Post-Matching Two-Way Fixed Effects Estimation*

Yihong Liu† Gonzalo Vazquez-Bare‡

October 23, 2026

Abstract

Identification in difference-in-differences (DiD) settings is often justified under a conditional parallel trends assumption, under which units with similar observed covariates are expected to have followed similar outcome trends in the absence of treatment. In such settings, researchers often combine matching with two-way fixed effects (2WFE) methods. Specifically, in a first step, each treated unit is matched to one or more never-treated units based on observed pre-treatment covariates. In the second step, treatment effects are estimated with a 2WFE regression in the matched sample, reweighting the never-treated units by the number of times they are matched. We formally analyze this common practice and highlight three problems. First, with variation in treatment timing, the post-matching 2WFE estimator that pools all treated cohorts has an asymptotic "pooling" 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. Third, imperfect matches generate a matching bias that can be first-order for inference, but the required bias correction to address it is commonly omitted when using post-matching 2WFE estimators in empirical work. Thus, conventional t-statistics and confidence intervals based on the uncorrected estimator may be invalid. We provide consistent, asymptotically normal bias-corrected matched DiD estimators of treatment effects, event-study coefficients and other aggregate magnitudes, with consistent standard errors that account for the matching step. We illustrate our results with simulations and with two empirical applications.

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

Identification in difference-in-differences (DiD) settings is often justified under a conditional parallel trends assumption, under which units that are similar in their observed covariates are assumed to have followed the same outcome trends in the absence of treatment. In such settings, researchers often use a combination of matching and two-way fixed effects (2WFE) methods, matching each treated unit to one or more untreated units using pre-treatment observed covariates first, and then estimating a 2WFE specification reweighting the untreated units by the number of times they are used as a match, excluding untreated units that are not matched to any treated unit.

To illustrate the motivation for this strategy, Figure 1 compares the estimates from an unconditional event study without covariates to those from an event study combined with matching.[1] As the figure shows, the pre-trends coefficients are much smaller and become jointly insignificant after matching, which provides additional support for the plausibility of conditional parallel trends assumption.

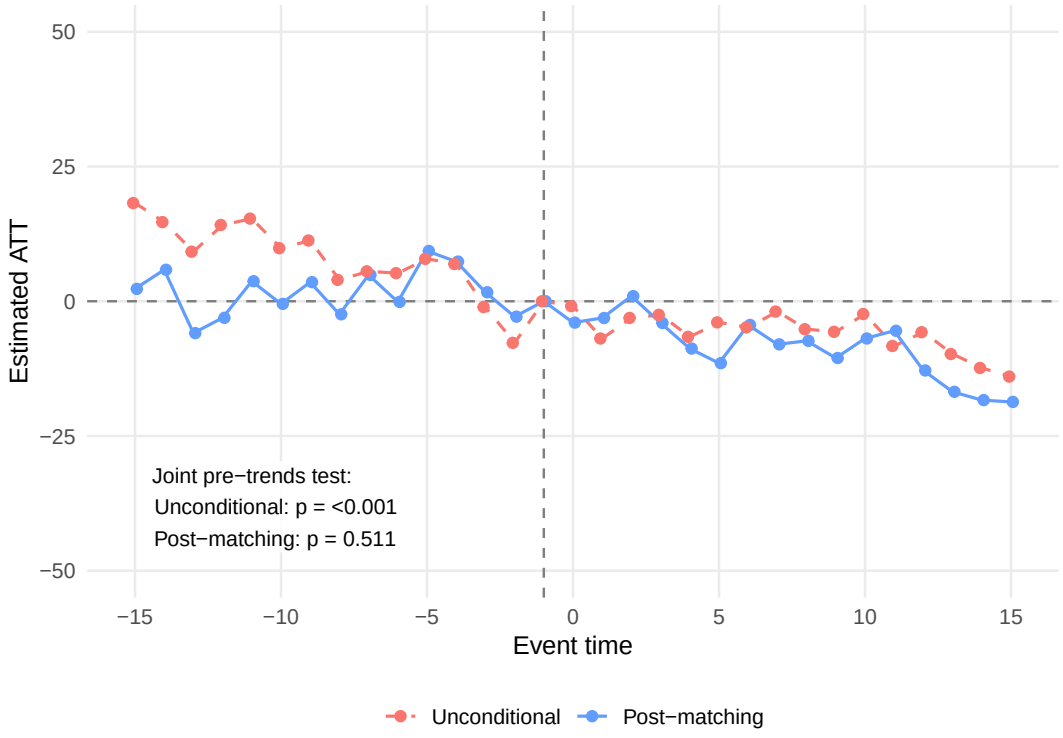


Figure 1: Comparison of unconditional and post-matching event-study estimates

While several alternative methods are available for estimating treatment effects under a conditional parallel trends assumption, matching-based methods have the advantages of being intuitively appealing, easy to implement, computationally stable and fully nonparametric, not requiring the

[1] These estimates are computed using data from Bailey and Goodman-Bacon (2015) and measure the effect of the construction of Community Health Centers on county-level mortality. See Section 6 for details.

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. Despite its advantages and popularity in empirical practice, this method has received little attention by the methodological and theoretical literature.

In this paper, we formally analyze this commonly employed strategy and highlight three problems that affect point estimation and inference. First, when different treatment cohorts enter treatment in different time periods, the post-matching 2WFE estimator has 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 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 cohort-specific ATTs 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 are invalid, even when the comparison units are not yet treated.

Second, the first-step matching procedure introduces variability that needs to be accounted for when conducting inference, and failing to account for this additional variability results in generally inconsistent 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.

Third, imperfect matches generate a matching bias that can be first-order for inference. While this fact is well-known in the theoretical matching literature and can be addressed using bias correction, this issue is commonly omitted when using post-matching 2WFE estimators in empirical work. Thus, conventional t-statistics and confidence intervals based on the uncorrected estimator may be invalid.

To address these issues, we characterize the joint asymptotic distribution of the vector of cohort means over time and their matched counterparts. We show that under the conditional parallel trends assumption and mild regularity conditions, our results can be used to construct consistent, asymptotically normal bias-corrected estimators of treatment effects, event-study coefficients and other aggregate magnitudes, with valid standard errors that account for the matching step. We rely on recent advancements on measures of high-order Voronoi cells by Chen and Han (2026) to provide a closed-form representation of the true asymptotic variance matrix of our proposed

estimators, a result previously unavailable in the matching literature, and propose consistent variance estimators. Our results also allow us to compare the true asymptotic variance to the semiparametric efficiency bound for a general dimension of the covariate vector. We show that the post-matching DiD estimator does not attain the semiparametric efficiency bound for a fixed number of matches, and characterize the resulting efficiency gap as a function of M and q.

We illustrate our results with simulations and with two empirical applications. The first one is 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. Our second application uses data from Bailey and Goodman-Bacon (2015) to measure the effect of the construction of Community Health Centers on county-level mortality in the US. Our results show how matching successfully makes the pre-treatment trends more similar between cohorts compared to an unconditional event study designs, and increases the magnitude of the overall estimated ATT by about 78% (from -3.03 to -5.40), although the effect is statistically insignificant at the 5% level.

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'Haultfœuille, 2023; 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 a single post-matching 2WFE regression creates an additional bias that has not been studied in the literature. Specifically, our Theorem B1 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 Proposition 1 and Theorem 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. Theorem 2 also allows to construct estimators, standard errors and confidence intervals for event-study-type estimators and aggregate magnitudes. In addition, our closed-form characterization of the asymptotic variance based on the results by Chen and Han (2026) 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 with a fixed number of matches, and we describe this efficiency gap as a function of M and q.

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 our proposed 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 and from Bailey and Goodman-Bacon (2015). 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 $i = 1, \ldots, n$ over T time periods $t = 1, 2, \ldots, T$. Let t_i^* be a random variable with support $\mathcal{S} = \{s \in \{2, \ldots, T, \infty\} : \mathbb{P}[t_i^* = s] > 0\}$ that indicates the period in which unit i receives the treatment for the first time. The values of t_i^* identify different treatment cohorts. Following the DiD literature, 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 $\mathbf{d}_{t:t'} = (d_t, d_{t+1}, \ldots, d_{t'})$ 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 restrictions, the potential outcome for each unit in each time period may be a function of the whole treatment path, $Y_{it}(\mathbf{d}_T)$, with observed outcomes denoted by $Y_{it} = Y_{it}(D_{i1}, \ldots, D_{iT})$. We introduce the following assumption, which is standard in the DiD literature, that 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 in each treated period, $\mathbb{E}[\tau_{it} | t_i^* = s]$, $t \geq s$.

Let X_i be a time-invariant vector of covariates of dimension $q \geq 1$, measured at a pure pre-treatment period $t < \min\{\mathcal{S} \setminus \{\infty\}\}$. We assume that the parallel trends assumption holds conditional on covariates.

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

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

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 to guarantee a non-zero mass of never treated at each of the possible values of the covariates for the treated cohorts.

Assumption 3 (Overlap) *There exists $\eta > 0$ such that, for every $s \in \mathcal{S} \setminus \{\infty\}$, $e_\infty(X_i) \geq \eta$ a.s. on the support of $X_i | t_i^* = s$.*

Under these assumptions, the ATT for cohort s in each period $t \geq s$ is identified by comparing the observed evolution of outcomes over two periods between cohort s and another group that remains untreated in both periods, after adjusting for differences in covariates. This result is standard in the literature, and we state it as a lemma below for reference. We slightly modify Assumption 3 to allow for a comparison group that is not necessarily the never treated; this same idea can be used throughout the paper to allow for different comparison groups, as discussed further below.

Lemma 1 *Under Assumptions 1 and 2, for any $t' < s \leq t < s'$ such that $e_{s'}(X_i) > 0$ a.s. on the support of $X_i | t_i^* = s$,*

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

3 Post-matching 2WFE estimation

Consider a sample $(Y_{i1}, \ldots, Y_{iT}, t_i^*, X_i')_{i=1}^n$. 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.[2]

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

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 $j \in \{1, \ldots, n\}$ of the unit in comparison cohort $s' \neq t_i^*$ such that:

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

[2]Throughout the paper, we implicitly assume that $N_s > 0$ for all s and that $N_{s'} \geq M$ whenever cohort s' is used as a comparison group. These conditions hold with probability approaching one.

that is, the index of the m-th closest unit to i in covariate values among the comparison cohort s'. Let $\mathcal{J}_M(i,s') = \{j_1(i,s'), j_2(i,s'), \ldots, j_M(i,s')\}$ 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:

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

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

Once the matching procedure 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 $w = (w_1, \ldots, w_n)$. For each time period $t = 1, \ldots, T$, 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 $\lambda = (\lambda_1, \ldots, \lambda_T)$ 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:[3]

$$Y_{it} = \alpha_i + \delta_t + \tau D_{it} + u_{it} \tag{1}$$

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

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

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 reweights the never-treated units by the number of times they are used as a match, scaled by the number of neighbors M, assigning a weight of zero to unmatched untreated units. We refer to this estimator as the full-sample matched 2WFE estimator. The following result characterizes this estimator.

[3] We assume that the weights (w, λ) are such 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.

Proposition 1 *Let*

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

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

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

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

$$\begin{aligned}
\hat{\tau}_{\mathsf{num}}^{\mathsf{pool}}(\lambda) &= \sum_{s<\infty} \hat{p}_s \left\{ \sum_{t \geq s} \sum_{t'<s} \lambda_t \lambda_{t'} \hat{\tau}_{t,t'}^{s,\mathsf{pool}} + \sum_{s' \neq s,\infty} \frac{\hat{p}_{s'}}{1-\hat{p}_\infty} \sum_{t \geq s} \sum_{t'<s} \lambda_t \lambda_{t'} \hat{\tau}_{t,t'}^{s,s'} \right\} \\
\hat{\tau}_{\mathsf{den}}^{\mathsf{pool}}(\lambda) &\to_{\mathbb{P}} \sum_{s<\infty} p_s \left\{ \Lambda_s(1-\Lambda_s) + \sum_{s'<s} \frac{p_{s'}}{1-p_\infty} \Lambda_s(\Lambda_{s'} - \Lambda_s) + \sum_{s<s'<\infty} \frac{p_{s'}}{1-p_\infty} (1-\Lambda_s)(\Lambda_s - \Lambda_{s'}) \right\} \\
&\quad \times \left(\sum_t \lambda_t \right)^2 \\
\hat{\tau}_{t,t'}^{s,\mathsf{pool}} &= \frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = s) \left(Y_{it} - Y_{it'}\right) - \frac{1}{n - N_0} \sum_i \mathbb{1}(t_i^* = \infty) \frac{K_M(i)}{M} \left(Y_{it} - Y_{it'}\right) \\
\hat{\tau}_{t,t'}^{s,s'} &= \frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = s) \left(Y_{it} - Y_{it'}\right) - \frac{1}{N_{s'}} \sum_i \mathbb{1}(t_i^* = s') \left(Y_{it} - Y_{it'}\right), \quad s' \neq \infty.
\end{aligned}$$

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 reweighted by $K_M(i)/M$ in different post- and pre-treatment periods, $\hat{\tau}_{t,t'}^{s,\mathsf{pool}}$, and comparing each treated cohort s to other (unweighted) treated cohorts, $\hat{\tau}_{t,t'}^{s,s'}$ for $s' \neq \infty$.

This result highlights three problems with this estimator, one of which is known in the DiD literature and two of which are new and specific to the matching DiD setup. The first one is that 2x2 estimators comparing different eventually-treated cohorts, $\hat{\tau}_{t,t'}^{s,s'}$ for $s' \neq \infty$ involve cases where the comparison group is already treated, the so-called "forbidden comparisons" problem (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). Our result shows that these issues persist in matching DiD settings.

The second problem is these 2x2 comparisons between eventually-treated cohorts are unweighted and ignore the covariates that guarantee the conditional parallel trends assumption. These unweighted comparisons generally bias the estimator, as we show in our next result.

The third problem is that the weighted comparisons between treated and never treated units use

aggregate weights $K_M(i)/M$ instead of the cohort-specific weights $K_M(i,s)/M$. Thus, intuitively, this estimator matches never-treated units to an "aggregate" treated unit, ignoring potential differences in covariate distributions across treatment cohorts.

We now study these issues further by characterizing the probability limit of the estimator. To this end, 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 on their respective convex and compact supports, 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 t' and any treatment paths $\mathbf{d}_t$ and $\mathbf{d}_{t'}$, $\mathbb{C}\text{ov}(Y_{it}(\mathbf{d}_t), Y_{it'}(\mathbf{d}_{t'})|t_i^* = s, X_i = x)$ is Lipschitz-continuous and $\mathbb{V}[Y_{it}(\mathbf{d}_t)|t_i^* = s, X_i = x]$ 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.*

Our next result shows that the full-sample matched 2WFE does not generally recover a causally interpretable parameter. To simplify the exposition and abstract from biases of 2WFE estimators that have already been analyzed in the literature, for our next result we make the simplifying assumption of homogeneous ATTs, and set $\lambda = (1, 1, \ldots, 1) := \mathbf{1}$ to include all time periods in the analysis, letting $\hat{\tau}(w^{\mathsf{pool}}) := \hat{\tau}(w^{\mathsf{pool}}, \mathbf{1})$. These simplifications are included only for brevity and clarity, and Theorem B1 in the appendix generalizes the result to the case with unrestricted treatment effect heterogeneity and general time weights λ.

Theorem 1 *Suppose that $\mathbb{E}[\tau_{it}|t_i^* = s] = \tau$ for all $s < \infty$ and $t \geq s$. Under Assumptions 1-5,*

$$\hat{\tau}(w^{\mathsf{pool}}) \to_{\mathbb{P}} \tau + B_{\mathsf{pool}}$$

where

$$B_{\mathsf{pool}} = \frac{2 \sum_{s<\infty} \sum_{s' \neq s,\infty} p_s \frac{p_{s'}}{1-p_\infty} \sum_{t \geq s} \sum_{t' < s} B_{t,t'}^{s,s'}}{\sum_{s<\infty} p_s \left\{ (T+1-s)(s-1) + \sum_{s'<s} \frac{p_{s'}}{1-p_\infty}(s-s')(T+1-s) + \sum_{s<s'<\infty} \frac{p_{s'}}{1-p_\infty}(s'-s)(s-1) \right\}}.$$

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

Theorem 1 shows that the full sample matched 2WFE estimator has an asymptotic bias, which we refer to as the "pooling bias", which is a linear combination of unweighted differences in counterfactual trends between cohorts, $\mathbb{E}[Y_{it}(0) - Y_{it'}(0)|t_i^* = s] - \mathbb{E}[Y_{it}(0) - Y_{it'}(0)|t_i^* = s']$.

This bias arises from the fact that in the 2x2 comparisons implicit in the 2WFE estimator are either unweighted or incorrectly weighted, as discussed in Proposition 1. Because different treated cohorts are not matched to each other, 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 a 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. To get some further intuition, notice that under the assumptions in the theorem, these terms can be rewritten as:

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

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.

While the matching procedure corrects this lack of comparablity by matching treated and never treated units, doing so requires reweighting the never treated with cohort-specific weights $K_M(i,s)/M$. As mentioned, however, the comparisons between treated and never treated implicit in the 2WFE estimator use the weights $K_M(i)/M$, this reweighting the never treated to match the distribution of covariates in the pooled treatment group, instead of each cohort-specific distribution. As a result, this set of comparisons is biased, unless the covariate distributions are the same in all treatment cohorts.

The nature of the bias we characterize differs from that of the biases studied in the staggered DiD literature, where the negative weights problem arises from "forbidden comparisons" between just-treated and already-treated units. The issues caused by forbidden comparisons disappear when the treatment effect is homogeneous, as long as the parallel trends assumption holds. In our case, the bias is due to differences in covariate distributions across groups that yield comparisons for which the identification assumption does not hold, regardless of whether the treatment effect is homogeneous or heterogeneous.

