Lambda_s (1 - \Lambda_s) \frac{\sum_i w_i \mathbb{1}(t_i^* = s) (\bar{Y}_i^{\text{post},s}(\lambda) - \bar{Y}_i^{\text{pre},s}(\lambda))}{\sum_i w_i \mathbb{1}(t_i^* = s)} \left(\sum_i \sum_t w_i \lambda_t \right) \\
&\quad - \sum_{s < \infty} \hat{p}_s^w (1 - \hat{p}_s^w) \Lambda_s (1 - \Lambda_s) \frac{\sum_i w_i \mathbb{1}(t_i^* \neq s) (\bar{Y}_i^{\text{post},s}(\lambda) - \bar{Y}_i^{\text{pre},s}(\lambda))}{\sum_i w_i \mathbb{1}(t_i^* \neq s)} \left(\sum_i \sum_t w_i \lambda_t \right) \\
&= \left(\sum_i \sum_t w_i \lambda_t \right) \sum_{s < \infty} \hat{p}_s^w (1 - \hat{p}_s^w) \Lambda_s (1 - \Lambda_s) \\
&\quad \times \left\{ \frac{\sum_i w_i \mathbb{1}(t_i^* = s) (\bar{Y}_i^{\text{post},s}(\lambda) - \bar{Y}_i^{\text{pre},s}(\lambda))}{\sum_i w_i \mathbb{1}(t_i^* = s)} - \frac{\sum_i w_i \mathbb{1}(t_i^* \neq s) (\bar{Y}_i^{\text{post},s}(\lambda) - \bar{Y}_i^{\text{pre},s}(\lambda))}{\sum_i w_i \mathbb{1}(t_i^* \neq s)} \right\} \\
&= \left(\sum_i \sum_t w_i \lambda_t \right) \sum_{s < \infty} \hat{p}_s^w (1 - \hat{p}_s^w) \Lambda_s (1 - \Lambda_s) \\
&\quad \times \left\{ \frac{\sum_i w_i \mathbb{1}(t_i^* = s) (\bar{Y}_i^{\text{post},s}(\lambda) - \bar{Y}_i^{\text{pre},s}(\lambda))}{\sum_i w_i \mathbb{1}(t_i^* = s)} \right. \\
&\quad \left. - \sum_{s' \neq s} \frac{\sum_i w_i \mathbb{1}(t_i^* = s') (\bar{Y}_i^{\text{post},s}(\lambda) - \bar{Y}_i^{\text{pre},s}(\lambda))}{\sum_i w_i \mathbb{1}(t_i^* = s')} \frac{\hat{p}_{s'}^w}{\sum_{s' \neq s} \hat{p}_{s'}^w} \right\}.
\end{aligned}$$

Defining:

$$\overline{\Delta Y}_i^s(\lambda) = \bar{Y}_i^{\text{post},s}(\lambda) - \bar{Y}_i^{\text{pre},s}(\lambda)$$

and using that $1 - \hat{p}_s^w = \sum_{s' \neq s} \hat{p}_{s'}^w$ we can rewrite this as:

$$\begin{aligned}
\sum_i \sum_t w_i \lambda_t Y_{it} \ddot{D}_{it} &= \sum_{s < \infty} \sum_{s' \neq s} \hat{p}_s^w \hat{p}_{s'}^w \Lambda_s (1 - \Lambda_s) \left\{ \frac{\sum_i w_i \mathbb{1}(t_i^* = s) \overline{\Delta Y}_i^s(\lambda)}{\sum_i w_i \mathbb{1}(t_i^* = s)} - \frac{\sum_i w_i \mathbb{1}(t_i^* = s') \overline{\Delta Y}_i^{s'}(\lambda)}{\sum_i w_i \mathbb{1}(t_i^* = s')} \right\} \\
&\quad \times \left(\sum_i \sum_t w_i \lambda_t \right).
\end{aligned}$$