While Theorem 1 generally invalidates the causal interpretability of the full sample matched 2WFE estimand, the pooling bias can be avoided by computing cohort-period-specific ATT estimators separately, comparing each cohort with the matched never-treated between two periods, and then aggregating the resulting estimates if desired. This strategy was suggested by Callaway and Sant'Anna (2021); Sun and Abraham (2021); Steigerwald, Vazquez-Bare, and Maier (2021), among others, in unmatched panel designs. In the next section, we show how to construct consistent and asymptotically normal estimators for cohort-period-specific ATTs and averages thereof, and provide valid variance estimators and inference procedures that account for the matching process.

4 Asymptotic Distribution and Inference

We now provide consistent estimation and valid inference procedures for matching DiD estimators, accounting for the variability of the matching weights used for the never-treated units. We start by defining some additional notation.

Let $G = |\mathcal{S} \setminus \{\infty\}|$ be the number of treated cohorts, indexed by $s_1, \ldots, s_G$. Let $\mu_t(s_g) = \mathbb{E}[Y_{it}|t_i^* = s_g]$, $\mu_t(s_g, X_i) = \mathbb{E}[Y_{it}|t_i^* = s_g, X_i]$, $\mu_t(0) = \mathbb{E}[Y_{it}|t_i^* = \infty]$, and $\mu_t(0, X_i) = \mathbb{E}[Y_{it}|t_i^* = \infty, X_i]$. The conditional variance and covariance function of the outcome are $\sigma_t^2(s_g, X_i) = \mathbb{V}[Y_{it}|t_i^* = s_g, X_i]$, $\sigma_{tt'}(s_g, X_i) = \mathbb{C}\text{ov}(Y_{it}, Y_{it'}|t_i^* = s_g, X_i)$, $\sigma_t^2(0, X_i) = \mathbb{V}[Y_{it}|t_i^* = \infty, X_i]$, and $\sigma_{tt'}(0, X_i) = \mathbb{C}\text{ov}(Y_{it}, Y_{it'}|t_i^* = \infty, X_i)$.

Next, define the sample means for the treated and reweighted never treated:

$$\widetilde{Y}_t^s = \frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = s) Y_{it}, \quad \widetilde{Y}_t^{\infty,s} = \frac{1}{N_s} \sum_i \mathbb{1}(t_i^* = \infty) \frac{K_M(i,s)}{M} Y_{it}.$$

Also let

$$\widetilde{B}_t^{s,\infty} = \frac{1}{N_s} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{K_M(i,s)}{M} \right) \mu_t(0, X_i)$$
$$\widehat{B}_t^{s,\infty} = \frac{1}{N_s} \sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty) \frac{K_M(i,s)}{M} \right) \widehat{\mu}_t(0, X_i)$$

where $\widehat{\mu}_t(0, X_i)$ is an estimator of $\mu_t(0, X_i)$. These terms constitute the oracle and estimated matching bias, as discussed in more detail below. Define the oracle and bias-corrected contrasts $\hat{\delta}_t^{s,or} = \widetilde{Y}_t^s - \widetilde{Y}_t^{\infty,s} - \widetilde{B}_t^{s,\infty}$ and $\hat{\delta}_t^{s,bc} = \widetilde{Y}_t^s - \widetilde{Y}_t^{\infty,s} - \widehat{B}_t^{s,\infty}$ and the target estimand $\delta_t^s = \mu_t(s) - \mathbb{E}[\mu_t(0, X_i)|t_i^* = s]$. Finally, define the vectors $\hat{\delta}^{s,or} = (\hat{\delta}_1^{s,or}, \ldots, \hat{\delta}_T^{s,or})'$ and $\hat{\delta}^{or} = (\hat{\delta}^{s_1,or}, \ldots, \hat{\delta}^{s_G,or})$, with analogous definitions for $\hat{\delta}^{bc}$, $\widetilde{B}$, $\widehat{B}$ and δ, and let $\mathbf{0}_{GT}$ be a $GT \times 1-$dimensional vector of zeros. Then we have the following result.

Theorem 2 *Under Assumptions 1-5,*

$$\sqrt{n}(\hat{\delta}^{or} - \delta) \to_{\mathcal{D}} \mathcal{N}(\mathbf{0}_{GT}, \mathbf{V}), \quad \mathbf{V} = \begin{bmatrix} \boldsymbol{V}^{s_1} + \boldsymbol{V}^{\infty,s_1} & \boldsymbol{V}^{\infty,s_1 s_2} & \cdots \\ \boldsymbol{V}^{\infty,s_1 s_2} & \boldsymbol{V}^{s_2} + \boldsymbol{V}^{\infty,s_2} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

where

$$\boldsymbol{V}^{\infty,s_g} = \left[V_{tt'}^{\infty,s_g}\right]_{t,t'=1}^{T}, \quad \boldsymbol{V}^{\infty,s_g s_{g'}} = \left[V_{tt'}^{\infty,s_g s_{g'}}\right]_{t,t'=1}^{T}, \quad \boldsymbol{V}^{s_g} = \left[V_{tt'}^{s_g}\right]_{t,t'=1}^{T},$$

$$V_{tt'}^{\infty,s_g} = \begin{cases} \frac{1}{p_{s_g}^2}\left(\frac{1}{M}\mathbb{E}\left[e_{s_g}(X)\sigma_t^2(0,X)\right] + \frac{\alpha(M,q)}{M^2}\mathbb{E}\left[\frac{e_{s_g}(X)^2}{e_\infty(X)}\sigma_t^2(0,X)\right]\right), & t = t', \\ \frac{1}{p_{s_g}^2}\left(\frac{1}{M}\mathbb{E}\left[e_{s_g}(X)\sigma_{tt'}(0,X)\right] + \frac{\alpha(M,q)}{M^2}\mathbb{E}\left[\frac{e_{s_g}(X)^2}{e_\infty(X)}\sigma_{tt'}(0,X)\right]\right), & t \neq t', \end{cases}$$

$$V_{tt'}^{\infty,s_g s_{g'}} = \begin{cases} \frac{1}{p_{s_g}}\frac{1}{p_{s_{g'}}}\left(\frac{\alpha(M,q)}{M^2}\mathbb{E}\left[\frac{e_{s_g}(X)e_{s_{g'}}(X)}{e_\infty(X)}\sigma_t^2(0,X)\right]\right), & t = t', \\ \frac{1}{p_{s_g}}\frac{1}{p_{s_{g'}}}\left(\frac{\alpha(M,q)}{M^2}\mathbb{E}\left[\frac{e_{s_g}(X)e_{s_{g'}}(X)}{e_\infty(X)}\sigma_{tt'}(0,X)\right]\right), & t \neq t', \end{cases}$$

$$V_{tt'}^{s_g} = \begin{cases} \frac{1}{p_{s_g}}\left(\mathbb{V}\left[\mu_t(s_g,X) - \mu_t(0,X)|t^* = s_g\right] + \mathbb{E}[\sigma_t^2(s_g,X)|t^* = s_g]\right), & t = t', \\ \frac{1}{p_{s_g}}\left(\mathbb{C}\mathrm{ov}(\mu_t(s_g,X) - \mu_t(0,X), \mu_{t'}(s_g,X) - \mu_{t'}(0,X)|\, t^* = s_g\right) & \\ + \mathbb{E}[\sigma_{tt'}(s_g,X)|t^* = s_g]\right), & t \neq t' \end{cases}$$

and where $\alpha(M,q)$ is a distribution-free, nonrandom function defined in Chen and Han (2026). If, in addition, the estimation error $r_t(X_i) = \widehat{\mu}_t(0,X_i) - \mu_t(0,X_i)$ satisfies:

$$\max_{\substack{s\in\mathcal{S}\setminus\{\infty\} \\ t=1,\ldots,T}} \frac{1}{N_s}\sum_i \mathbb{1}(t_i^* = s)\frac{1}{M}\sum_{j\in\mathcal{J}_M(i,\infty)} |r_t(X_i) - r_t(X_j)| = o_{\mathbb{P}}(n^{-1/2}), \tag{3}$$

then $\sqrt{n}(\hat{\delta}^{bc} - \hat{\delta}^{or}) = o_{\mathbb{P}}(1)$ and as a result, $\sqrt{n}(\hat{\delta}^{bc} - \delta) \to_{\mathcal{D}} \mathcal{N}(\mathbf{0}_{GT}, \mathbf{V})$.

Theorem 2 shows that the period- and cohort-specific differences between treated and reweighted never treated sample means are asymptotically normal after the appropriate centering. Specifically, this centering involves the matching bias terms $\widetilde{B}_t^{s,\infty}$ due to the fact that, when covariates are continuous, it is not possible to find perfect matches for treated units among the never treated in finite samples. This type of bias, which is common in nonparametric estimation, was pointed out for cross-sectional matching estimators by Abadie and Imbens (2006, 2011). Under the assumptions in the theorem, $\widetilde{B}_t^{s,\infty} = O_{\mathbb{P}}(n^{-1/q})$ where q is the dimension of X_i.

Because the matching bias is unknown, it has to be estimated in order to make Theorem 2 operational. Our result allows for a general estimator of the conditional mean $\mu_t(0,X_i)$, $\widehat{\mu}(0,X_i)$, under the high-level condition (3), under which the feasible bias-corrected estimator is asymptotically equivalent to the oracle estimator and thus the bias correction does not affect the asymptotic variance. This condition holds, for example, under a correctly specified parametric model for $\mu_t(0,X_i)$, or under the conditions in Theorem 2 of Abadie and Imbens (2011) for fully nonparametric series estimators. We also note that 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 vanish with the sample size.

There are specific settings in which the matching bias may be ignored asymptotically. For example, when matching on a single covariate, $\sqrt{n}\widetilde{B}_t^{s,\infty} = o_{\mathbb{P}}(1)$. 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 a related discussion. Finally, when matching on discrete covariates, matches can be exact so the matching bias is zero, as discussed in Section 7.1.

Theorem 2 also characterizes the asymptotic variance matrix $\mathbf{V}$ that accounts for the first-step matching procedure in closed form, a result that our knowledge was previously unavailable. This asymptotic variance matrix depends on a distribution-free function of the number of matches and the dimension of the covariates, $\alpha(M, q)$. Intuitively, this function characterizes the asymptotic variability of an untreated unit's catchment area (the set of covariate values for which the untreated unit is one of the M nearest neighbors) as the pool of untreated units grows. See Chen and Han (2026) for the formal definition and further discussion.

To see how Theorem 2 can be used to construct consistent and asymptotically normal bias-corrected ATT estimators, notice that taking one treated cohort s and two time periods $\tilde{t} < s \leq t$, so that $\tilde{t}$ is a pre-treatment period and t is a post-treatment period for cohort s,

$$\hat{\delta}_t^{s,bc} - \hat{\delta}_{\tilde{t}}^{s,bc} = \hat{\tau}_{t,\tilde{t}}^{s,\infty} - \widehat{B}_{t,\tilde{t}}^{s,\infty} - \mathbb{E}[\tau_{it}|t_i^* = s]$$

where $\widehat{B}_{t,\tilde{t}}^{s,\infty} = \widehat{B}_t^{s,\infty} - \widehat{B}_{\tilde{t}}^{s,\infty}$ and

$$\hat{\tau}_{t,\tilde{t}}^{s,\infty} = \frac{1}{N_s}\sum_i \mathbb{1}(t_i^* = s)(Y_{it} - Y_{i\tilde{t}}) - \frac{1}{N_s}\sum_i \mathbb{1}(t_i^* = \infty)\frac{K_M(i,s)}{M}(Y_{it} - Y_{i\tilde{t}}).$$

In the above expression, $\hat{\tau}_{t,\tilde{t}}^{s,\infty}$ is a pairwise matched DiD estimator that compares the change in outcomes between periods $\tilde{t}$ and t for cohort s to the change in outcomes for the never-treated cohort, reweighting the never treated by the matching weights. This pairwise matched DiD estimator can be computed as the 2WFE estimator (2) with weights:

$$w_i^s = \mathbb{1}(t_i^* = s) + \mathbb{1}(t_i^* = \infty)\frac{K_M(i,s)}{M}, \quad \lambda_t = \lambda_{\tilde{t}} = 1, \quad \lambda_k = 0, k \neq t, \tilde{t}, \tag{4}$$

or equivalently as a cross-sectional matching estimator with $Y_{it} - Y_{i\tilde{t}}$ as the outcome. Theorem 2 then implies that, for $\tilde{t} < s \leq t$, $\hat{\tau}_{t,\tilde{t}}^s \to_{\mathbb{P}} \mathbb{E}[\tau_{it}|t_i^* = s]$, so that the pairwise matched DiD estimator is consistent for the ATT for cohort s in period $t \geq s$. Similarly, when $\tilde{t} < t < s$, $\hat{\tau}_{t,\tilde{t}}^s \to_{\mathbb{P}} 0$ under the conditional parallel trends assumption, which serves as a pre-trends test. In addition,

$$\sqrt{n}\left(\hat{\tau}_{t,\tilde{t}}^{s,\infty} - \widehat{B}_{t,\tilde{t}}^{s,\infty} - \mathbb{E}[\tau_{it}|t_i^* = s]\right) \to_{\mathcal{D}} \mathcal{N}(0, \Omega_{t,\tilde{t}}^s) \tag{5}$$

where the asymptotic variance is:

$$\Omega^s_{t,\tilde{t}} = \frac{1}{p_s}\mathbb{V}\left[\mathbb{E}[\tau_{it}|t^*_i = s, X_i]|t^*_i = s\right] + \frac{1}{p_s}\mathbb{E}\left[\mathbb{V}[Y_{it} - Y_{i\tilde{t}}|t^*_i = s, X_i]|t^*_i = s\right] + \frac{1}{p_s^2 M^2}\mathbb{E}\left[\left(M + \alpha(M,q)\frac{e_s(X_i)}{e_\infty(X_i)}\right)e_s(X_i)\mathbb{V}[Y_{it} - Y_{i\tilde{t}}|t^*_i = \infty, X_i]\right].$$

Importantly, this result shows that, when constructing $\hat{\tau}^{s,\infty}_{t,\tilde{t}}$ with 2WFE methods as commonly done in empirical practice, the matching bias is ignored, which does not affect consistency but may generally invalidate inference based on the conventional t-statistics and confidence intervals.

In many cases, researchers may be interested in weighted averages of cohort- and period-specific ATTs:

$$\hat{\tau}(\omega) = \sum_g \sum_{t \geq s_g} \omega(s_g, t)\hat{\tau}^{s_g,\infty}_{t,\tilde{t}(g)}, \quad \omega = (\omega(s_1,s_1), \ldots, \omega(s_1,T), \ldots, \omega(s_G,s_G), \ldots, \omega(s_G,T)).$$

For example, Callaway and Sant'Anna (2021) and Sun and Abraham (2021) propose averaging ATTs by treatment exposure, aggregating the effects over cohorts in the first period of treatment, second period of treatment, and so on, as in an event-study design. Alternatively, one may average the ATTs over time for each cohort separately to assess the cohort heterogeneity in treatment effects (Callaway and Sant'Anna, 2021). Our Theorem 2 allows researchers to construct these and any other type of aggregate causal parameter based on $\hat{\tau}^{s,\infty}_{t,\tilde{t}}$ and derive their limiting variance accounting for the matching procedure and for the estimation of the aggregation weights ω, if needed, through the delta method.

Remark 1 (Normalized weights) *Because weighted OLS is invariant to rescaling all weights by a constant, it follows that the weights in (4) can be replaced by the normalized weights:*

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

which replicate those in Abadie and Imbens (2006).

Remark 2 (Not-yet-treated) *With additional notation and under the appropriate overlap conditions, our results can be generalized to the case where units in cohort s are matched to units in not-yet-treated cohort $s' \neq s$ instead of, or in addition to, the never-treated, as discussed in Callaway and Sant'Anna (2021) and Sun and Abraham (2021).*

4.1 Variance estimation

To obtain consistent estimators for the asymptotic variance matrix, we define the following objects.

$$\widehat{V}^{\infty,s_g} = \left[\widehat{V}^{\infty,s_g}_{tt'}\right]^T_{t,t'=1}, \quad \widehat{V}^{\infty,s_g s_{g'}} = \left[\widehat{V}^{\infty,s_g s_{g'}}_{tt'}\right]^T_{t,t'=1}, \quad \widehat{V}^{s_g} = \left[\widehat{V}^{s_g}_{tt'}\right]^T_{t,t'=1}$$

$$\widehat{V}_{tt'}^{\infty,s_g} = \begin{cases} \frac{n^2}{N_{s_g}^2}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{1}(t_i^*=\infty)\frac{K_M^2(i,s_g)}{M^2}\widehat{\sigma}_t^2(0,X_i)\right), & t=t', \\ \frac{n^2}{N_{s_g}^2}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{1}(t_i^*=\infty)\frac{K_M^2(i,s_g)}{M^2}\widehat{\sigma}_{tt'}(0,X_i)\right), & t\neq t', \end{cases}$$

$$\widehat{V}_{tt'}^{\infty,s_g s_{g'}} = \begin{cases} \frac{n}{N_{s_g}}\frac{n}{N_{s_{g'}}}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{1}(t_i^*=\infty)\frac{K_M(i,s_g)K_M(i,s_{g'})}{M^2}\widehat{\sigma}_t^2(0,X_i)\right), & t=t', \\ \frac{n}{N_{s_g}}\frac{n}{N_{s_{g'}}}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{1}(t_i^*=\infty)\frac{K_M(i,s_g)K_M(i,s_{g'})}{M^2}\widehat{\sigma}_{tt'}(0,X_i)\right), & t\neq t'. \end{cases}$$

$$\widehat{V}_{tt'}^{s_g} = \begin{cases} \frac{n^2}{N_{s_g}^2}\frac{1}{n}\sum_{i=1}^n \mathbb{1}(t_i^*=s_g)\left(Y_{it}-\widehat{Y}_{it}-\left(\widetilde{Y}_t^{s_g}-\widetilde{Y}_t^{\infty,s_g}\right)\right)^2 & \\ -\frac{n^2}{N_{s_g}^2}\frac{1}{n}\sum_{i=1}^n \mathbb{1}(t_i^*=\infty)\frac{K_M(i,s_g)}{M^2}\widehat{\sigma}_t^2(0,X_i), & t=t', \\ \frac{n^2}{N_{s_g}^2}\frac{1}{n}\sum_{i=1}^n \mathbb{1}(t_i^*=s_g)\left(Y_{it}-\widehat{Y}_{it}-\left(\widetilde{Y}_t^{s_g}-\widetilde{Y}_t^{\infty,s_g}\right)\right)\left(Y_{it'}-\widehat{Y}_{it'}-\left(\widetilde{Y}_{t'}^{s_g}-\widetilde{Y}_{t'}^{\infty,s_g}\right)\right) & \\ -\frac{n^2}{N_{s_g}^2}\frac{1}{n}\sum_{i=1}^n \mathbb{1}(t_i^*=\infty)\frac{K_M(i,s_g)}{M^2}\widehat{\sigma}_{tt'}^2(0,X_i), & t\neq t' \end{cases}$$

where

$$\widehat{Y}_{it} = \frac{1}{M}\sum_{m=1}^M Y_{j_m(i,\infty)t}, \quad \widetilde{Y}_t^{s_g} = \frac{1}{N_{s_g}}\sum_{i=1}^n \mathbb{1}(t_i^*=s_g)Y_{it}, \quad \widetilde{Y}_t^{\infty,s_g} = \frac{1}{N_{s_g}}\sum_{i=1}^n \mathbb{1}(t_i^*=s_g)\widehat{Y}_{it}.$$

To estimate the unknown conditional variance and covariance function in the variance estimator, we follow Abadie and Imbens (2006). For each cohort, we conduct an additional $J-$nearest-neighbor matching within the same treatment cohort and denote the index of j-th nearest neighbor for unit i within the same treatment cohort as $l_j(i)$ and denote $K_J(i)$ as the number of times unit i is used as a match in the within-cohort nearest-neighbor matching. Then we define:

$$\widehat{\sigma}_t^2(s_g,X_i) = \frac{J}{J+1}\left(Y_{it}-\frac{1}{J}\sum_{j=1}^J Y_{l_j(i)t}\right)^2$$
$$\widehat{\sigma}_{tt'}(s_g,X_i) = \frac{J}{J+1}\left(Y_{it}-\frac{1}{J}\sum_{j=1}^J Y_{l_j(i)t}\right)\left(Y_{it'}-\frac{1}{J}\sum_{j=1}^J Y_{l_j(i)t'}\right).$$

The following shows that these estimators are consistent for the elements in the asymptotic variance matrix.

Proposition 2 *Under Assumptions 1-5,* $\widehat{V}_{tt'}^{\infty,s_g} - V_{tt'}^{\infty,s_g} \to_{\mathbb{P}} 0$, $\widehat{V}_{tt'}^{\infty,s_g s_{g'}} - V_{tt'}^{\infty,s_g s_{g'}} \to_{\mathbb{P}} 0$ *and* $\widehat{V}_{tt'}^{s_g} - V_{tt'}^{s_g} \to_{\mathbb{P}} 0$.

4.2 Semiparametric Efficiency

The analytic formula for the asymptotic variance we derive in Theorem 2 allows us to study the efficiency of the pairwise matched DiD estimator $\hat{\tau}_{t,\tilde{t}}^{s,\infty}$ 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.[4] When comparing cohort s against the never-treated between periods t and $\tilde{t}$ with $\tilde{t} < s \leq t$, the bound is given by:

$$\begin{aligned}\Omega^s_{\mathsf{SEB}} = \mathbb{E}\left[\frac{\sigma^2_\Delta(s, X_i)e_s(X_i)}{p_s^2} + \frac{(\mathbb{E}[\tau_{it}|t_i^* = s, X_i] - \mathbb{E}[\tau_{it}|t_i^* = s])^2 e_s(X_i)}{p_s^2}\right] \\ + \mathbb{E}\left[\frac{e_s(X_i)^2\sigma^2_\Delta(0, X_i)}{p_s^2 e_\infty(X_i)}\right]\end{aligned} \tag{6}$$

where $\sigma^2_\Delta(s, X_i) = \mathbb{V}[Y_{it} - Y_{i\tilde{t}}|t_i^* = s, X_i]$ and $\sigma^2_\Delta(0, X_i) = \mathbb{V}[Y_{it} - Y_{i\tilde{t}}|t_i^* = \infty, X_i]$ and the dependence on t and $\tilde{t}$ is left implicit. Then we have the following result.

Proposition 3 *Under Assumptions 1-5,*

$$\Omega^s_{t,\tilde{t}} - \Omega^s_{\mathsf{SEB}} = \frac{1}{M}\mathbb{E}\left[\frac{e_s(X_i)\sigma^2_\Delta(0, X_i)}{p_s^2}\right] + \left(\frac{\alpha(M, q)}{M^2} - 1\right)\mathbb{E}\left[\frac{e_s(X_i)^2\sigma^2_\Delta(0, X_i)}{e_\infty(X_i)p_s^2}\right]$$

where $\alpha(M, q) \geq M^2$.

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, 2026) and thus:

$$\Omega^s_{t,\tilde{t}} - \Omega^s_{\mathsf{SEB}} = \frac{1}{M}\mathbb{E}\left[\frac{e_s(X_i)\sigma^2_\Delta(0, X_i)}{p_s^2}\right] + \frac{1}{2M}\mathbb{E}\left[\frac{e_s(X_i)^2\sigma^2_\Delta(0, X_i)}{e_\infty(X_i)p_s^2}\right].$$

It can be seen that the efficiency gap decreases with M, in line with the results in Abadie and Imbens (2006) and Lin, Ding, and Han (2023) who consider asymptotics with $M \to \infty$ 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 (2026) show that for $M \in \{2, 3, \ldots 10\}$ and $q \in \{2, 3, \ldots, 10\}$, $\alpha(M, q)/M^2 - 1 < 0.15$, and that $\alpha(M, q)/M^2 - 1$ is positive and decreasing in M for each of these values of M and q.

[4] For simplicity, we focus on the two-group, two-period model. With multiple periods and groups, the model is over-identified when the conditional parallel trends assumption holds between all pairs of periods, but we do not consider that case here. See Chen, Sant'Anna, and Xie (2025) for a related discussion.

4.3 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 shown in Theorem 2. In practice, however, researchers often rely on the cluster-robust variance estimator from the second-stage weighted 2WFE regression, ignoring the fact that the weights are estimated in the matching step, which results in inconsistent variance estimators.[5] 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. 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 weighted 2WFE variance estimator that ignores the matching step, which we refer to as the naive cluster-robust variance estimator.

Specifically, the cluster-robust estimator of the variance of $\hat{\tau}^{s,\infty}_{t,\tilde{t}}$, computed from the weighted 2WFE regression (1) with weights (4), can be written as:

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

where:

$$\hat{\varepsilon}_i = \Delta Y_i - \left(\frac{1}{N_s}\sum_j \mathbb{1}(t_j^* = s)\Delta Y_j\right)\mathbb{1}(t_i^* = s) - \left(\frac{1}{N_s}\sum_j \mathbb{1}(t_j^* = \infty)\frac{K_M(j,s)}{M}\Delta Y_j\right)\mathbb{1}(t_i^* = \infty)$$

is the regression residual, $\Delta Y_i = Y_{it} - Y_{i\tilde{t}}$ and the dependence on $t, \tilde{t}$ is left implicit to reduce notation. The following result characterizes the probability limit of this estimator.

Proposition 4 *Under Assumptions 1-5,*

$$\begin{aligned} n\hat{V}_{\mathsf{ncr}} \to_{\mathbb{P}} \Omega^s_{t,\tilde{t}} \\ &+ \frac{1}{p_s}\mathbb{V}[\mathbb{E}[\Delta Y_i(0)|t_i^* = s, X_i]|t_i^* = s] \\ &+ \frac{2}{p_s}\mathbb{C}\mathrm{ov}(\mathbb{E}[\tau_{it}|t_i^* = s, X_i], \mathbb{E}[\Delta Y_i(0)|t_i^* = s, X_i]|t_i^* = s) \\ &+ \frac{1}{p_s^2 M^2}\mathbb{E}\left[\left(M + \alpha(M,q)\frac{e_s(X_i)}{e_\infty(X_i)}\right)e_s(X_i)(\mathbb{E}[\Delta Y_i(0)|t_i^* = s, X_i] - \mathbb{E}[\Delta Y_i(0)|t_i^* = s])^2\right]. \end{aligned}$$

[5] A notable exception is Deryugina, MacKay, and Reif (2020), who specifically account for the matching step using the methods in Abadie and Imbens (2006, 2011) in their DiD analysis.

where $\Delta Y_i(0) = Y_{it}(0) - Y_{i\tilde{t}}(0)$.

Proposition 4 shows that if $\mathbb{E}[\Delta Y_i(0)|t_i^* = s, X_i]$ is constant in X_i over the support of $X_i|t_i^* = s$, so that covariates do not affect the average counterfactual trend over the support of cohort s, 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:

$$\begin{aligned}
&\frac{1}{p_s}\mathbb{V}[\mathbb{E}[\Delta Y_i(0)|t_i^* = s, X_i]|t_i^* = s] \\
&+ \frac{2}{p_s}\mathbb{C}\text{ov}(\mathbb{E}[\tau_{it}|t_i^* = s, X_i], \mathbb{E}[\Delta Y_i(0)|t_i^* = s, X_i]|t_i^* = s) \\
&+ \frac{1}{p_s^2 M^2}\mathbb{E}\left[\left(M + \alpha(M,q)\frac{e_s(X_i)}{e_\infty(X_i)}\right) e_s(X_i)(\mathbb{E}[\Delta Y_i(0)|t_i^* = s, X_i] - \mathbb{E}[\Delta Y_i(0)|t_i^* = s])^2\right].
\end{aligned}$$

The direction and magnitude of this bias depend on the data generating process. The first and third terms are always nonnegative, 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 B1, we assume that the conditional ATT 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:

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

which implies that the conditional ATT is $\mathbb{E}[Y_{it}(1) - Y_{it}(0)|t_i^* = 3, X_i] = 5 + 7X_i(t-1)$ 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 $\mathbb{C}\text{ov}(\mathbb{E}[Y_{it}(1) - Y_{it}(0)|t_i^* = 3, X_i], \mathbb{E}[\Delta Y_{it}(0)|X_i, t_i^* = 3]|t_i^* = 3) = -14(t-1)\mathbb{V}[X_i|t_i^* = 3]$ 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 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 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 Applications

6.1 The National Supported Work Demonstration

We first 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)[6] 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

[6]Specifically, we use Westat's matched CPS, which includes earnings data from multiple periods. See Smith and Todd (2005) for details. The data was obtained from Sant'Anna and Zhao (2020)'s replication files.

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.

6.2 The Effect of Community Health Centers on Mortality

We now use the data from Bailey and Goodman-Bacon (2015) to estimate the effect of the construction of Community Health Centers (CHCs) on county-level mortality rates in the US. In the original paper, the authors use an event-study design to estimate the long-term health effects of increased primary care access by exploiting the plausibly exogenous timing in the rollout of the first CHCs starting in 1965.[7]

Because the treatment adoption design is staggered, recent literature has shown that estimating a 2WFE regression is subject to weighting issues and forbidden comparisons. These issues can be avoided by separately estimating cohort-specific ATTs in each time period, and then aggregating these estimated ATTs by event time. The results from this specification can be seen in Figure 1. The validity of these estimates relies on the unconditional parallel trends assumption. The figure, however, shows that the pre-trends coefficients are non-zero and jointly statistically significant, with a p-value for the joint null of no pre-treatment trends differences close to zero, which makes the identification assumption harder to defend.

To address these pre-trends differences, we apply the methods discussed in the previous sections. We start by matching each treated unit to a unit in the never-treated group, using the following time-invariant covariates measured in 1960 at the county level: percentage of urban population, percentage of rural population, percentage of non-white population, percentage of population ages 0-4, percentage of population aged 65 or more, median years of schooling for persons aged 25 and older, percentage of population aged 25 and older with less than 4 years of schooling, percentage of population aged 25 and older with more 12 years of schooling or more, percentage of population with family income below $3,000, percentage of population with family income above $10,000, median family income, retirement transfers and public assistance transfers. We also include the outcome in 1966 (a pure pre-treatment period in the sample) as a matching covariate. The outcome of interest is adjusted mortality rate. We refer the reader to the original paper and replication package for details about the data.

Based on the matching weights, we construct ATT estimators following Theorem 2 and their corresponding standard errors based on Proposition 2, aggregating the estimated effects by event time. We restrict the results to the event window $[-15, 15]$ to exclude very long term estimates

[7] While the identification strategy in the original paper relies on covariates, their estimation procedure does not use matching. We note that our goal is not to replicate or assess the results in the original paper, but only to use this data set to illustrate our proposed methods.

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.

computed using very few observations. The estimates in Figure 1 show that matching successfully makes the pre-trends coefficients closer to zero and jointly insignificant. The full event study with 95% confidence intervals is shown in Figure 2. The estimates suggest negative effects that tend to increase in magnitude over time, although they are mostly insignificant at the 5% level.

Finally, Table 3 shows the estimated overall ATT (aggregated over cohorts and time periods) for the unconditional DiD estimator, the pooled matching estimator characterized in Proposition 1 and Theorem 1, and the post-matching DiD estimator based on Theorem 2. The results show that adjusting for covariates through matching increases the magnitude of the effect by about 78%, although the overall ATT remains insignificant at the 5% level. In this application, the pooling bias seen from comparing columns (2) and (3) is not substantial. The fact that the standard error is larger in column (3) compared to column (2) is partly explained by the fact that the estimates in column (3) incorporate the variability introduced by the matching procedure.

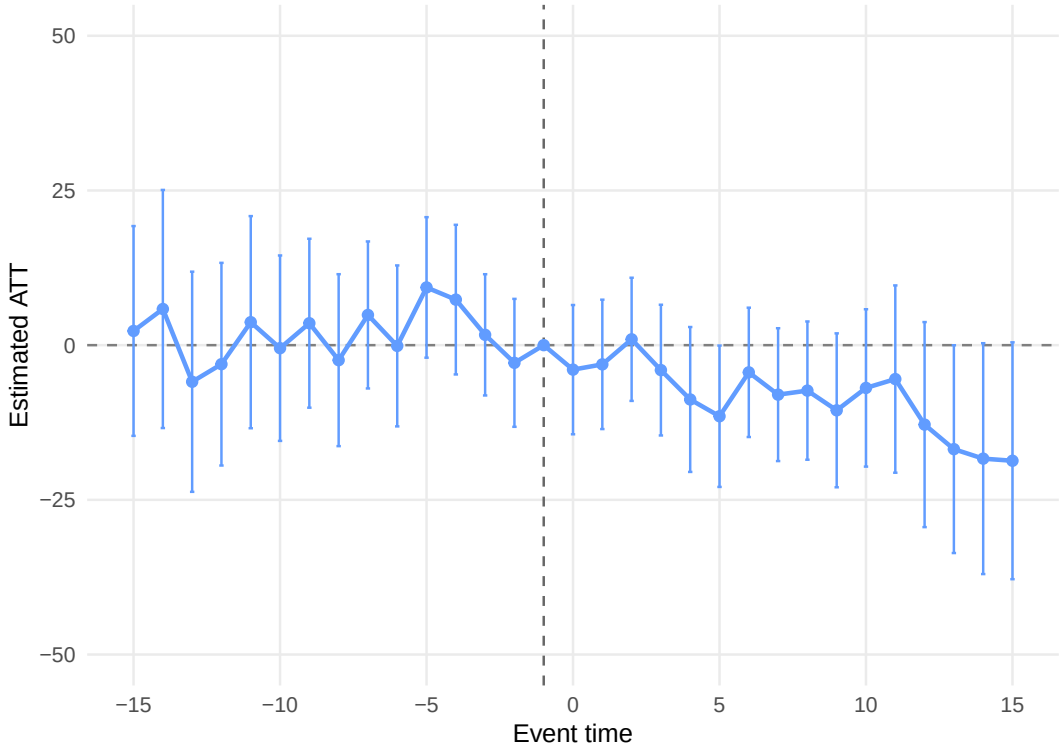


Figure 2: Post-matching event study estimates with 95% confidence intervals

7 Extensions to Other Types of Matching

In this section we discuss several extensions to our results, namely, matching when all covariates are discrete, matching without replacement and propensity score matching. For brevity, we consider the case with two periods $\tilde{t} < t$ and two cohorts, $t_i^* \in \{s, \infty\}$, with $\tilde{t} < s \leq t$, although all the results we present here can straightforwardly be generalized to multiple periods. In what

Table 3: Overall ATT estimates

	(1) Unconditional DiD	(2) Naive pooled matched 2WFE	(3) Post-matching DiD
Estimate	-3.03	-5.15	-5.40
Std. err.	3.19	3.94	4.41
95% CI	[-9.28, 3.22]	[-12.87, 2.56]	[-14.05, 3.25]

Notes: the estimate in column (1) is an average of pairwise ATTs over treatment exposure. The estimate in column (2) is obtained through a weighted 2WFE regression using the matching weights obtained in the first stage. The standard error in this column ignores the matching procedure. The estimate in column (3) is computed as explained in Theorem 2 and subsequent discussion, accounting for the matching procedure when computing standard errors.

follows we let $\Delta Y_i = Y_{it} - Y_{i\tilde{t}}$, $\mu_\Delta(s, X_i) = \mathbb{E}[\Delta Y_i | t_i^* = s, X_i]$, $\sigma^2_\Delta(s, X_i) = \mathbb{V}[\Delta Y_i | t_i^* = s, X_i]$ and $\sigma^2_\Delta(0, X_i) = \mathbb{V}[\Delta Y_i | t_i^* = \infty, X_i]$ where the dependence on t, $\tilde{t}$ is left implicit to simplify the notation.

7.1 Matching on Discrete Covariates

When the vector X_i is a low-dimensional vector of discrete covariates, each treated unit can be matched exactly to an untreated unit, 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 cohort-specific ATT estimator is given by:

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

where

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

This estimator that reweights never-treated units by the ratio of the estimated propensity scores is equivalent to a matched DiD estimator 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 $n \to \infty$, 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 (6), as shown in the following result.