Finally note that:

$$\begin{aligned}
\overline{\Delta Y}_i^s(\lambda) &= \frac{\sum_t \lambda_t Y_{it} \mathbb{1}(t \geq s)}{\sum_t \lambda_t \mathbb{1}(t \geq s)} - \bar{Y}_i^{\text{pre},s}(\lambda) = \frac{\sum_t \lambda_t (Y_{it} - \bar{Y}_i^{\text{pre},s}(\lambda)) \mathbb{1}(t \geq s)}{\sum_t \lambda_t \mathbb{1}(t \geq s)} \\
&= \frac{\sum_t \lambda_t \left(Y_{it} - \frac{\sum_{t' < s} \lambda_{t'} Y_{it'} \mathbb{1}(t' < s)}{\sum_{t' < s} \lambda_{t'} \mathbb{1}(t' < s)} \right) \mathbb{1}(t \geq s)}{\sum_t \lambda_t \mathbb{1}(t \geq s)} \\
&= \frac{\sum_t \sum_{t' < s} \lambda_t \lambda_{t'} (Y_{it} - Y_{it'}) \mathbb{1}(t \geq s) \mathbb{1}(t' < s)}{\sum_t \lambda_t \mathbb{1}(t \geq s) \sum_{t' < s} \lambda_{t'} \mathbb{1}(t' < s)}
\end{aligned}$$

and thus

$$\frac{\sum_i w_i \mathbb{1}(t_i^* = s) \overline{\Delta Y}_i^s(\lambda)}{\sum_i w_i \mathbb{1}(t_i^* = s)} = \frac{\sum_t \sum_{t' < s} \lambda_t \lambda_{t'} \frac{\sum_i w_i \mathbb{1}(t_i^* = s) (Y_{it} - Y_{it'}) \mathbb{1}(t \geq s) \mathbb{1}(t' < s)}{\sum_i w_i \mathbb{1}(t_i^* = s)}}{\sum_t \lambda_t \mathbb{1}(t \geq s) \sum_{t' < s} \lambda_{t'} \mathbb{1}(t' < s)}$$

from which

$$\frac{\sum_i w_i \mathbb{1}(t_i^* = s) \overline{\Delta Y}_i^s(\lambda)}{\sum_i w_i \mathbb{1}(t_i^* = s)} - \frac{\sum_i w_i \mathbb{1}(t_i^* = s') \overline{\Delta Y}_i^{s'}(\lambda)}{\sum_i w_i \mathbb{1}(t_i^* = s')} = \frac{\sum_t \sum_{t' < s} \lambda_t \lambda_{t'} \hat{\tau}_{s'}^s(t, t') \mathbb{1}(t \geq s) \mathbb{1}(t' < s)}{\sum_t \lambda_t \mathbb{1}(t \geq s) \sum_{t' < s} \lambda_{t'} \mathbb{1}(t' < s)}$$

where

$$\hat{\tau}_{s'}^s(t, t') = \frac{\sum_i w_i \mathbb{1}(t_i^* = s) (Y_{it} - Y_{it'})}{\sum_i w_i \mathbb{1}(t_i^* = s)} - \frac{\sum_i w_i \mathbb{1}(t_i^* = s') (Y_{it} - Y_{it'})}{\sum_i w_i \mathbb{1}(t_i^* = s')}.$$

Therefore,

$$\begin{aligned}
\sum_i \sum_t w_i \lambda_t Y_{it} \ddot{D}_{it} &= \sum_{s < \infty} \sum_{s' \neq s} \sum_{t \geq s} \sum_{t' < s} \frac{\lambda_t \lambda_{t'} \hat{p}_s^w \hat{p}_{s'}^w \Lambda_s (1 - \Lambda_s)}{\sum_t \lambda_t \mathbb{1}(t \geq s) \sum_{t' < s} \lambda_{t'} \mathbb{1}(t' < s)} \hat{\tau}_{s'}^s(t, t') \\
&\quad \times \left(\sum_i \sum_t w_i \lambda_t \right) \\
&= \left(\frac{\sum_i w_i}{\sum_t \lambda_t} \right) \sum_{s < \infty} \sum_{s' \neq s} \sum_{t \geq s} \sum_{t' < s} \lambda_t \lambda_{t'} \hat{p}_s^w \hat{p}_{s'}^w \hat{\tau}_{s'}^s(t, t').
\end{aligned}$$