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

$$\sqrt{n}(\hat{\tau}_{\mathsf{disc}} - \mathbb{E}[\tau_{it} | t_i^* = s]) \to_{\mathcal{D}} \mathcal{N}(0, \Omega^s_{\mathsf{SEB}})$$

where Ω^s_{SEB} *is defined in* (6).

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 untreated 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). In this case, the post-matching 2WFE estimator using matching without replacement is:

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

For brevity, we will consider an alternative, conditional sampling scheme requiring that $N_s = O(N_0^{1/r})$ with $r > q$, which intuitively implies that the untreated group is "sufficiently larger" than the treated group. Under this sampling scheme, the matching bias is $o_{\mathbb{P}}(N_s^{-1/2})$ (Abadie and Spiess, 2022) and thus bias correction is not required for asymptotic normality. This is not specific to matching without replacement, and can also be applied to matching with replacement, as used in the previous sections of the paper, under the rate condition $r > q/2$. 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* $Y_1, \ldots, Y_T, X | t^* = s$ *and* $Y_1, \ldots, Y_T, X | t^* = \infty$*, respectively, with* $N_0 \geq MN_s$ *and* $N_s = O(N_0^{1/r})$ *for some* $r > q$*. Then, for matching without replacement under Assumptions 1, 2, 3 and 5, as* $N_s \to \infty$*,*

$$\sqrt{(M+1)N_s}(\hat{\tau}_{\mathsf{nr}} - \mathbb{E}[\tau_{it} | t_i^* = s]) \to_{\mathcal{D}} \mathcal{N}(0, \Omega_{\mathsf{nr}})$$

where

$$\begin{aligned}\Omega_{\mathsf{nr}} = (1+M)\mathbb{E}\left[(\mathbb{E}[\tau_{it} | t_i^* = s, X_i] - \mathbb{E}[\tau_{it} | t_i^* = s])^2 \middle| t_i^* = s \right] \\ + (1+M)\mathbb{E}\left[\sigma^2_\Delta(s, X_i) + \frac{\sigma^2_\Delta(0, X_i)}{M} \middle| t_i^* = s \right].\end{aligned}$$

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. We leave the formal treatment of this case for future research

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.

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 and from Bailey and Goodman-Bacon (2015).

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 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.

Bailey, M. J., and A. Goodman-Bacon (2015): "The War on Poverty's Experiment in Public Medicine: Community Health Centers and the Mortality of Older Americans," *American Economic Review*, 105(3), 1067–1104.

Borusyak, K., X. Jaravel, and J. Spiess (2024): "Revisiting Event-Study Designs: Robust and Efficient Estimation," *The Review of Economic 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 (2026): "On the limiting variance of matching estimators," *working paper*.

Chen, X., P. H. C. Sant'Anna, and H. Xie (2025): "Efficient Difference-in-Differences and Event Study 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'Haultfœuille (2023): "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey," *The Econometrics Journal*, 26(3), C1–C30.

Deryugina, T., A. MacKay, and J. Reif (2020): "The Long-Run Dynamics of Electricity Demand: Evidence from Municipal Aggregation," *American Economic Journal: Applied Economics*, 12(1), 86–114.

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*, 65(2), 261–294.

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.

Smith, J. A., and P. E. Todd (2005): "Does matching overcome LaLonde's critique of nonexperimental estimators?," *Journal of Econometrics*, 125(1), 305–353.

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 2013 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 (AEJAE), American Economic Journal: Economic Policy (AEJEP), American Economic Review (AER), American Political Science Review (APSR) and Quarterly Journal of Economics (QJE).

B Additional results

Lemma B1 *Let*

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

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

$$\hat{\tau}_{\mathsf{num}}(w,\lambda) = \sum_{s<\infty}\sum_{s'\neq s}\sum_{t\geq s}\sum_{t'<s} \lambda_t\lambda_{t'}\hat{p}_s^w\hat{p}_{s'}^w\hat{\tau}_{t,t'}^{s,s'}(w)$$

$$\hat{\tau}_{\mathsf{den}}(w,\lambda) = \left(\sum_t \lambda_t\right)^2 \sum_{s<\infty}\hat{p}_s^w\left\{\hat{p}_\infty^w\Lambda_s(1-\Lambda_s) + \sum_{s'<s}\hat{p}_{s'}^w\Lambda_s(\Lambda_{s'}-\Lambda_s) + \sum_{s<s'<\infty}\hat{p}_{s'}^w(1-\Lambda_s)(\Lambda_s-\Lambda_{s'})\right\}$$

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

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

$$\hat{p}_s^w = \frac{\sum_i w_i\mathbb{1}(t_i^*=s)}{\sum_i w_i}.$$

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*

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-Urena, 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 neighbor matching, Mahalanobis 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, Wachter, Heining	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, Spiller, Timmins	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, MacKay, Reif	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	Ticketing and 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

Assumption 3 and 5(1),

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

and

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

Lemma B3 *Suppose we conduct two nearest neighbor matching with M neighbors on a subset of the sample with three cohorts s, s', and s'', where in each matching, each unit in cohorts s, and s' is matched to M nearest neighbors from cohort s''. Let $K_M(i,s)$ and $K_M(i,s')$ denote the number of times unit i is used as a match in the nearest neighbor matching with cohort s and s'. Let $(Z_i)_{i=1}^n$ be a sequence of i.i.d random variables with $\mathbb{E}[Z_i|t_i^*=s,X_i=x]=m(s,x)$, 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, s', and s''. In addition, assume that the conditional densities of $X_i|t_i^*=s$, $X_i|t_i^*=s'$, and $X_i|t_i^*=s''$, $f_s(x)$, $f_{s'}(x)$ and $f_{s''}(x)$ respectively, are continuous and have convex and compact support, and $f_{s''}(x)$ is bounded away from zero. Then,*

$$\frac{1}{n}\sum_{i=1}^n \mathbb{1}(t_i^*=s'')\frac{K_M(i,s)K_M(i,s')}{M^2}Z_i \to_{\mathbb{P}} \frac{\alpha(M,q)}{M^2}\mathbb{E}\left[\frac{e_s(X_i)e_{s'}(X_i)}{e_{s''}(X_i)}m(s'',X_i)\right].$$

Lemma B4 *Performing nearest-neighbor matching to match each unit in cohort s to units in cohort s', with possibly $s=s'$, let $j_m(i,s')$ denote the index of the m-th closest match in subgroup s' for unit i in subgroup s. Then $\max_{t_i^*=s}\left\|X_i-X_{j_m(i,s')}\right\|=o_p(1)$.*

Lemma B5 *For any $\alpha>0$, $\frac{1}{n}\sum_{i=1}^n \mathbb{1}(t_i^*=\infty)K_M^\alpha(i,s_g)=O_p(1)$.*

Theorem B1 *Under Assumptions 1-5,*

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

where $B^{\mathsf{het}}_{\mathsf{time}}$ *and* B_{pool} *are bias terms defined as:*

$$B^{\mathsf{het}}_{\mathsf{time}} = \sum_{s<\infty}\sum_{s'<s} \eta_1(s,s',\lambda) \frac{\sum_{t\geq s}\sum_{s'\leq t'<s}\lambda_t\lambda_{t'}\mathbb{E}[\tau_{it}-\tau_{it'}|t_i^*=s']}{\sum_{t\geq s}\sum_{s'\leq t'<s}\lambda_t\lambda_{t'}}$$

$$B_{\mathsf{pool}} = 2\sum_{s<\infty}\sum_{s'\neq s,\infty} \eta_2(s,s',\lambda) \frac{\sum_{t\geq s}\sum_{t'<s}\lambda_t\lambda_{t'}B^{s,s'}_{t,t'}}{\sum_{t\geq s}\sum_{t'<s}\lambda_t\lambda_{t'}}$$

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

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

$$\phi_1(s,\lambda) = \frac{p_s\Lambda_s\left[(1-\Lambda_s)+\sum_{s'<s}p_{s'}(\Lambda_{s'}-\Lambda_s)/(1-p_\infty)\right]}{\sum_{s<\infty}\left\{p_s\Lambda_s\left[(1-\Lambda_s)+\frac{\sum_{s'<s}p_{s'}(\Lambda_{s'}-\Lambda_s)}{1-p_\infty}\right]+\sum_{s<s'<\infty}\frac{p_sp_{s'}}{1-p_\infty}(1-\Lambda_s)(\Lambda_s-\Lambda_{s'})\right\}}$$

$$\phi_2(s,s',\lambda) = \frac{p_sp_{s'}(1-\Lambda_s)(\Lambda_s-\Lambda_{s'})/(1-p_\infty)}{\sum_{s<\infty}\left\{p_s\Lambda_s\left[(1-\Lambda_s)+\frac{\sum_{s'<s}p_{s'}(\Lambda_{s'}-\Lambda_s)}{1-p_\infty}\right]+\sum_{s<s'<\infty}\frac{p_sp_{s'}}{1-p_\infty}(1-\Lambda_s)(\Lambda_s-\Lambda_{s'})\right\}}$$

$$\eta_1(s,s',\lambda) = \frac{p_sp_{s'}\Lambda_s(\Lambda_{s'}-\Lambda_s)/(1-p_\infty)}{\sum_{s<\infty}\left\{p_s\Lambda_s\left[(1-\Lambda_s)+\frac{\sum_{s'<s}p_{s'}(\Lambda_{s'}-\Lambda_s)}{1-p_\infty}\right]+\sum_{s<s'<\infty}\frac{p_sp_{s'}}{1-p_\infty}(1-\Lambda_s)(\Lambda_s-\Lambda_{s'})\right\}}$$

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

Theorem B1 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 three components. The first component, in the first line of the expression for $\hat{\tau}(w^{\mathsf{pool}},\lambda)$ 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 two bias terms. The first bias, $B^{\mathsf{het}}_{\mathsf{time}}$, 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 $B^{\mathsf{het}}_{\mathsf{time}}$ has been discussed in existing literature analyzing unmatched DiD estimators. Theorem B1 shows that these issues persist when using matched 2WFE estimators. The second bias is the pooling bias discussed in Theorem 1.

C Proofs of Main Results

Proof of Lemma 1

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

as required. □

Proof of Proposition 1

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

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

which implies that $\sum_i w_i^{\mathsf{pool}} = 2(n - N_0)$, $\hat{p}_\infty^w = 1/2$ 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 after defining:

$$\begin{aligned}
\hat{\tau}_{t,t'}^{s,\infty}(w^{\mathsf{pool}}) = \hat{\tau}_{t,t'}^{s,\mathsf{pool}} &:= \frac{1}{N_s}\sum_i \mathbb{1}(t_i^* = s)\,(Y_{it} - Y_{it'}) - \frac{1}{n - N_0}\sum_i \mathbb{1}(t_i^* = \infty)\frac{K_M(i)}{M}\,(Y_{it} - Y_{it'}) \\
\hat{\tau}_{t,t'}^{s,s'}(w^{\mathsf{pool}}) = \hat{\tau}_{t,t'}^{s,s'} &:= \frac{1}{N_s}\sum_i \mathbb{1}(t_i^* = s)\,(Y_{it} - Y_{it'}) - \frac{1}{N_{s'}}\sum_i \mathbb{1}(t_i^* = s')\,(Y_{it} - Y_{it'}), \quad s' \neq \infty
\end{aligned}$$

as required. □

Proof of Theorem 1

This result follows as a particular case of Theorem B1. □

Proof of Theorem 2

We begin by establishing a more general joint asymptotic result for the vector of treated means and their matched counterparts. Specifically, define the re-centered mean for each cohort s_g and

the corresponding matched never-treated cohort in period t as:

$$\overline{Y}_t^{s_g} = \frac{1}{N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g) \left[(Y_{it} - \mu_t(0, X_i)) - (\mu_t(s_g) - \mathbb{E}[\mu_t(0, X) | t^* = s_g]) \right]$$
$$\overline{Y}_t^{\infty, s_g} = \frac{1}{N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s_g)}{M} (Y_{it} - \mu_t(0, X_i)).$$

By construction, $\mathbb{E}[\overline{Y}_t^{s_g}] = \mathbb{E}[\overline{Y}_t^{\infty, s_g}] = 0$. Stack the re-centered means at each time period into the vectors $\overline{Y}^{s_g} = (\overline{Y}_1^{s_g}, \ldots, \overline{Y}_T^{s_g})'$ and $\overline{Y}^{\infty, s_g} = (\overline{Y}_1^{\infty, s_g}, \ldots, \overline{Y}_T^{\infty, s_g})'$. In what follows, $p_{s_g} = \mathbb{P}[t_i^* = s_g]$ and $\mathbf{0}$ represents a $T \times 1$ vector of zeros. We will begin by showing that:

$$\sqrt{n} \begin{pmatrix} \overline{Y}^{\infty, s_1} \\ \vdots \\ \overline{Y}^{\infty, s_G} \\ \overline{Y}^{s_1} \\ \vdots \\ \overline{Y}^{s_G} \end{pmatrix} \to_{\mathcal{D}} \mathcal{N} \left(\begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{V}^{\infty, s_1} & \cdots & \boldsymbol{V}^{\infty, s_1 s_G} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{V}^{\infty, s_G s_1} & \cdots & \boldsymbol{V}^{\infty, s_G} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{V}^{s_1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{V}^{s_G} \end{pmatrix} \right)$$

where the elements in the variance-covariance matrix are defined in the statement of the theorem. We establish the result using the Cramér-Wold device together with the martingale representation of the matching estimator. Define $\lambda = (\lambda^{\infty, s_1}, \ldots, \lambda^{\infty, s_G}, \lambda^{s_1}, \ldots, \lambda^{s_G})'$, where $\lambda^{\infty, s_g} = (\lambda_1^{\infty, s_g}, \ldots, \lambda_T^{\infty, s_g})'$, $\lambda^{s_g} = (\lambda_1^{s_g}, \ldots, \lambda_T^{s_g})'$, and define the residual as $\epsilon_{it} = Y_{it} - \mathbb{E}[Y_{it} \mid t_i^*, X_i] = Y_{it} - \mu_t(t_i^*, X_i)$. Fix any λ such that $\|\lambda\| = 1$. Then

$$\begin{aligned}
\sqrt{n} \lambda' \overline{Y} &= \sqrt{n} \sum_{g=1}^{G} \sum_{t=1}^{T} \left(\lambda_t^{\infty, s_g} \overline{Y}_t^{\infty, s_g} + \lambda_t^{s_g} \overline{Y}_t^{s_g} \right) \\
&= \sqrt{n} \sum_{g=1}^{G} \sum_{t=1}^{T} \left[\frac{\lambda_t^{\infty, s_g}}{N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s_g)}{M} (Y_{it} - \mu_t(0, X_i)) \right. \\
&\quad \left. + \frac{\lambda_t^{s_g}}{N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g) \left[(Y_{it} - \mu_t(0, X_i)) - (\mu_t(s_g) - \mathbb{E}[\mu_t(0, X) | t^* = s_g]) \right] \right] \\
&= \sum_{i=1}^{n} \sum_{g=1}^{G} \sum_{t=1}^{T} \frac{n}{N_{s_g}} \frac{\lambda_t^{s_g}}{\sqrt{n}} \mathbb{1}(t_i^* = s_g) \left[\mu_t(s_g, X_i) - \mu_t(0, X_i) - (\mu_t(s_g) - \mathbb{E}[\mu_t(0, X_i) | t^* = s_g]) \right] \\
&\quad + \sum_{i=1}^{n} \sum_{g=1}^{G} \sum_{t=1}^{T} \frac{n}{N_{s_g}} \frac{1}{\sqrt{n}} \left(\lambda_t^{s_g} \mathbb{1}(t_i^* = s_g) + \lambda_t^{\infty, s_g} \mathbb{1}(t_i^* = \infty) \frac{K_M(i, s_g)}{M} \right) \epsilon_{it}
\end{aligned}$$

To obtain a martingale representation of this scalar projection, define $Z_{n,i} = \sum_{g=1}^{G}\sum_{t=1}^{T} Z_{n,i}^{s_g t}$, where

$$Z_{n,i}^{s_g t} = \begin{cases} \frac{n}{N_{s_g}} \frac{\lambda_t^{s_g}}{\sqrt{n}} \mathbb{1}(t_i^* = s_g) \left[\mu_t(s_g, X_i) - \mu_t(0, X_i) - (\mu_t(s_g) - \mathbb{E}[\mu_t(0, X_i)|t^* = s_g])\right] & \text{if } 1 \leq i \leq n \\ \frac{n}{N_{s_g}} \frac{1}{\sqrt{n}} \left(\lambda_t^{s_g} \mathbb{1}(t_{i-n}^* = s_g) + \lambda_t^{\infty, s_g} \mathbb{1}(t_{i-n}^* = \infty) \frac{K_M(i-n, s_g)}{M}\right) \epsilon_{i-n,t} & \text{if } n+1 \leq i \leq 2n \end{cases}$$

and introduce the corresponding $\sigma-$ fields by

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

where $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iT})'$. By construction, $\mathbb{E}[Z_{n,i}^{s_g t}] = 0$, $\mathbb{E}[Z_{n,i}] = 0$, $\sqrt{n}\lambda'\overline{Y} = \sum_{i=1}^{2n} Z_{n,i}$, and therefore 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 sequence. We first compute its conditional quadratic variation. For $1 \leq i \leq n$,

$$\begin{aligned}
&\mathbb{E}[Z_{n,i}^2 | \mathcal{F}_{n,i-1}] \\
&= \mathbb{E}\left[\left(\sum_{g=1}^{G}\sum_{t=1}^{T} Z_{n,i}^{s_g t} \right)^2 \middle| \mathcal{F}_{n,i-1} \right] \\
&= \mathbb{E}\left[\left(\sum_{g=1}^{G}\sum_{t=1}^{T} \frac{n}{N_{s_g}} \frac{\lambda_t^{s_g}}{\sqrt{n}} \mathbb{1}(t_i^* = s_g) \left[\mu_t(s_g, X_i) - \mu_t(0, X_i) - (\mu_t(s_g) - \mathbb{E}[\mu_t(0, X_i)|t^* = s_g])\right] \right)^2 \middle| \mathcal{F}_{n,i-1} \right] \\
&= \sum_{g=1}^{G}\sum_{t=1}^{T} \frac{n^2}{N_{s_g}^2} \frac{(\lambda_t^{s_g})^2}{n} \mathbb{1}(t_i^* = s_g) \mathbb{V}\left[\mu_t(s_g, X_i) - \mu_t(0, X_i) \middle| t_i^* = s_g\right] \\
&+ \sum_{g=1}^{G}\sum_{t=1}^{T}\sum_{t' \neq t} \frac{n^2}{N_{s_g}^2} \frac{\lambda_t^{s_g}}{\sqrt{n}} \frac{\lambda_{t'}^{s_g}}{\sqrt{n}} \mathbb{1}(t_i^* = s_g) \mathbb{C}\text{ov}(\mu_t(s_g, X_i) - \mu_t(0, X_i), \mu_{t'}(s_g, X_i) - \mu_{t'}(0, X_i) \middle| t_i^* = s_g)
\end{aligned}$$