Collecting the results, we have that:

$$\hat{\tau}(w, \lambda) = \frac{\hat{\tau}_{\text{num}}(w, \lambda)}{\hat{\tau}_{\text{den}}(w, \lambda)}$$

where

$$\begin{aligned}
\hat{\tau}_{\text{num}}(w, \lambda) &= \sum_{s < \infty} \sum_{s' \neq s} \sum_{t \geq s} \sum_{t' < s} \lambda_t \lambda_{t'} \hat{p}_s^w \hat{p}_{s'}^w \hat{\tau}_{s'}^s(t, t') \\
\hat{\tau}_{\text{den}}(w, \lambda) &= \left(\sum_t \lambda_t \right)^2 \sum_{s < \infty} \hat{p}_s^w \left\{ \hat{p}_\infty^w \Lambda_s (1 - \Lambda_s) + \sum_{s' < s} \hat{p}_{s'}^w \Lambda_s (\Lambda_{s'} - \Lambda_s) + \sum_{s < s' < \infty} \hat{p}_{s'}^w (1 - \Lambda_s) (\Lambda_s - \Lambda_{s'}) \right\}.
\end{aligned}$$

which completes the proof. \square

Proof of Lemma B2

Let $\mathbf{t}^* = (t_1^*, \dots, t_n^*)$ and $\mathbf{X} = (X_1', \dots, X_n')'$.

$$\begin{aligned} \frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = s') \frac{K_M(i, s)}{M} Z_i &= \frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = s) Z_i - \frac{1}{N_s} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = s') \frac{K_M(i, s)}{M} \right) Z_i \\ &= \mathbb{E}[Z_i | t_i^* = s] + o_{\mathbb{P}}(1) - T_n \end{aligned}$$

where

$$T_n = \frac{1}{N_s} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = s') \frac{K_M(i, s)}{M} \right) Z_i.$$

Now,

$$\begin{aligned} T_n &= \frac{n}{N_s} \left\{ \frac{1}{n} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = s') \frac{K_M(i, s)}{M} \right) (Z_i - \mu_Z(t_i^*, X_i)) \right. \\ &\quad + \frac{1}{n} \sum_i \mathbb{1}(t_i^* = s) (\mu_Z(s, X_i) - \mu_Z(s', X_i)) \\ &\quad \left. + \frac{1}{n} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = s') \frac{K_M(i, s)}{M} \right) \mu_Z(s', X_i) \right\} \\ &:= \frac{n}{N_s} (T_{1,n} + T_{2,n} + T_{3,n}). \end{aligned}$$

We look at these three terms separately. First, $\mathbb{E}[T_{1,n} | \mathbf{t}^*, \mathbf{X}] = 0$ and

$$\mathbb{V}[T_{1,n} | \mathbf{t}^*, \mathbf{X}] = \frac{1}{n^2} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = s') \frac{K_M(i, s)}{M} \right)^2 \sigma_Z^2(t_i^*, X_i)$$

from which

$$\begin{aligned} \mathbb{V}[T_{1,n}] &= \frac{1}{n} \mathbb{E} \left[\left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = s') \frac{K_M(i, s)}{M} \right)^2 \sigma_Z^2(t_i^*, X_i) \right] \\ &\leq \frac{C}{n} \left(1 + \frac{1}{M^2} \mathbb{E} [K_M(i, s)^2] \right) \\ &\rightarrow 0 \end{aligned}$$

because by Lemma 3 in [Abadie and Imbens \(2006\)](#), $\mathbb{E}[K_M(i, s)^2]$ is uniformly bounded. Thus, $T_{1,n} = o_{\mathbb{P}}(1)$. Next,

$$\begin{aligned} T_{2,n} &= \frac{1}{n} \sum_i \mathbb{1}(t_i^* = s) (\mu_Z(s, X_i) - \mu_Z(s', X_i)) \\ &\rightarrow_{\mathbb{P}} \mathbb{E} [\mu_Z(s, X_i) - \mu_Z(s', X_i) | t_i^* = s] p_s \\ &= (\mathbb{E}[Z_i | t_i^* = s] - \mathbb{E}[\mu_Z(s', X_i) | t_i^* = s]) p_s. \end{aligned}$$

Finally, for the third term,

$$\begin{aligned} T_{3,n} &= \frac{1}{n} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = s') \frac{K_M(i, s)}{M} \right) \mu_Z(s', X_i) \\ &= \frac{1}{n} \sum_i \mathbb{1}(t_i^* = s) \frac{1}{M} \sum_{m=1}^M (\mu_Z(s', X_i) - \mu_Z(s', X_{j_m(i, s')})) \end{aligned}$$

By a Taylor expansion,

$$T_{3,n} = \frac{1}{n} \sum_i \mathbb{1}(t_i^* = s) \frac{1}{M} \sum_{m=1}^M \dot{\mu}_Z(s', \tilde{X}) \|X_i - X_{j_m(i, s')}\|$$

where $\dot{\mu}_Z(s, x) = \partial \mu_Z(s, x) / \partial x$ and \tilde{X} is a (random) intermediate point between X_i and $X_{j_m(i)}$. Note that by continuity of $\mu_Z(s, x)$ and compactness, the derivative is uniformly bounded and thus

$$T_{3,n} \leq \frac{1}{n} \sum_i \mathbb{1}(t_i^* = s) \frac{C}{M} \sum_{m=1}^M \|X_i - X_{j_m(i, s')}\| \leq \frac{C}{n} \sum_i \|X_i - X_{j_M(i, s')}\|.$$

Finally, note that

$$\begin{aligned} \mathbb{E} \left[\left(\frac{1}{n} \sum_i \|X_i - X_{j_M(i, s')}\| \right)^2 \right] &= \frac{1}{n^2} \sum_i \mathbb{E} \left[\|X_i - X_{j_M(i, s')}\|^2 \right] \\ &\quad + \frac{2}{n^2} \sum_i \sum_{l>i} \mathbb{E} \left[\|X_i - X_{j_M(i, s')}\| \|X_l - X_{j_M(l, s')}\| \right] \\ &= \frac{1}{n} \mathbb{E} \left[\|X_i - X_{j_M(i, s')}\|^2 \right] + \frac{n(n-1)}{n^2} \mathbb{E} \left[\|X_i - X_{j_M(i, s')}\| \|X_l - X_{j_M(l)}\| \right] \\ &\leq \mathbb{E} \left[\|X_i - X_{j_M(i, s')}\|^2 \right] \end{aligned}$$

by identical distributions and Cauchy-Schwarz. Thus,

$$T_{3,n} \leq \mathbb{E} \left[\|X_i - X_{j_M(i, s')}\|^2 \right] = \frac{1}{n^{1/q}} \mathbb{E} \left[n^{1/q} \|X_i - X_{j_M(i, s')}\|^2 \right] \rightarrow 0$$

by Lemma 2 in [Abadie and Imbens \(2006\)](#). Thus $T_{3,n} = o_{\mathbb{P}}(1)$. Therefore,

$$T_n = \frac{n}{N_s} (T_{1,n} + T_{2,n} + T_{3,n}) \rightarrow_{\mathbb{P}} \mathbb{E}[Z_i | t_i^* = s] - \mathbb{E}[\mu_Z(s', X_i) | t_i^* = s]$$

and

$$\frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = s') \frac{K_M(i, s)}{M} Z_i \rightarrow_{\mathbb{P}} \mathbb{E}[\mu_Z(s', X_i) | t_i^* = s]$$

as required. For the second result, Let $\mathbf{X}_{s'} = (X_i)_{i:t_i^*=s'}$ be the subset of covariates in the comparison cohort $t_i^* = s'$ and let $m(s', x) = \mathbb{E}[Z_i^2 | t_i^* = s', X_i = x]$. We have that

$$\frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = s') \frac{K_M(i)^2}{M^2} Z_i^2 = \frac{n}{N_s} \cdot \frac{1}{M^2} \cdot T_n, \quad T_n = \frac{1}{n} \sum_i \mathbb{1}(t_i^* = s') K_M(i, s)^2 Z_i^2$$

and

$$\begin{aligned} \mathbb{E}[T_n] &= \mathbb{E}[\mathbb{1}(t_i^* = s') K_M(i, s)^2 Z_i^2] = \mathbb{E}[\mathbb{1}(t_i^* = s') K_M(i, s)^2 m(s', X_i)] \\ &= \mathbb{E}[\mathbb{E}[K_M(i, s)^2 | \mathbf{t}^*, \mathbf{X}_{s'}, t_i^* = s'] m(s', X_i) | t_i^* = s'] p_{s'} \end{aligned}$$