For $n+1 \leq i \leq 2n$, the corresponding conditional second moment is

$$
\begin{aligned}
&\mathbb{E}[Z_{n,i}^2 | \mathcal{F}_{n,i-1}] \\
&= \mathbb{E}\left[\left(\sum_{g=1}^{G} \sum_{t=1}^{T} Z_{n,i}^{s_g t} \right)^2 \middle| \mathcal{F}_{n,i-1} \right] \\
&= \mathbb{E}\left[\left(\sum_{g=1}^{G} \sum_{t=1}^{T} \frac{n}{N_{s_g}} \frac{1}{\sqrt{n}} \left(\lambda_t^{s_g} \mathbb{1}(t_{i-n}^* = s_g) + \lambda_t^{\infty, s_g} \mathbb{1}(t_{i-n}^* = \infty) \frac{K_M(i-n, s_g)}{M} \right) \epsilon_{i-n,t} \right)^2 \middle| \mathcal{F}_{n,i-1} \right] \\
&= \sum_{g=1}^{G} \sum_{t=1}^{T} \frac{n^2}{N_{s_g}^2} \frac{(\lambda_t^{s_g})^2}{n} \mathbb{1}(t_{i-n}^* = s_g) \mathbb{E}[\epsilon_{i-n,t}^2 | t_{i-n}^* = s_g, X_{i-n}] \\
&+ \sum_{g=1}^{G} \sum_{t=1}^{T} \sum_{t' \neq t} \frac{n^2}{N_{s_g}^2} \frac{\lambda_t^{s_g}}{\sqrt{n}} \frac{\lambda_{t'}^{s_g}}{\sqrt{n}} \mathbb{1}(t_{i-n}^* = s_g) \mathbb{E}[\epsilon_{i-n,t} \epsilon_{i-n,t'} | t_{i-n}^* = s_g, X_{i-n}] \\
&+ \sum_{g=1}^{G} \sum_{t=1}^{T} \frac{n^2}{N_{s_g}^2} \frac{(\lambda_t^{\infty, s_g})^2}{n} \mathbb{1}(t_{i-n}^* = \infty) \frac{K_M^2(i-n, s_g)}{M^2} \mathbb{E}[\epsilon_{i-n,t}^2 | t_{i-n}^* = \infty, X_{i-n}] \\
&+ \sum_{g=1}^{G} \sum_{t=1}^{T} \sum_{t' \neq t} \frac{n^2}{N_{s_g}^2} \frac{\lambda_t^{\infty, s_g}}{\sqrt{n}} \frac{\lambda_{t'}^{\infty, s_g}}{\sqrt{n}} \mathbb{1}(t_{i-n}^* = \infty) \frac{K_M^2(i-n, s_g)}{M^2} \mathbb{E}[\epsilon_{i-n,t} \epsilon_{i-n,t'} | t_{i-n}^* = \infty, X_{i-n}] \\
&+ \sum_{g=1}^{G} \sum_{s_{g'} \neq s_g} \sum_{t=1}^{T} \frac{n}{N_{s_g}} \frac{n}{N_{s_{g'}}} \frac{\lambda_t^{\infty, s_g}}{\sqrt{n}} \frac{\lambda_t^{\infty, s_{g'}}}{\sqrt{n}} \mathbb{1}(t_{i-n}^* = \infty) \frac{K_M(i-n, s_g) K_M(i-n, s_{g'})}{M^2} \mathbb{E}[\epsilon_{i-n,t}^2 | t_{i-n}^* = \infty, X_{i-n}] \\
&+ \sum_{g=1}^{G} \sum_{s' \neq s} \sum_{t=1}^{T} \sum_{t' \neq t} \frac{n}{N_{s_g}} \frac{n}{N_{s_{g'}}} \frac{\lambda_t^{\infty, s_g}}{\sqrt{n}} \frac{\lambda_{t'}^{\infty, s_{g'}}}{\sqrt{n}} \mathbb{1}(t_{i-n}^* = \infty) \frac{K_M(i-n, s_g) K_M(i-n, s_{g'})}{M^2} \\
&\mathbb{E}[\epsilon_{i-n,t} \epsilon_{i-n,t'} | t_{i-n}^* = \infty, X_{i-n}]
\end{aligned}
$$

Summing these conditional second moments over both parts of the martingale array yields

$$
\begin{aligned}
&\sum_{i=1}^{2n} \mathbb{E}[Z_{n,i}^2 | \mathcal{F}_{n,i-1}] \\
&= \sum_{g=1}^{G} \sum_{t=1}^{T} \frac{n}{N_{s_g}} (\lambda_t^{s_g})^2 \mathbb{V}\left(\mu_t(s_g, X) - \mu_t(0, X) | t^* = s_g\right) \\
&+ \sum_{g=1}^{G} \sum_{t=1}^{T} \sum_{t' \neq t} \frac{n}{N_{s_g}} \lambda_t^{s_g} \lambda_{t'}^{s_g} \mathbb{C}\mathrm{ov}(\mu_t(s_g, X) - \mu_t(0, X), \mu_{t'}(s_g, X) - \mu_{t'}(0, X) | t^* = s_g) \\
&+ \sum_{g=1}^{G} \sum_{t=1}^{T} \frac{n^2}{N_{s_g}^2} (\lambda_t^s)^2 \frac{1}{n} \sum_{i=n+1}^{2n} \mathbb{1}(t_{i-n}^* = s_g) \sigma_t^2(s_g, X_{i-n}) \\
&+ \sum_{g=1}^{G} \sum_{t=1}^{T} \sum_{t' \neq t} \frac{n^2}{N_{s_g}^2} \lambda_t^{s_g} \lambda_{t'}^{s_g} \frac{1}{n} \sum_{i=n+1}^{2n} \mathbb{1}(t_{i-n}^* = s_g) \sigma_{tt'}(s_g, X_{i-n}) \\
&+ \sum_{g=1}^{G} \sum_{t=1}^{T} \frac{n^2}{N_{s_g}^2} (\lambda_t^{\infty, s_g})^2 \frac{1}{n} \sum_{i=n+1}^{2n} \mathbb{1}(t_{i-n}^* = \infty) \frac{K_M^2(i-n, s_g)}{M^2} \sigma_t^2(0, X_{i-n}) \\
&+ \sum_{g=1}^{G} \sum_{t=1}^{T} \sum_{t' \neq t} \frac{n^2}{N_{s_g}^2} \lambda_t^{\infty, s_g} \lambda_{t'}^{\infty, s_g} \frac{1}{n} \sum_{i=n+1}^{2n} \mathbb{1}(t_{i-n}^* = \infty) \frac{K_M^2(i-n, s_g)}{M^2} \sigma_{tt'}(0, X_{i-n}) \\
&+ \sum_{g=1}^{G} \sum_{s' \neq s} \sum_{t=1}^{T} \frac{n}{N_{s_g}} \frac{n}{N_{s_{g'}}} \lambda_t^{\infty, s_g} \lambda_t^{\infty, s_{g'}} \frac{1}{n} \sum_{i=n+1}^{2n} \mathbb{1}(t_{i-n}^* = \infty) \frac{K_M(i-n, s_g) K_M(i-n, s_{g'})}{M^2} \sigma_t^2(0, X_{i-n}) \\
&+ \sum_{g=1}^{G} \sum_{s' \neq s} \sum_{t=1}^{T} \sum_{t' \neq t} \frac{n}{N_{s_g}} \frac{n}{N_{s_{g'}}} \lambda_t^{\infty, s_g} \lambda_{t'}^{\infty, s_{g'}} \frac{1}{n} \sum_{i=n+1}^{2n} \mathbb{1}(t_{i-n}^* = \infty) \frac{K_M(i-n, s_g) K_M(i-n, s_{g'})}{M^2} \sigma_{tt'}(0, X_{i-n})
\end{aligned}
$$

$$
\begin{aligned}
&\to_{\mathbb{P}} \sum_{g=1}^{G}\sum_{t=1}^{T} \frac{1}{p_{s_g}} (\lambda_t^{s_g})^2 Var\left(\mu_t(s_g, X) - \mu_t(0, X) \middle| t^* = s_g\right) \\
&+ \sum_{g=1}^{G}\sum_{t=1}^{T}\sum_{t'\neq t} \frac{1}{p_{s_g}} \lambda_t^{s_g}\lambda_{t'}^{s_g} Cov(\mu_t(s_g, X) - \mu_t(0, X), \mu_{t'}(s_g, X) - \mu_{t'}(0, X) | t^* = s_g) \\
&+ \sum_{g=1}^{G}\sum_{t=1}^{T} \frac{1}{p_{s_g}} (\lambda_t^{s_g})^2 \mathbb{E}[\sigma_t^2(s_g, X) | t^* = s_g] \\
&+ \sum_{g=1}^{G}\sum_{t=1}^{T}\sum_{t'\neq t} \frac{1}{p_{s_g}} \lambda_t^{s_g}\lambda_{t'}^{s_g} \mathbb{E}[\sigma_{tt'}(s_g, X) | t^* = s_g] \\
&+ \sum_{g=1}^{G}\sum_{t=1}^{T} \frac{1}{p_{s_g}^2} (\lambda_t^{\infty, s_g})^2 \left(\frac{1}{M}\mathbb{E}\left[e_{s_g}(X)\sigma_t^2(0, X)\right] + \frac{\alpha(M, q)}{M^2}\mathbb{E}\left[\frac{e_{s_g}(X)^2}{e_\infty(X)}\sigma_t^2(0, X)\right]\right) \\
&+ \sum_{g=1}^{G}\sum_{t=1}^{T}\sum_{t'\neq t} \frac{1}{p_{s_g}^2} \lambda_t^{\infty, s_g}\lambda_{t'}^{\infty, s_g} \left(\frac{1}{M}\mathbb{E}\left[e_{s_g}(X)\sigma_{tt'}(0, X)\right] + \frac{\alpha(M, q)}{M^2}\mathbb{E}\left[\frac{e_{s_g}(X)^2}{e_\infty(X)}\sigma_{tt'}(0, X)\right]\right) \\
&+ \sum_{g=1}^{G}\sum_{s'\neq s}\sum_{t=1}^{T} \frac{1}{p_{s_g}}\frac{1}{p_{s_{g'}}} \lambda_t^{\infty, s_g}\lambda_t^{\infty, s_{g'}} \left(\frac{\alpha(M, q)}{M^2}\mathbb{E}\left[\frac{e_{s_g}(X)e_{s_{g'}}(X)}{e_\infty(X)}\sigma_t^2(0, X)\right]\right) \\
&+ \sum_{g=1}^{G}\sum_{s'\neq s}\sum_{t=1}^{T}\sum_{t'\neq t} \frac{1}{p_{s_g}}\frac{1}{p_{s_{g'}}} \lambda_t^{\infty, s_g}\lambda_{t'}^{\infty, s_{g'}} \left(\frac{\alpha(M, q)}{M^2}\mathbb{E}\left[\frac{e_s(X)e_{s_{g'}}(X)}{e_\infty(X)}\sigma_{tt'}(0, X)\right]\right) \\
&= \lambda' V \lambda
\end{aligned}
$$

The stated convergence follows from the WLLN, Lemma B2 and Lemma B3. It remains to verify the relevant moment condition for the martingale central limit theorem. Under the assumptions in the paper, for $1 \leq i \leq n$,

$$
\begin{aligned}
&\mathbb{E}[Z_{n,i}^4 | \mathcal{F}_{n,i-1}] \\
&= \mathbb{E}\left[\left(\sum_{g=1}^{G} \sum_{t=1}^{T} Z_{n,i}^{s_g t} \right)^4 \middle| \mathcal{F}_{n,i-1} \right] \\
&\leq (GT)^3 \sum_{g=1}^{G} \sum_{t=1}^{T} \mathbb{E}[(Z_{n,i}^{s_g t})^4 | \mathcal{F}_{n,i-1}] \\
&= (GT)^3 \sum_{g=1}^{G} \sum_{t=1}^{T} \mathbb{E}\left[\left(\frac{n}{N_{s_g}} \frac{\lambda_t^{s_g}}{\sqrt{n}} \mathbb{1}(t_i^* = s_g) \left[\mu_t(s_g, X_i) - \mu_t(0, X_i) - (\mu_t(s_g) - \mathbb{E}[\mu_t(0, X_i) | t^* = s_g]) \right] \right)^4 \middle| \mathcal{F}_{n,i-1} \right] \\
&= (GT)^3 \sum_{g=1}^{G} \sum_{t=1}^{T} \frac{n^4}{N_{s_g}^4} \frac{(\lambda_t^{s_g})^4}{n^2} \mathbb{1}(t_i^* = s_g) \mathbb{E}\left[\left([\mu_t(s_g, X_i) - \mu_t(0, X_i) - (\mu_t(s_g) - \mathbb{E}[\mu_t(0, X_i) | t^* = s_g])] \right)^4 \middle| t_i^* = s_g \right] \\
&\leq (GT)^3 \sum_{g=1}^{G} \sum_{t=1}^{T} \frac{n^4}{N_{s_g}^4} \frac{1}{n^2} \mathbb{1}(t_i^* = s_g) C \\
&= \frac{G^3 T^4 C}{n^2} \sum_{g=1}^{G} \frac{n^4}{N_s^4} \mathbb{1}(t_i^* = s_g)
\end{aligned}
$$

The first inequality follows from Hölder's inequality, while the final inequality uses the facts that $\|\lambda\| = 1$ and that the fourth moment of the outcome is uniformly bounded. For $n + 1 \leq i \leq 2n$, an analogous argument gives:

$$
\begin{aligned}
&\mathbb{E}[Z_{n,i}^4|\mathcal{F}_{n,i-1}] \\
&= \mathbb{E}\left[\left(\sum_{g=1}^{G}\sum_{t=1}^{T} Z_{n,i}^{s_g t}\right)^4 \middle| \mathcal{F}_{n,i-1}\right] \\
&\leq (GT)^3 \sum_{g=1}^{G}\sum_{t=1}^{T} \mathbb{E}[(Z_{n,i}^{s_g t})^4|\mathcal{F}_{n,i-1}] \\
&= (GT)^3 \sum_{g=1}^{G}\sum_{t=1}^{T} \mathbb{E}\left[\left(\frac{n}{N_{s_g}}\frac{1}{\sqrt{n}}\left(\lambda_t^{s_g}\mathbb{1}(t_{i-n}^* = s_g) + \lambda_t^{\infty,s_g}\mathbb{1}(t_{i-n}^* = \infty)\frac{K_M(i-n,s_g)}{M}\right)\epsilon_{i-n,t}\right)^4 \middle| \mathcal{F}_{n,i-1}\right] \\
&= (GT)^3 \sum_{g=1}^{G}\sum_{t=1}^{T} \frac{n^4}{N_{s_g}^4}\frac{(\lambda_t^{s_g})^4}{n^2}\mathbb{1}(t_{i-n}^* = s_g)\mathbb{E}[\epsilon_{i-n,t}^4|t_{i-n}^* = s_g, X_{i-n}] \\
&+ (GT)^3 \sum_{g=1}^{G}\sum_{t=1}^{T} \frac{n^4}{N_{s_g}^4}\frac{(\lambda_t^{\infty,s_g})^4}{n^2}\mathbb{1}(t_{i-n}^* = \infty)\frac{K_M^4(i-n,s_g)}{M^4}\mathbb{E}[\epsilon_{i-n,t}^4|t_{i-n}^* = \infty, X_{i-n}] \\
&\leq (GT)^3 C' \sum_{g=1}^{G}\sum_{t=1}^{T} \frac{n^4}{N_{s_g}^4}\frac{1}{n^2}\left(\mathbb{1}(t_{i-n}^* = s_g) + \mathbb{1}(t_{i-n}^* = \infty)\frac{K_M^4(i-n,s_g)}{M^4}\right) \\
&= \frac{G^3T^4C'}{n^2}\sum_{t=1}^{T} \frac{n^4}{N_{s_g}^4}\left(\mathbb{1}(t_{i-n}^* = s_g) + \mathbb{1}(t_{i-n}^* = \infty)\frac{K_M^4(i-n,s_g)}{M^4}\right)
\end{aligned}
$$

Once again, the first inequality follows from Hölder's inequality, and the final inequality uses the facts that $\|\lambda\| = 1$ and that the fourth moment of the outcome is uniformly bounded. Therefore,

$$
\begin{aligned}
&\sum_{i=1}^{2n} \mathbb{E}[Z_{n,i}^4|\mathcal{F}_{n,i-1}] \\
&= \frac{G^3T^4C}{n}\sum_{g=1}^{G} \frac{n^4}{N_{s_g}^4}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^* = s_g)\right) \\
&+ \frac{G^3T^4C'}{n}\sum_{g=1}^{G} \frac{n^4}{N_{s_g}^4}\left(\frac{1}{n}\sum_{i=n+1}^{2n}\mathbb{1}(t_{i-n}^* = s_g) + \frac{1}{n}\sum_{i=n+1}^{2n}\mathbb{1}(t_{i-n}^* = \infty)\frac{K_M^4(i-n,s_g)}{M^4}\right) \\
&\to_{\mathbb{P}} 0.
\end{aligned}
$$

This convergence follows from the WLLN and the uniform boundedness of $\mathbb{E}[K_M^4(i-n,s_g)]$. The martingale central limit theorem therefore implies that $\sqrt{n}(\lambda'\overline{Y}) \to_{\mathcal{D}} \mathcal{N}(0, \lambda' V \lambda)$. Because this conclusion holds for every λ such that $\|\lambda\| = 1$, the Cramér-Wold device gives $\sqrt{n}\overline{Y} \to_{\mathcal{D}} \mathcal{N}(0, V)$. The theorem then follows from the fact that $\widehat{\delta}^{or} - \delta = L\sqrt{n}(\overline{Y}^{\infty,s_1}, \ldots, \overline{Y}^{\infty,s_G}, \overline{Y}^{s_1}, \ldots, \overline{Y}^{s_G})$ where

$L = \begin{pmatrix} -I_{GT} & I_{GT} \end{pmatrix}$. Finally, note that

$$\begin{aligned}
\sqrt{n}\left\|\widehat{\delta}^{bc} - \widehat{\delta}^{or}\right\| &= \sqrt{n}\left\|\widehat{B} - \widetilde{B}\right\| \\
&\leq GT\sqrt{n} \max_{\substack{s\in\mathcal{S}\setminus\{\infty\} \\ t=1,\dots,T}} \left| \frac{1}{N_s}\sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty)\frac{K_M(i,s)}{M} \right) r_t(X_i) \right| \\
&= GT\sqrt{n} \max_{\substack{s\in\mathcal{S}\setminus\{\infty\} \\ t=1,\dots,T}} \left| \frac{1}{N_s}\sum_i \mathbb{1}(t_i^* = s)\frac{1}{M}\sum_{j\in\mathcal{J}_M(i,\infty)} [r_t(X_i) - r_t(X_j)] \right| \\
&\leq GT\sqrt{n} \max_{\substack{s\in\mathcal{S}\setminus\{\infty\} \\ t=1,\dots,T}} \frac{1}{N_s}\sum_i \mathbb{1}(t_i^* = s)\frac{1}{M}\sum_{j\in\mathcal{J}_M(i,\infty)} |r_t(X_i) - r_t(X_j)| \\
&= o_{\mathbb{P}}(1)
\end{aligned}$$

under Condition (3), which completes the proof. □

Proof of Proposition 2

We split the proof into three parts, one for each convergence statement in the theorem. For brevity, we present the proofs for the case $t' = t$. The proofs for the case $t' \neq t$ follow a very similar structure and they are omitted, but they are available from the authors upon request.