Now,

$$K_M(i, s) | \mathbf{t}^*, \mathbf{X}_{s'}, t_i^* = s' \sim \text{Binomial}(N_s, \nu_s(\mathcal{A}_M(X_i)))$$

where $\nu_s(\cdot)$ is the conditional distribution of $X | t^* = s$ and $\mathcal{A}_M(x)$ is the catchment area, see [Abadie and Imbens \(2006\)](#) and [Chen and Han \(2024\)](#) for details. This implies that

$$\mathbb{E}[K_M(i, s)^2 | \mathbf{t}^*, \mathbf{X}_{s'}, t_i^* = s'] = N_s \nu_s(\mathcal{A}_M(X_i)) + N_s(N_s - 1) \nu_s(\mathcal{A}_M(X_i))^2$$

and thus

$$\begin{aligned} \mathbb{E}[K_M(i, s)^2 m(s', X_i) | t_i^* = s'] &= \mathbb{E}[N_s \nu_s(\mathcal{A}_M(X_i)) m(s', X_i) | t_i^* = s'] \\ &\quad + \mathbb{E}[N_s(N_s - 1) \nu_s(\mathcal{A}_M(X_i))^2 m(s', X_i) | t_i^* = s']. \end{aligned}$$

For the first term, note that:

$$\begin{aligned} \mathbb{E}[N_s \nu_s(\mathcal{A}_M(X_i)) m(s', X_i) | t_i^* = s'] &= \mathbb{E}\left[\frac{N_s}{N_{s'}} N_{s'} \mathbb{E}[\nu_s(\mathcal{A}_M(X_i)) | t_i^* = s', X_i] m(s', X_i) | t_i^* = s'\right] \\ &= \frac{p_s}{p_{s'}} \mathbb{E}[N_{s'} \mathbb{E}[\nu_s(\mathcal{A}_M(X_i)) | t_i^* = s', X_i] m(s', X_i) | t_i^* = s'] \\ &\quad + \mathbb{E}\left[\left(\frac{N_s}{N_{s'}} - \frac{p_s}{p_{s'}}\right) N_{s'} \mathbb{E}[\nu_s(\mathcal{A}_M(X_i)) | t_i^* = s', X_i] m(s', X_i) | t_i^* = s'\right]. \end{aligned}$$

By Assumption 5(1) and Lemma 5.3 in [Chen and Han \(2024\)](#), $N_{s'}\mathbb{E}[\nu_s(\mathcal{A}_M(X_i))|t_i^* = s', X_i]m(s', X_i)$ is uniformly bounded. Then,

$$\mathbb{E} \left[\left(\frac{N_s}{N_{s'}} - \frac{p_s}{p_{s'}} \right) N_{s'} \mathbb{E}[\nu_s(\mathcal{A}_M(X_i))|t_i^* = s', X_i] m(s', X_i) | t_i^* = s' \right] \leq C \mathbb{E} \left[\left| \frac{N_s}{N_{s'}} - \frac{p_s}{p_{s'}} \right| \right].$$

Now,

$$\begin{aligned} \mathbb{E} \left[\left| \frac{N_s}{N_{s'}} - \frac{p_s}{p_{s'}} \right| \right] &= \mathbb{E} \left[\left| \frac{N_s}{N_{s'}} - \frac{p_s}{p_{s'}} \right| \mathbb{1} \left(\left| \frac{N_s}{N_{s'}} - \frac{p_s}{p_{s'}} \right| > \delta \right) \right] \\ &\quad + \mathbb{E} \left[\left| \frac{N_s}{N_{s'}} - \frac{p_s}{p_{s'}} \right| \mathbb{1} \left(\left| \frac{N_s}{N_{s'}} - \frac{p_s}{p_{s'}} \right| \leq \delta \right) \right] \\ &\leq \mathbb{E} \left[\left| \frac{N_s}{N_{s'}} - \frac{p_s}{p_{s'}} \right| \mathbb{1} \left(\left| \frac{N_s}{N_{s'}} - \frac{p_s}{p_{s'}} \right| > \delta \right) \right] + \delta \\ &\leq \mathbb{E} \left[\left| \frac{N_s}{N_{s'}} - \frac{p_s}{p_{s'}} \right|^2 \right]^{1/2} \mathbb{P} \left[\left| \frac{N_s}{N_{s'}} - \frac{p_s}{p_{s'}} \right| > \delta \right]^{1/2} + \delta \\ &\rightarrow \delta \end{aligned}$$