First part: $\widehat{V}_{tt'}^{\infty,s_g} - V_{tt'}^{\infty,s_g} \to_{\mathbb{P}} 0$**.**

Adding and subtracting,

$$\begin{aligned}
\widehat{\sigma}_t^2(0, X_i) &= \frac{J}{J+1}\left(Y_{it} - \frac{1}{J}\sum_{j=1}^{J} Y_{l_j(i)t} \right)^2 \\
&= \frac{J}{J+1}\left(Y_{it} - \mu_t(0, X_i) - \frac{1}{J}\sum_{j=1}^{J} Y_{l_j(i)t} - \mu_t(0, X_{l_{j(i)}}) + \frac{1}{J}\sum_{j=1}^{J}[\mu_t(0, X_i) - \mu_t(0, X_{l_{j(i)}})] \right)^2 \\
&= \frac{J}{J+1}\left(\epsilon_{it} - \frac{1}{J}\sum_{j=1}^{J} \epsilon_{l_{j(i)}t} + \frac{1}{J}\sum_{j=1}^{J}[\mu_t(0, X_i) - \mu_t(0, X_{l_{j(i)}})] \right)^2 .
\end{aligned}$$

Therefore,

$$
\begin{aligned}
&\mathbb{E}[\widehat{\sigma}_t^2(0,X_i)|\mathbf{t}^*,\mathbf{X}] \\
&= \frac{J}{J+1}\left(\mathbb{E}[\epsilon_{it}^2|\mathbf{t}^*,\mathbf{X}] + \frac{1}{J^2}\sum_{j=1}^{J}\mathbb{E}\left[\epsilon_{l_{j(i)}t}^2\middle|\mathbf{t}^*,\mathbf{X}\right] + \mathbb{E}\left[\left(\frac{1}{J}\sum_{j=1}^{J}[\mu_t(0,X_i)-\mu_t(0,X_{l_{j(i)}})]\right)^2\middle|\mathbf{t}^*,\mathbf{X}\right]\right) \\
&= \frac{J}{J+1}\left(\sigma_t^2(0,X_i) + \frac{1}{J^2}\sum_{j=1}^{J}\sigma_t^2(0,X_{l_{j(i)}}) + \left(\frac{1}{J}\sum_{j=1}^{J}[\mu_t(0,X_i)-\mu_t(0,X_{l_{j(i)}})]\right)^2\right) \\
&= \frac{J}{J+1}\left(\frac{J+1}{J}\sigma_t^2(0,X_i) + \frac{1}{J^2}\sum_{j=1}^{J}\left(\sigma_t^2(0,X_{l_{j(i)}})-\sigma_t^2(0,X_i)\right) + \left(\frac{1}{J}\sum_{j=1}^{J}[\mu_t(0,X_i)-\mu_t(0,X_{l_{j(i)}})]\right)^2\right).
\end{aligned}
$$

The first equality follows from the fact all the interaction terms are zero after conditioning on $(\mathbf{t}^*,\mathbf{X})$. Next, by Lipschitz continuity of $\sigma_t^2(0,X_i)$, and $\mu_t(0,X_i)$,

$$
\begin{aligned}
&\max_{t_i^*=\infty}\left|\mathbb{E}[\widehat{\sigma}_t^2(0,X_i)|\mathbf{t}^*,\mathbf{X}] - \sigma_t^2(0,X_i)\right| \\
&= \max_{t_i^*=\infty}\left|\frac{1}{J(J+1)}\sum_{j=1}^{J}\sigma_t^2(0,X_{l_{j(i)}}) - \sigma_t^2(0,X_i) + \frac{1}{J(J+1)}\left(\sum_{j=1}^{J}\mu_t(0,X_i)-\mu_t(0,X_{l_{j(i)}})\right)^2\right| \\
&\le \max_{t_i^*=\infty}\frac{1}{J(J+1)}\sum_{j=1}^{J}\left|\sigma_t^2(0,X_{l_{j(i)}}) - \sigma_t^2(0,X_i)\right| + \frac{1}{J(J+1)}\left(\sum_{j=1}^{J}\left|\mu_t(0,X_i)-\mu_t(0,X_{l_{j(i)}})\right|\right)^2 \\
&\le \max_{t_i^*=\infty}\frac{1}{J(J+1)}\sum_{j=1}^{J}C\left\|X_i - X_{l_{j(i)}}\right\| + \frac{1}{J(J+1)}\left(\sum_{j=1}^{J}C\left\|X_i - X_{l_{j(i)}}\right\|\right)^2 \\
&\le \frac{C}{J+1}\max_{t_i^*=\infty}\left\|X_i - X_{l_{J(i)}}\right\| + C^2\max_{t_i^*=\infty}\left\|X_i - X_{l_{J(i)}}\right\|^2 = o_{\mathbb{P}}(1).
\end{aligned}
$$

This first inequality follows from the triangle inequality. The second inequality follows from the Lipschitz continuity of $\sigma_t^2(0,X_i)$, and $\mu_t(0,X_i)$. The last equality follows from Lemma B4. Therefore for any $\alpha > 0$,

$$
\begin{aligned}
&\left|\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{\alpha}(i,s_g)\left[\mathbb{E}[\widehat{\sigma}_t^2(0,X_i)|\mathbf{t}^*,\mathbf{X}] - \sigma_t^2(0,X_i)\right]\right| \\
&\le \frac{1}{N_{s_g}}\sum_{i=1}^{n}\left|\mathbb{1}(t_i^*=\infty)K_M^{\alpha}(i,s_g)\left[\mathbb{E}[\widehat{\sigma}_t^2(0,X_i)|\mathbf{t}^*,\mathbf{X}] - \sigma_t^2(0,X_i)\right]\right| \\
&\le \max_{t_i^*=\infty}\left|\widehat{\sigma}_t^2(0,X_i)|\mathbf{t}^*,\mathbf{X}] - \sigma_t^2(0,X_i)\right|\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbf{1}(t_i^*=\infty)K_M^{\alpha}(i,s_g) = o_{\mathbb{P}}(1).
\end{aligned}
$$

The first inequality follows from the triangular inequality. The last equality follows from the fact that $\max_{t_i^*=\infty} |\mathbb{E}[\widehat{\sigma}_t^2(0,X_i)|\mathbf{t}^*,\mathbf{X}] - \sigma_t^2(0,X_i)| = o_{\mathbb{P}}(1)$ and Lemma B5. Next we want to show that:

$$\frac{1}{N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = s) K_M^{\alpha}(i, s_g) \left[\mathbb{E}[\widehat{\sigma}_t^2(0,X_i)|\mathbf{t}^*,\mathbf{X}] - \widehat{\sigma}_t^2(0,X_i) \right] = o_{\mathbb{P}}(1)$$

By algebra,

$$\begin{aligned}
&\frac{J+1}{J} \frac{1}{N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) K_M^{\alpha}(i, s_g) \left[\widehat{\sigma}_t^2(0,X_i) - \mathbb{E}[\widehat{\sigma}_t^2(0,X_i)|\mathbf{t}^*,\mathbf{X}] \right] \\
&= \frac{1}{N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) K_M^{\alpha}(i, s_g) \left[\left(\epsilon_{it} - \frac{1}{J} \sum_{j=1}^{J} \epsilon_{l_{j(i)}t} + \frac{1}{J} \sum_{j=1}^{J} \mu_t(0,X_i) - \mu_t(0,X_{l_{j(i)}}) \right)^2 \right. \\
&\left. - \left(\sigma_t^2(0,X_i) + \frac{1}{J^2} \sum_{j=1}^{J} \sigma_t^2(0,X_{l_{j(i)}}) + \left(\frac{1}{J} \sum_{j=1}^{J} \mu_t(0,X_i) - \mu_t(0,X_{l_{j(i)}}) \right)^2 \right) \right] \\
&= \frac{1}{N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) K_M^{\alpha}(i, s_g) \left(\epsilon_{it}^2 - \sigma_t^2(0,X_i) \right) \\
&+ \frac{1}{N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) K_M^{\alpha}(i, s_g) \frac{1}{J^2} \sum_{j=1}^{J} \left(\epsilon_{l_{j(i)}t}^2 - \sigma_t^2(0,X_{l_{j(i)}t}) \right) \\
&+ \frac{2}{J^2 N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) K_M^{\alpha}(i, s_g) \sum_{j=1}^{J} \sum_{h>j} \epsilon_{l_{j(i)}t} \epsilon_{l_{h(i)}t} \\
&- \frac{2}{J N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) K_M^{\alpha}(i, s_g) \sum_{j=1}^{J} \epsilon_{it} \epsilon_{l_{j(i)}t} \\
&+ \frac{2}{J N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) K_M^{\alpha}(i, s_g) \epsilon_{it} \left(\sum_{j=1}^{J} \mu_t(0,X_i) - \mu_t(0,X_{l_{j(i)}}) \right) \\
&- \frac{2}{J^2 N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) K_M^{\alpha}(i, s_g) \sum_{j=1}^{J} \epsilon_{l_{j(i)}t} \left(\sum_{j'=1}^{J} \mu_t(0,X_i) - \mu_t(0,X_{l_{j'(i)}}) \right).
\end{aligned}$$

The second equality follows from direct calculation. Now we will show that the above six terms are all $o_{\mathbb{P}}(1)$.

First term:

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) K_M^{\alpha}(i, s_g) \left(\epsilon_{it}^2 - \sigma_t^2(0,X_i) \right) \middle| \mathbf{t}^*, \mathbf{X} \right] \\
&= \frac{1}{N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) K_M^{\alpha}(i, s_g) (\mathbb{E}[\epsilon_{it}^2|\mathbf{t}^*,\mathbf{X}] - \sigma_t^2(0,X_i)) = 0
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{V}\left[\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{\alpha}(i,s_g)\left(\epsilon_{it}^2-\sigma_t^2(0,X_i)\right)\middle|\,\mathbf{t}^*,\mathbf{X}\right]\\
&=\frac{1}{N_{s_g}^2}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{2\alpha}(i,s)\mathbb{V}\left[\left(\epsilon_{it}^2-\sigma_t^2(0,X_i)\right)\middle|\,\mathbf{t}^*,\mathbf{X}\right]\\
&\leq\frac{1}{N_{s_g}^2}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{2\alpha}(i,s)\mathbb{E}\left[\left(\epsilon_{it}^2-\sigma_t^2(0,X_i)\right)^2\middle|\,\mathbf{t}^*,\mathbf{X}\right]\\
&\leq\frac{2}{N_{s_g}^2}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{2\alpha}(i,s)\mathbb{E}\left[\epsilon_{it}^4+(\sigma_t^2(0,X_i))^2\middle|\,\mathbf{t}^*,\mathbf{X}\right]\\
&\leq\frac{2C}{N_{s_g}}\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{2\alpha}(i,s)=o_{\mathbb{P}}(1).
\end{aligned}$$

Thus, by the conditional Markov inequality and the bounded convergence theorem,

$$\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{\alpha}(i,s_g)\left(\epsilon_{it}^2-\sigma_t^2(0,X_i)\right)=o_{\mathbb{P}}(1).$$

Second term:

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{\alpha}(i,s_g)\frac{1}{J^2}\sum_{j=1}^{J}\left(\epsilon_{l_{j(i)}t}^2-\sigma_t^2(0,X_{l_{j(i)}t})\right)\middle|\,\mathbf{t}^*,\mathbf{X}\right]\\
&=\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{\alpha}(i,s_g)\frac{1}{J^2}\sum_{j=1}^{J}\mathbb{E}\left[\left(\epsilon_{l_{j(i)}t}^2-\sigma_t^2(0,X_{l_{j(i)}t})\right)\middle|\,\mathbf{t}^*,\mathbf{X}\right]=0
\end{aligned}$$

and

$$
\begin{aligned}
&\mathbb{V}\left[\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{\alpha}(i,s_g)\frac{1}{J^2}\sum_{j=1}^{J}\left(\epsilon^2_{l_{j(i)}t}-\sigma_t^2(0,X_{l_{j(i)}})\right)\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&=\frac{1}{J^4N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{\alpha}(i,s_g)K_M^{\alpha}(i',s_g)\sum_{j=1}^{J}\sum_{j'=1}^{J}\mathbb{C}\mathrm{ov}\left(\epsilon^2_{l_{j(i)}t},\ \epsilon^2_{l_{j'(i')}t}\middle|\mathbf{t}^*,\mathbf{X}\right)\\
&=\frac{1}{J^4N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{\alpha}(i,s_g)K_M^{\alpha}(i',s_g)\sum_{j=1}^{J}\sum_{j'=1}^{J}\mathbb{1}(l_{j(i)}=l_{j'(i')})\mathbb{V}\left(\epsilon^2_{l_{j(i)}t}\middle|\mathbf{t}^*,\mathbf{X}\right)\\
&\leq\frac{C}{J^4N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{\alpha}(i,s_g)K_M^{\alpha}(i',s_g)\sum_{j=1}^{J}\sum_{j'=1}^{J}\mathbb{1}(l_{j(i)}=l_{j'(i')})\\
&=\frac{C}{J^3N_{s_g}}\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{2\alpha}(i,s_g)\\
&+\frac{C}{J^4N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{\alpha}(i,s_g)K_M^{\alpha}(i',s_g)\sum_{j=1}^{J}\sum_{j'=1}^{J}\mathbb{1}(l_{j(i)}=l_{j'(i')})\\
&\leq\frac{C}{2J^4N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'\neq i}^{n}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)\left(K_M^{2\alpha}(i,s_g)+K_M^{2\alpha}(i',s_g)\right)\sum_{j=1}^{J}\sum_{j'=1}^{J}\mathbb{1}(l_{j(i)}=l_{j'(i')})+o_{\mathbb{P}}(1)\\
&=\frac{C}{J^4N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{2\alpha}(i,s_g)\sum_{j=1}^{J}\sum_{j'=1}^{J}\mathbb{1}(l_{j(i)}=l_{j'(i')})+o_{\mathbb{P}}(1)\\
&=\frac{C}{J^4N_{s_g}^2}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{2\alpha}(i,s_g)\sum_{j=1}^{J}\sum_{i'\neq i}^{n}\mathbb{1}(t_{i'}^*=\infty)\sum_{j'=1}^{J}\mathbb{1}(l_{j(i)}=l_{j'(i')})+o_{\mathbb{P}}(1)\\
&=\frac{C}{J^4N_{s_g}^2}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{2\alpha}(i,s_g)\sum_{j=1}^{J}K_J(l_{j(i)})\\
&\leq\frac{C}{2J^4N_{s_g}^2}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)\sum_{j=1}^{J}\left(K_M^{4\alpha}(i,s_g)+K_J^2(l_{j(i)})\right)+o_{\mathbb{P}}(1)\\
&=\frac{C}{2J^3N_{s_g}}\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{4\alpha}(i,s_g)+\frac{C}{2J^4N_{s_g}}\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_J^3(i,0)+o_{\mathbb{P}}(1)\\
&=o_{\mathbb{P}}(1)
\end{aligned}
$$

The second equality follows from the fact that the covariance between $\epsilon^2_{l_{j(i)}t}$ and $\epsilon^2_{l_{j'(i')}t}$ is nonzero if and only if $l_{j(i)}=l_{j'(i')}$ after conditioning on $\mathbf{t}^*,\mathbf{X}$. The second inequality follows from Lemma B5 and the inequality $ab\leq\frac{a^2+b^2}{2}$. The fourth equality follows from the fact that the two terms in summation are equal. The sixth equality follows from the definition of $K_J(l_j(i))$. The third inequality again follows from $ab\leq\frac{a^2+b^2}{2}$. For the last equality, the first term follows from direct calculation. The second term follows from the fact that these two expressions are two equivalent

ways to express the same summation. Again combining the conditional Markov inequality and bounded convergence,

$$\frac{1}{N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) K_M^{\alpha}(i, s_g) \frac{1}{J^2} \sum_{j=1}^{J} \left(\epsilon_{l_{j(i)}t}^2 - \sigma_t^2(0, X_{l_{j(i)}}) \right) = o_{\mathbb{P}}(1).$$