because $N_s/N_{s'} \rightarrow_{\mathbb{P}} p_s/p_{s'}$ and $\mathbb{E}[|N_s/N_{s'} - p_s/p_{s'}|^2]$ is bounded by Lemma S3 in [Abadie and Imbens \(2016\)](#). Because δ is arbitrary, it follows that:

$$\mathbb{E} \left[\left(\frac{N_s}{N_{s'}} - \frac{p_s}{p_{s'}} \right) N_{s'} \mathbb{E}[\nu_s(\mathcal{A}_M(X_i))|t_i^* = s', X_i] m(s', X_i) | t_i^* = s' \right] \rightarrow 0$$

and thus

$$\begin{aligned} \mathbb{E} [N_s \nu_s(\mathcal{A}_M(X_i)) \sigma^2(s', X_i) | t_i^* = s'] &= \frac{p_s}{p_{s'}} \mathbb{E} [N_{s'} \mathbb{E}[\nu_s(\mathcal{A}_M(X_i)) | t_i^* = s', X_i] m(s', X_i) | t_i^* = s'] + o(1) \\ &\rightarrow \frac{p_s}{p_{s'}} \mathbb{E} \left[M \frac{f_s(X_i)}{f_{s'}(X_i)} m(s', X_i) \middle| t_i^* = s' \right] \end{aligned}$$

by Theorem 4.1 and Lemma 5.3 in [Chen and Han \(2024\)](#) and the dominated convergence theorem. Also note that:

$$\frac{f_s(X_i)}{f_{s'}(X_i)} = \frac{e_s(X_i)/p_s}{e_{s'}(X_i)/p_{s'}}$$

from which

$$\begin{aligned} \mathbb{E} [N_s \nu_s(\mathcal{A}_M(X_i)) m(s', X_i) | t_i^* = s'] &\rightarrow_{\mathbb{P}} M \mathbb{E} \left[\frac{e_s(X_i)}{e_{s'}(X_i)} m(s', X_i) \middle| t_i^* = s' \right] \\ &= \frac{M}{p_{s'}} \mathbb{E} [e_s(X_i) m(s', X_i)]. \end{aligned}$$

Similarly, for the second term,

$$\begin{aligned} \mathbb{E} [N_s(N_s - 1)\nu_s(\mathcal{A}_M(X_i))^2 m(s', X_i) | t_i^* = s'] &= \left(\frac{p_s}{p_{s'}}\right)^2 \mathbb{E} [N_{s'}^2 \mathbb{E}[\nu_s(\mathcal{A}_M(X_i))^2 | t_i^* = s', X_i] m(s', X_i) | t_i^* = s'] \\ &+ \mathbb{E} \left[\left(\frac{N_s(N_s - 1)}{N_{s'}^2} - \left(\frac{p_s}{p_{s'}}\right)^2 \right) N_{s'}^2 \mathbb{E}[\nu_s(\mathcal{A}_M(X_i))^2 | t_i^* = s', X_i] m(s', X_i) | t_i^* = s' \right]. \end{aligned}$$

By Theorem 4.1 and Lemma 5.3 in [Chen and Han \(2024\)](#),

$$\begin{aligned} \mathbb{E} [N_{s'}^2 \mathbb{E}[\nu_s(\mathcal{A}_M(X_i))^2 | t_i^* = s', X_i] m(s', X_i) | t_i^* = s'] &\rightarrow \alpha(M, q) \mathbb{E} \left[\left(\frac{f_s(X_i)}{f_{s'}(X_i)} \right)^2 m(s', X_i) \middle| t_i^* = s' \right] \\ &= \alpha(M, q) \left(\frac{p_{s'}}{p_s^2} \right) \mathbb{E} \left[\frac{e_s(X_i)^2}{e_{s'}(X_i)} m(s', X_i) \right] \end{aligned}$$