Third term:

$$\begin{aligned}
&\mathbb{E}\left[\frac{2}{J^2 N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) K_M^{\alpha}(i, s_g) \sum_{j=1}^{J} \sum_{h>j} \epsilon_{l_{j(i)}t} \epsilon_{l_{h(i)}t} \middle| \mathbf{t}^*, \mathbf{X} \right] \\
&= \frac{2}{J^2 N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) K_M^{\alpha}(i, s_g) \sum_{j=1}^{J} \sum_{h>j} \mathbb{E}\left[\epsilon_{l_{j(i)}t} \epsilon_{l_{h(i)}t} \middle| \mathbf{t}^*, \mathbf{X} \right] \\
&= \frac{2}{J^2 N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) K_M^{\alpha}(i, s_g) \sum_{j=1}^{J} \sum_{h>j} \mathbb{E}\left[\epsilon_{l_{j(i)}t} \middle| \mathbf{t}^*, \mathbf{X} \right] \mathbb{E}\left[\epsilon_{l_{h(i)}t} \middle| \mathbf{t}^*, \mathbf{X} \right] = 0
\end{aligned}$$

and

$$
\begin{aligned}
&\mathbb{V}\left[\frac{2}{J^2N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{\alpha}(i,s_g)\sum_{j=1}^{J}\sum_{h>j}\epsilon_{l_{j(i)}t}\epsilon_{l_{h(i)}t}\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&=\mathbb{E}\left[\left(\frac{2}{J^2N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{\alpha}(i,s_g)\sum_{j=1}^{J}\sum_{h>j}\epsilon_{l_{j(i)}t}\epsilon_{l_{h(i)}t}\right)^2\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&=\frac{4}{J^4N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{\alpha}(i,s_g)K_M^{\alpha}(i',s_g)\sum_{j=1}^{J}\sum_{h>j}\sum_{j'=1}^{J}\sum_{h'>j'}\mathbb{E}\left[\epsilon_{l_{j(i)}t}\epsilon_{l_{h(i)}t}\epsilon_{l_{j'(i')}t}\epsilon_{l_{h'(i')}t}\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&=\frac{4}{J^4N_{s_g}^2}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{2\alpha}(i,s_g)\sum_{j=1}^{J}\sum_{h>j}\mathbb{E}\left[\epsilon_{l_{j(i)}t}^2\epsilon_{l_{h(i)}t}^2\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&+\frac{4}{J^4N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{\alpha}(i,s_g)K_M^{\alpha}(i',s_g)\sum_{j=1}^{J}\sum_{h>j}\sum_{j'=1}^{J}\sum_{h'>j'}\mathbb{E}\left[\epsilon_{l_{j(i)}t}\epsilon_{l_{h(i)}t}\epsilon_{l_{j'(i')}t}\epsilon_{l_{h'(i')}t}\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&\leq\frac{8C(J-1)}{J^3N_{s_g}}\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{2\alpha}(i,s_g)\\
&+\frac{4C}{J^4N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{\alpha}(i,s_g)K_M^{\alpha}(i',s_g)\sum_{j=1}^{J}\sum_{h>j}\sum_{j'=1}^{J}\sum_{h'>j'}\mathbb{1}\left(\{l_j(i),l_h(i)\}=\{l_{j'}(i'),l_{h'}(i')\}\right)\\
&\leq\frac{4C}{J^4N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{\alpha}(i,s_g)K_M^{\alpha}(i',s_g)\\
&\sum_{j=1}^{J}\sum_{h>j}\sum_{j'=1}^{J}\sum_{h'>j'}\mathbb{1}\left(l_j(i)=l_{j'}(i')\right)+\mathbb{1}\left(l_j(i)=l_{h'}(i')\right)+o_{\mathbb{P}}(1)\\
&\leq\frac{4C}{J^4N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{\alpha}(i,s_g)K_M^{\alpha}(i',s_g)\\
&\sum_{j=1}^{J}\sum_{h=1}^{J}\sum_{j'=1}^{J}\sum_{h'=1}^{J}\mathbb{1}\left(l_j(i)=l_{j'}(i')\right)+\mathbb{1}\left(l_j(i)=l_{h'}(i')\right)+o_{\mathbb{P}}(1)\\
&=\frac{8C}{J^2N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{\alpha}(i,s_g)K_M^{\alpha}(i',s_g)\sum_{j=1}^{J}\sum_{j'=1}^{J}\mathbb{1}\left(l_j(i)=l_{j'}(i')\right)+o_{\mathbb{P}}(1)\\
&\leq\frac{4C}{JN_{s_g}}\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{4\alpha}(i,s_g)+\frac{4C}{J^2N_{s_g}}\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_J^3(i,0)\\
&=o_{\mathbb{P}}(1)
\end{aligned}
$$

The first inequality follows from the fact the fourth moment of Y is uniformly bounded and the fact that $\mathbb{E}\left[\epsilon_{l_{j(i)}t}\epsilon_{l_{h(i)}t}\epsilon_{l_{j'(i')}t}\epsilon_{l_{h'(i')}t}\middle|\mathbf{t}^*,\mathbf{X}\right]$ is nonzero if and only if $\{l_j(i),l_h(i)\}=\{l_{j'}(i'),l_{h'}(i')\}$. The second inequality follows from the fact that the event $\{l_j(i),l_h(i)\}=\{l_{j'}(i'),l_{h'}(i')\}$ is included in the event $\{l_j(i)=l_{j'}(i')\}\cup\{l_j(i)=l_{h'}(i')\}$. The fourth equality follows from the fact the two

terms in the summation are symmetric. The last inequality follows from the same argument in the second term. As before, this implies

$$\frac{2}{J^2 N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) K_M^{\alpha}(i, s_g) \sum_{j=1}^{J} \sum_{h>j} \epsilon_{l_{j(i)}t} \epsilon_{l_{h(i)}t} = o_{\mathbb{P}}(1).$$

Fourth term:

$$\begin{aligned}
&\mathbb{E}\left[\frac{2}{J N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) K_M^{\alpha}(i, s_g) \sum_{j=1}^{J} \epsilon_{it} \epsilon_{l_{j(i)}t} \middle| \mathbf{t}^*, \mathbf{X} \right] \\
&= \frac{2}{J N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) K_M^{\alpha}(i, s_g) \sum_{j=1}^{J} \mathbb{E}\left[\epsilon_{it} \epsilon_{l_{j(i)}t} \middle| \mathbf{t}^*, \mathbf{X} \right] \\
&= \frac{2}{J N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty) K_M^{\alpha}(i, s_g) \sum_{j=1}^{J} \mathbb{E}\left[\epsilon_{it} \middle| \mathbf{t}^*, \mathbf{X} \right] E\left[\epsilon_{l_{j(i)}t} \middle| \mathbf{t}^*, \mathbf{X} \right] = 0
\end{aligned}$$

and

$$
\begin{aligned}
&\mathbb{V}\left[\frac{2}{JN_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{\alpha}(i,s_g)\sum_{j=1}^{J}\epsilon_{it}\epsilon_{l_{j(i)}t}\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&=\mathbb{E}\left[\left(\frac{2}{JN_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{\alpha}(i,s_g)\sum_{j=1}^{J}\epsilon_{it}\epsilon_{l_{j(i)}t}\right)^2\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&=\frac{4}{J^2N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{\alpha}(i,s_g)K_M^{\alpha}(i',s_g)\sum_{j=1}^{J}\sum_{j'=1}^{J}\mathbb{E}\left[\epsilon_{it}\epsilon_{i't}\epsilon_{l_{j(i)}t}\epsilon_{l_{j'(i')}t}\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&=\frac{4}{J^2N_{s_g}^2}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{2\alpha}(i,s_g)\sum_{j=1}^{J}\mathbb{E}\left[\epsilon_{it}^2\epsilon_{l_{j(i)}t}^2\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&+\frac{4}{J^2N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{\alpha}(i,s_g)K_M^{\alpha}(i',s_g)\sum_{j=1}^{J}\sum_{j'=1}^{J}\mathbb{E}\left[\epsilon_{it}\epsilon_{i't}\epsilon_{l_{j(i)}t}\epsilon_{l_{j'(i')}t}\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&\leq\frac{4C}{JN_{s_g}}\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{2\alpha}(i,s_g)\\
&+\frac{4C}{J^2N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{\alpha}(i,s_g)K_M^{\alpha}(i',s_g)\sum_{j=1}^{J}\sum_{j'=1}^{J}\mathbb{1}(l_j(i)=i')\mathbb{1}(l_{j'}(i')=i)\\
&\leq\frac{2C}{J^2N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)\left(K_M^{2\alpha}(i,s_g)+K_M^{2\alpha}(i',s_g)\right)\sum_{j=1}^{J}\sum_{j'=1}^{J}\mathbb{1}(l_j(i)=i')\mathbb{1}(l_{j'}(i')=i)+o_{\mathbb{P}}(1)\\
&=\frac{4C}{J^2N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{2\alpha}(i,s_g)\sum_{j=1}^{J}\sum_{j'=1}^{J}\mathbb{1}(l_j(i)=i')\mathbb{1}(l_{j'}(i')=i)+o_{\mathbb{P}}(1)\\
&\leq\frac{4C}{JN_{s_g}^2}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{2\alpha}(i,s_g)\sum_{j=1}^{J}\mathbb{1}(l_j(i)=i')+o_{\mathbb{P}}(1)\\
&=\frac{4C}{JN_{s_g}^2}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{2\alpha}(i,s_g)\sum_{i'\neq i}\mathbb{1}(t_{i'}^*=\infty)\sum_{j=1}^{J}\mathbb{1}(l_j(i)=i')+o_{\mathbb{P}}(1)\\
&=\frac{4C}{N_{s_g}}\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{2\alpha}(i,s_g)+o_{\mathbb{P}}(1)\\
&=o_{\mathbb{P}}(1)
\end{aligned}
$$

The first inequality follows from the fact that the fourth moment of Y is uniformly bounded and the fact that $\mathbb{E}\left[\epsilon_{it}\epsilon_{i't}\epsilon_{l_{j(i)}t}\epsilon_{l_{j'(i')}t}\middle|\mathbf{t}^*,\mathbf{X}\right]$ is non-zero if and only if $l_{j(i)}=i'$ and $l_{j'(i')}=i$. The second inequality follows from the inequality $ab\leq\frac{a^2+b^2}{2}$. The fifth equality follows from the fact that the two terms in summation are equal. The third inequality follows from the fact that $\mathbb{1}(l_j(i)=i')\mathbb{1}(l_{j'}(i')=i)\leq\mathbb{1}(l_j(i)=i')$. The second last equality follows from the fact that

$\sum_{i' \neq i} \mathbb{1}(t^*_{i'} = \infty) \sum_{j=1}^{J} \mathbb{1}(l_j(i) = i') = \sum_{j=1}^{J} \sum_{i' \neq i} \mathbb{1}(t^*_{i'} = \infty) \mathbb{1}(l_j(i) = i') = J$. Thus,

$$\frac{2}{JN_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t^*_i = \infty) K^{\alpha}_M(i, s_g) \sum_{j=1}^{J} \epsilon_{it} \epsilon_{l_{j(i)}t} = o_{\mathbb{P}}(1).$$

Before proving the results for the fifth and sixth term, first define:

$$\Phi_{J,i} = \sum_{j=1}^{J} \mu_t(0, X_i) - \mu_t(0, X_{l_{j(i)}})$$

and note that:

$$\begin{aligned} |\Phi_{J,i}| &= \left| \sum_{j=1}^{J} \mu_t(0, X_i) - \mu_t(0, X_{l_{j(i)}}) \right| \leq \sum_{j=1}^{J} \left| \mu_t(0, X_i) - \mu_t(0, X_{l_{j(i)}}) \right| \\ &\leq \sum_{j=1}^{J} |\mu_t(0, X_i)| + \left| \mu_t(0, X_{l_{j(i)}}) \right| \leq C. \end{aligned}$$

The first and second inequality follow from triangle inequality. The third inequality follows from the Lipschitz continuity of the conditional mean of the outcome and the bounded support of X.

Fifth term:

$$\mathbb{E}\left[\frac{2}{JN_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t^*_i = \infty) K^{\alpha}_M(i, s_g) \epsilon_{it} \Phi_{J,i} \middle| \mathbf{t}^*, \mathbf{X} \right] = \frac{2}{JN_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t^*_i = \infty) K^{\alpha}_M(i, s_g) \Phi_{J,i} \mathbb{E}\left[\epsilon_{it} \middle| \mathbf{t}^*, \mathbf{X} \right] = 0$$

and

$$\begin{aligned} &\mathbb{V}\left[\frac{2}{JN_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t^*_i = \infty) K^{\alpha}_M(i, s_g) \epsilon_{it} \Phi_{J,i} \middle| \mathbf{t}^*, \mathbf{X} \right] \\ &= \frac{4}{J^2 N^2_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t^*_i = \infty) K^{2\alpha}_M(i, s_g) \Phi^2_{J,i} Var\left(\epsilon_{it} \middle| \mathbf{t}^*, \mathbf{X} \right) \\ &\leq \frac{4C}{J^2 N_{s_g}} \frac{1}{N_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t^*_i = \infty) K^{2\alpha}_M(i, s_g) \\ &= o_{\mathbb{P}}(1) \end{aligned}$$

and thus

$$\frac{2}{JN_{s_g}} \sum_{i=1}^{n} \mathbb{1}(t^*_i = \infty) K^{\alpha}_M(i, s_g) \epsilon_{it} = o_{\mathbb{P}}(1).$$

Sixth term:

$$\mathbb{E}\left[\frac{2}{J^2 N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{\alpha}(i,s_g)\Phi_{J,i}\sum_{j=1}^{J}\epsilon_{l_{j(i)}t}\middle|\mathbf{t}^*,\mathbf{X}\right]$$
$$=\frac{2}{J^2 N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{\alpha}(i,s_g)\Phi_{J,i}\sum_{j=1}^{J}\mathbb{E}[\epsilon_{l_{j(i)}t}|\mathbf{t}^*,\mathbf{X}]=0$$

and

$$\begin{aligned}
&\mathbb{V}\left[\frac{2}{J^2 N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{\alpha}(i,s_g)\Phi_{J,i}\sum_{j=1}^{J}\epsilon_{l_{j(i)}t}\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&=\frac{4}{J^4 N_{s_g}^2}\mathbb{E}\left[\left(\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{\alpha}(i,s_g)\Phi_{J,i}\sum_{j=1}^{J}\epsilon_{l_{j(i)}t}\right)^2\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&=\frac{4}{J^4 N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{\alpha}(i,s_g)K_M^{\alpha}(i',s_g)\Phi_{J,i}\Phi_{J,i'}\sum_{j=1}^{J}\sum_{j'=1}^{J}\mathbb{E}\left[\epsilon_{l_{j(i)}t}\epsilon_{l_{j'(i')}t}\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&=\frac{4}{J^4 N_{s_g}^2}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{2\alpha}(i,s_g)\Phi_{J,i}^2\sum_{j=1}^{J}\mathbb{E}\left[\epsilon_{l_{j(i)}t}^2\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&+\frac{4}{J^4 N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{\alpha}(i,s_g)K_M^{\alpha}(i',s_g)\Phi_{J,i}\Phi_{J,i'}\sum_{j=1}^{J}\sum_{j'=1}^{J}\mathbb{E}\left[\epsilon_{l_{j(i)}t}\epsilon_{l_{j'(i')}t}\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&\leq\frac{4C}{J^3 N_{s_g}}\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{2\alpha}(i,s_g)\\
&+\frac{4C}{J^4 N_{s_g}^2}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=\infty)\mathbb{1}(t_{i'}^*=\infty)K_M^{\alpha}(i,s_g)K_M^{\alpha}(i',s_g)\sum_{j=1}^{J}\sum_{j'=1}^{J}\mathbb{1}(l_j(i)=l_{j'}(i'))\\
&\leq\frac{2C}{J^3 N_{s_g}}\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{4\alpha}(i,s_g)+\frac{2C}{J^4 N_{s_g}}\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_J^3(i,0)+o_{\mathbb{P}}(1)\\
&=o_{\mathbb{P}}(1)
\end{aligned}$$

The first inequality follows from the fact that the conditional variance of Y and $\Phi_{J,i}$ are uniformly bounded, and $\mathbb{E}\left[\epsilon_{l_{j(i)}t}\epsilon_{l_{j'(i')}t}\middle|\mathbf{t}^*,\mathbf{X}\right]$ is non-zero if and only if $l_j(i)=l_{j'}(i')$. The last inequality follows from the same argument as the one in the second term. Thus,

$$\frac{2}{J^2 N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^{\alpha}(i,s_g)\Phi_{J,i}\sum_{j=1}^{J}\epsilon_{l_{j(i)}t}=o_{\mathbb{P}}(1).$$

Therefore,

$$\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s)K_M^{\alpha}(i,s_g)\left[\widehat{\sigma}_t^2(0,X_i)-\sigma_t^2(0,X_i)\right]$$
$$=\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s)K_M^{\alpha}(i,s_g)\left(\left[\widehat{\sigma}_t^2(0,X_i)-\mathbb{E}[\widehat{\sigma}_t^2(0,X_i)|\mathbf{t}^*,\mathbf{X}]\right]+\left[\mathbb{E}[\widehat{\sigma}_t^2(0,X_i)|\mathbf{t}^*,\mathbf{X}]-\sigma_t^2(0,X_i)\right]\right)$$
$$=o_{\mathbb{P}}(1).$$

Set $\alpha=2$ and apply the Slutsky's theorem and Lemma B2 to obtain:

$$\widehat{V}_{tt}^{\infty,s_g}=\frac{n}{N_{s_g}}\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s)K_M^2(i,s_g)\widehat{\sigma}_t^2(0,X_i)=\frac{n}{N_{s_g}}\frac{1}{N_{s_g}}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s)K_M^2(i,s_g)\sigma_t^2(0,X_i)+o_{\mathbb{P}}(1)$$
$$\to_{\mathbb{P}}\frac{1}{p_{s_g}^2}\left(\frac{1}{M}\mathbb{E}\left[e_{s_g}(X)\sigma_t^2(0,X)\right]+\frac{\alpha(M,q)}{M^2}\mathbb{E}\left[\frac{e_{s_g}(X)^2}{e_{\infty}(X)}\sigma_t^2(0,X)\right]\right)=V_{tt}^{\infty,s_g}.$$

This proves the first statement.