On the other hand, by previous arguments

$$\mathbb{E} \left[\left(\frac{N_s(N_s - 1)}{N_{s'}^2} - \left(\frac{p_s}{p_{s'}}\right)^2 \right) N_{s'}^2 \mathbb{E}[\nu_s(\mathcal{A}_M(X_i))^2 | t_i^* = s', X_i] m(s', X_i) | t_i^* = s' \right] \rightarrow 0.$$

Thus

$$\mathbb{E} [N_s(N_s - 1)\nu_s(\mathcal{A}_M(X_i))^2 m(s', X_i) | t_i^* = s'] \rightarrow_{\mathbb{P}} \frac{\alpha(M, q)}{p_{s'}} \mathbb{E} \left[\frac{e_s(X_i)^2}{e_{s'}(X_i)} m(s', X_i) \right].$$

and therefore

$$\mathbb{E} [K_M(i, s)^2 m(s', X_i) | t_i^* = s'] \rightarrow_{\mathbb{P}} \frac{M}{p_{s'}} \mathbb{E}[e_s(X_i) m(s', X_i)] + \frac{\alpha(M, q)}{p_{s'}} \mathbb{E} \left[\frac{e_s(X_i)^2}{e_{s'}(X_i)} m(s', X_i) \right].$$

Thus,

$$\mathbb{E}[T_n] \rightarrow_{\mathbb{P}} M \mathbb{E}[e_s(X_i) m(s', X_i)] + \alpha(M, q) \mathbb{E} \left[\frac{e_s(X_i)^2}{e_{s'}(X_i)} m(s', X_i) \right].$$

Finally, note that:

$$\begin{aligned}
\mathbb{V}[T_n] &= \frac{1}{n} \mathbb{V} [\mathbb{1}(t_i^* = s') K_M(i, s)^2 m(s', X_i)] + \frac{1}{n} \mathbb{E} [\mathbb{1}(t_i^* = s') K_M(i, s)^4 \mathbb{V}[Z_i^2 | t_i^* = s', X_i]] \\
&\quad + \frac{2}{n^2} \sum_i \sum_{j>i} \text{Cov} (\mathbb{1}(t_i^* = s') K_M(i, s)^2 m(s', X_i), \mathbb{1}(t_j^* = s') K_M(j, s)^2 m(s', X_j)) \\
&\leq \frac{1}{n} \mathbb{V} [\mathbb{1}(t_i^* = s') K_M(i, s)^2 m(s', X_i)] + \frac{1}{n} \mathbb{E} [\mathbb{1}(t_i^* = s') K_M(i, s)^4 \mathbb{V}[Z_i^2 | t_i^* = s', X_i]] \\
&\leq \frac{C}{n} \mathbb{E} [K_M(i, s)^4] \\
&\rightarrow 0
\end{aligned}$$

where the first inequality follows from the fact that $K_M(i, s)$ and $K_M(j, s)$ are negatively correlated ([Abadie and Imbens, 2002](#), p.40), the second inequality follows because the conditional variance is uniformly bounded and the convergence follows from the fact that $\mathbb{E}[K_M(i, s)^4]$ is bounded by Lemma 3 in [Abadie and Imbens \(2006\)](#). This implies that

$$T_n \xrightarrow{\mathbb{P}} M \mathbb{E}[e_s(X_i) m(s', X_i)] + \alpha(M, q) \mathbb{E} \left[\frac{e_s(X_i)^2}{e_{s'}(X_i)} m(s', X_i) \right]$$

and thus

$$\begin{aligned}
\frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = s') \frac{K_M(i, s)^2}{M^2} Z_i^2 &= \frac{n}{N_s} \cdot \frac{1}{M^2} \cdot T_n \\
&\xrightarrow{\mathbb{P}} \frac{1}{M} \mathbb{E} \left[e_s(X_i) \frac{m(s', X_i)}{p_s} \right] + \frac{\alpha(M, q)}{M^2} \mathbb{E} \left[\frac{e_s(X_i)^2 m(s', X_i)}{e_{s'}(X_i) p_s} \right]
\end{aligned}$$

which completes the proof. \square