Second part: $\widehat{V}_{tt'}^{\infty,s_gs_{g'}}-V_{tt'}^{\infty,s_gs_{g'}}\to_{\mathbb{P}}0$**.**

For any $\alpha>1$, by C-r inequality,

$$(K_M(i,s_g)+K_M(i,s_{g'}))^{\alpha}\leq 2^{\alpha-1}\left(K_M(i,s_g)^{\alpha}+K_M(i,s_{g'})^{\alpha}\right)$$

By Lemma 3 in Abadie and Imbens (2006) and Lemma B5,

$$\mathbb{E}\left[(K_M(i,s_g)+K_M(i,s_{g'}))^{\alpha}\right]\leq 2^{\alpha-1}\left(\mathbb{E}\left[K_M(i,s_g)^{\alpha}\right]+\left[K_M(i,s_{g'})^{\alpha}\right]\right)=o_{\mathbb{P}}(1)$$
$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)\left(K_M(i,s_g)+K_M(i,s_{g'})\right)^{\alpha}\leq\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)2^{\alpha-1}\left(K_M(i,s_g)^{\alpha}+K_M(i,s_{g'})^{\alpha}\right)=o_{\mathbb{P}}(1)$$

By the same argument as the proof of the first part, for any $\alpha>0$:

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)\left(K_M(i,s_g)+K_M(i,s_{g'})\right)^{\alpha}\left(\widehat{\sigma}_t^2(0,X_i)-\sigma_t^2(0,X_i)\right)=o_{\mathbb{P}}(1)$$

Therefore, set $\alpha = 2$,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^* = \infty)K_M(i,s_g)K_M(i,s_{g'})\left(\widehat{\sigma}_t^2(0,X_i) - \sigma_t^2(0,X_i)\right) \\
&= \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^* = \infty)\left(K_M(i,s_g) + K_M(i,s_{g'})\right)^2\left(\widehat{\sigma}_t^2(0,X_i) - \sigma_t^2(0,X_i)\right)\right. \\
&-\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^* = \infty)K_M^2(i,s_g)\left(\widehat{\sigma}_t^2(0,X_i) - \sigma_t^2(0,X_i)\right) \\
&\left.-\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^* = \infty)K_M^2(i,s_{g'})\left(\widehat{\sigma}_t^2(0,X_i) - \sigma_t^2(0,X_i)\right)\right) \\
&= o_{\mathbb{P}}(1)
\end{aligned}$$

By Lemma B3,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^* = \infty)\frac{K_M(i,s_g)K_M(i,s_{g'})}{M^2}\sigma_t^2(0,X_i) \to_{\mathbb{P}} \frac{\alpha(M,q)}{M^2}\mathbb{E}\left[\frac{e_{s_g}(X)e_{s_{g'}}(X)}{e_\infty(X)}\sigma_t^2(0,X)\right]$$

Therefore,

$$\begin{aligned}
\widehat{V}_{tt}^{\infty,s_g s_{g'}} &= \frac{n}{N_{s_g}}\frac{n}{N_{s_{g'}}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^* = \infty)\frac{K_M(i,s_g)K_M(i,s_{g'})}{M^2}\widehat{\sigma}_t^2(0,X_i) \\
&\to_{\mathbb{P}} \frac{1}{p_{s_g}}\frac{1}{p_{s_{g'}}}\frac{\alpha(M,q)}{M^2}\mathbb{E}\left[\frac{e_{s_g}(X)e_{s_{g'}}(X)}{e_\infty(X)}\sigma_t^2(0,X)\right] = V_{tt}^{\infty,s_g s_{g'}}
\end{aligned}$$

This proves the second part.

Third part: $\widehat{V}^{s_g}_{tt'} - V^{s_g}_{tt'} \to_{\mathbb{P}} 0$**.**

First,

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)\left(Y_{it} - \widehat{Y}_{it} - \left(\widetilde{Y}_t^{s_g} - \widetilde{Y}_t^{\infty,s_g}\right)\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)\Big(Y_{it} - \widehat{Y}_{it} - (\mu_t(s_g) - \mathbb{E}[\mu_t(0,X)|t^* = s_g]) \\
&- \left[\left(\widetilde{Y}_t^{s_g} - \widetilde{Y}_t^{\infty,s_g}\right) - (\mu_t(s_g) - \mathbb{E}[\mu_t(0,X)|t^* = s_g])\right]\Big)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)\left(Y_{it} - \frac{1}{M}\sum_{m=1}^{M} Y_{j_m(i,s_g)t} - (\mu_t(s_g) - \mathbb{E}[\mu_t(0,X)|t^* = s_g])\right)^2 \\
&+ \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)\left[\left(\widetilde{Y}_t^{s_g} - \widetilde{Y}_t^{\infty,s_g}\right) - (\mu_t(s_g) - \mathbb{E}[\mu_t(0,X)|t^* = s_g])\right]^2 \\
&- \frac{2N_{s_g}}{n}\left[\left(\widetilde{Y}_t^{s_g} - \widetilde{Y}_t^{\infty,s_g}\right) - (\mu_t(s_g) - \mathbb{E}[\mu_t(0,X)|t^* = s_g])\right]\frac{1}{N_{s_g}}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g) \\
&\left(Y_{it} - \frac{1}{M}\sum_{m=1}^{M} Y_{j_m(i,s_g)t} - (\mu_t(s_g) - \mathbb{E}[\mu_t(0,X)|t^* = s_g])\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)\left(Y_{it} - \frac{1}{M}\sum_{m=1}^{M} Y_{j_m(i,s_g)t} - (\mu_t(s_g) - \mathbb{E}[\mu_t(0,X)|t^* = s_g])\right)^2 \\
&- \frac{N_{s_g}}{n}\left[\left(\widetilde{Y}_t^{s_g} - \widetilde{Y}_t^{\infty,s_g}\right) - (\mu_t(s_g) - \mathbb{E}[\mu_t(0,X)|t^* = s_g])\right]^2
\end{aligned}
$$

Notice that by the WLLN and Lemma B2,

$$
\begin{aligned}
\widetilde{Y}_t^{s_g} - \widetilde{Y}_t^{\infty,s_g} &= \frac{1}{N_{s_g}}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)Y_{it} - \frac{1}{N_{s_g}}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)\frac{1}{M}\sum_{m=1}^{M} Y_{j_m(i,s_g)t} \\
&= \frac{1}{N_{s_g}}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)Y_{it} - \frac{1}{N_{s_g}}\sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty)\frac{K_M(i,s_g)}{M}Y_{it} \\
&\to_{\mathbb{P}} \mu_t(s_g) - \mathbb{E}[\mu_t(0,X)|t^* = s_g].
\end{aligned}
$$

Therefore,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(Y_{it}-\widehat{Y}_{it}-\left(\widetilde{Y}_t^{s_g}-\widetilde{Y}_t^{\infty,s_g}\right)\right)^2$$
$$=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(Y_{it}-\frac{1}{M}\sum_{m=1}^{M}Y_{j_m(i,s_g)t}-(\mu_t(s_g)-\mathbb{E}[\mu_t(0,X)|t^*=s_g])\right)^2+o_{\mathbb{P}}(1)$$
$$=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(\mu_t(s_g,X_i)+\epsilon_{it}-\frac{1}{M}\sum_{m=1}^{M}\left[\mu_t(0,X_{j_m(i,s_g)})+\epsilon_{j_m(i,s_g)t}\right)\right]-(\mu_t(s_g)-\mathbb{E}[\mu_t(0,X)|t^*=s_g])\Bigg)^2$$
$$=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(\mu_t(s_g,X_i)-\frac{1}{M}\sum_{m=1}^{M}\mu_t(0,X_{j_m(i,s_g)})-(\mu_t(s_g)-\mathbb{E}[\mu_t(0,X)|t^*=s_g])\right)^2$$
$$+\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(\epsilon_{it}-\frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_m(i,s_g)t}\right)^2$$
$$+\frac{2}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(\mu_t(s_g,X_i)-\frac{1}{M}\sum_{m=1}^{M}\mu_t(0,X_{j_m(i,s_g)})-(\mu_t(s_g)-\mathbb{E}[\mu_t(0,X)|t^*=s_g])\right)$$
$$\left(\epsilon_{it}-\frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_m(i,s_g)t}\right)$$

Now we analyze each of the three terms from the last equality.

First term: adding and subtracting,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(\mu_t(s_g,X_i)-\frac{1}{M}\sum_{m=1}^{M}\mu_t(0,X_{j_m(i,s_g)})-(\mu_t(s_g)-\mathbb{E}[\mu_t(0,X)|t^*=s_g])\right)^2$$
$$=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(\mu_t(s_g,X_i)-\mu_t(0,X_i)+\frac{1}{M}\sum_{m=1}^{M}\left[\mu_t(0,X_i)-\mu_t(0,X_{j_m(i,s_g)})\right]\right.$$
$$\left.-(\mu_t(s_g)-\mathbb{E}[\mu_t(0,X)|t^*=s_g])\right)^2$$
$$=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(\mu_t(s_g,X_i)-\mu_t(0,X_i)-(\mu_t(s_g)-\mathbb{E}[\mu_t(0,X)|t^*=s_g])\right)^2$$
$$+\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(\frac{1}{M}\sum_{m=1}^{M}\left[\mu_t(0,X_i)-\mu_t(0,X_{j_m(i,s_g)})\right]\right)^2$$
$$+\frac{2}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(\mu_t(s_g,X_i)-\mu_t(0,X_i)-(\mu_t(s_g)-\mathbb{E}[\mu_t(0,X)|t^*=s_g])\right)$$
$$\left(\frac{1}{M}\sum_{m=1}^{M}\left[\mu_t(0,X_i)-\mu_t(0,X_{j_m(i,s_g)})\right]\right).$$

Notice that:

$$\begin{aligned}\left|\frac{1}{M}\sum_{m=1}^{M}\left[\mu_t(0,X_i)-\mu_t(0,X_{j_m(i,s_g)})\right]\right| &\leq \frac{1}{M}\sum_{m=1}^{M}\left|\mu_t(0,X_i)-\mu_t(0,X_{j_m(i,s_g)})\right| \\ &\leq \frac{C}{M}\sum_{m=1}^{M}\left\|X_i-X_{j_m(i,s_g)}\right\| \\ &\leq C\left\|X_i-X_{j_M(i,s_g)}\right\|\end{aligned}$$

The first two inequalities follow from the triangle inequality. The third inequality follows from Lipschitz continuity of $\mu_t(0,X_i)$. This implies that:

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(\frac{1}{M}\sum_{m=1}^{M}\left[\mu_t(0,X_i)-\mu_t(0,X_{j_m(i,s_g)})\right]\right)^2 \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(C\left\|X_i-X_{j_M(i,s_g)}\right\|\right)^2 \\ &\leq \left(\max_{t_i^*=s_g} C\left\|X_i-X_{j_M(i,s_g)}\right\|\right)^2 = o_{\mathbb{P}}(1)\end{aligned}$$

where last line follows from Lemma B4. Next,

$$\begin{aligned}&\left|\frac{2}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(\mu_t(s_g,X_i)-\mu_t(0,X_i)-(\mu_t(s_g)-\mathbb{E}[\mu_t(0,X)|t^*=s_g])\right)\right. \\ &\left.\left(\frac{1}{M}\sum_{m=1}^{M}\left[\mu_t(0,X_i)-\mu_t(0,X_{j_m(i,s_g)})\right]\right)\right| \\ &\leq \max_{t_i^*=s_g} C\left\|X_i-X_{j_M(i,s_g)}\right\|\frac{2}{n}\sum_{i=1}^{n}\left|\mathbb{1}(t_i^*=s_g)\left(\mu_t(s_g,X_i)-\mu_t(0,X_i)-(\mu_t(s_g)-\mathbb{E}[\mu_t(0,X)|t^*=s_g])\right)\right| \\ &= o_{\mathbb{P}}(1).\end{aligned}$$

The first term of the above equation is $o_{\mathbb{P}}(1)$ by Lemma B4, and the second term is bounded by Lipschitz continuity of the conditional mean function. Combining the results above, we have:

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(\mu_t(s_g,X_i)-\frac{1}{M}\sum_{m=1}^{M}\mu_t(0,X_{j_m(i,s_g)})-(\mu_t(s_g)-\mathbb{E}[\mu_t(0,X)|t^*=s_g])\right)^2 \\ &=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(\mu_t(s_g,X_i)-\mu_t(0,X_i)-(\mu_t(s_g)-\mathbb{E}[\mu_t(0,X)|t^*=s_g])\right)^2+o_{\mathbb{P}}(1) \\ &\to_{\mathbb{P}}\mathbb{E}\left[\mathbb{1}(t^*=s_g)\left(\mu_t(s_g,X)-\mu_t(0,X)-(\mu_t(s_g)-\mathbb{E}[\mu_t(0,X)|t^*=s_g])\right)^2\right] \\ &=p_{s_g}\mathbb{V}\left[\mu_t(s_g,X)-\mu_t(0,X)|t^*=s_g\right].\end{aligned}$$

Second term: by algebra,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(\epsilon_{it}-\frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_m(i,s_g)t}\right)^2\\
&=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(\epsilon_{it}^2+\frac{1}{M^2}\sum_{m=1}^{M}\epsilon_{j_m(i,s_g)t}^2\right)+\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\frac{2}{M^2}\sum_{m=1}^{M}\sum_{r>m}\epsilon_{j_m(i,s_g)t}\epsilon_{j_r(i,s_g)t}\\
&-\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\frac{2}{M}\sum_{m=1}^{M}\epsilon_{it}\epsilon_{j_m(i,s_g)t}
\end{aligned}$$

By a WLLN,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(\epsilon_{it}^2-\sigma_t^2(s_g,X_i)\right)&\to_{\mathbb{P}} p_{s_g}\left(\mathbb{E}[\epsilon_{it}^2|t_i^*=s_g]-\mathbb{E}[\sigma_t^2(s_g,X_i)|t_i^*=s_g]\right)\\
&=p_{s_g}\left(\mathbb{E}[\mathbb{E}[\epsilon_{it}^2|t_i^*=s_g,X_i]|t_i^*=s_g]-\mathbb{E}[\sigma_t^2(s_g,X_i)|t_i^*=s_g]\right)=0.
\end{aligned}$$

From Lemma B2, we have:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\frac{1}{M^2}\sum_{m=1}^{M}\left(\epsilon_{j_m(i,s_g)t}^2-\sigma_t^2(0,X_{j_m(i,s_g)})\right)&=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)\frac{K_M(i,s_g)}{M^2}\left(\epsilon_{it}^2-\sigma_t^2(0,X_i)\right)\\
&\to_{\mathbb{P}} p_{s_g}\mathbb{E}[\mathbb{E}[\epsilon_{it}^2-\sigma_t^2(0,X_i)|t_i^*=\infty,X_i]|t_i^*=s_g]=0
\end{aligned}$$

For the third term,

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\frac{2}{M^2}\sum_{m=1}^{M}\sum_{r>m}\epsilon_{j_m(i,s_g)t}\epsilon_{j_r(i,s_g)t}\middle|\,\mathbf{t}^*,\mathbf{X}\right]\\
&=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\frac{2}{M^2}\sum_{m=1}^{M}\sum_{r>m}\mathbb{E}\left[\epsilon_{j_m(i,s_g)t}\middle|\,\mathbf{t}^*,\mathbf{X}\right]\mathbb{E}\left[\epsilon_{j_r(i,s_g)t}\middle|\,\mathbf{t}^*,\mathbf{X}\right]=0
\end{aligned}$$

Next consider its conditional variance,

$$
\begin{aligned}
&\mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\frac{2}{M^2}\sum_{m=1}^{M}\sum_{r>m}\epsilon_{j_m(i,s_g)t}\epsilon_{j_r(i,s_g)t}\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&=\frac{4}{n^2M^4}\mathbb{E}\left[\left(\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\sum_{m=1}^{M}\sum_{r>m}\epsilon_{j_m(i,s_g)t}\epsilon_{j_r(i,s_g)t}\right)^2\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&=\frac{4}{n^2M^4}\sum_{i=1}^{n}\mathbb{E}\left[\left(\mathbb{1}(t_i^*=s_g)\sum_{m=1}^{M}\sum_{r>m}\epsilon_{j_m(i,s_g)t}\epsilon_{j_r(i,s_g)t}\right)^2\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&+\frac{4}{n^2M^4}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{E}\left[\left(\mathbb{1}(t_i^*=s_g)\sum_{m=1}^{M}\sum_{r>m}\epsilon_{j_m(i,s_g)t}\epsilon_{j_r(i,s_g)t}\right)\left(\mathbb{1}(t_{i'}^*=s_g)\sum_{m'=1}^{M}\sum_{r'>m'}\epsilon_{j_{m'}(i',s_g)t}\epsilon_{j_{r'}(i',s_g)t}\right)\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&=\frac{4}{n^2M^4}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\sum_{m=1}^{M}\sum_{r>m}\mathbb{E}\left[\epsilon_{j_m(i,s_g)t}^2\epsilon_{j_r(i,s_g)t}^2\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&+\frac{4}{n^2M^4}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=s_g)\mathbb{1}(t_{i'}^*=s_g)\sum_{m=1}^{M}\sum_{r>m}\sum_{m'=1}^{M}\sum_{r'>m'}\mathbb{E}\left[\epsilon_{j_m(i,s_g)t}\epsilon_{j_r(i,s_g)t}\epsilon_{j_{m'}(i',s_g)t}\epsilon_{j_{r'}(i',s_g)t}\middle|\mathbf{t}^*,\mathbf{X}\right]\\
&\leq\frac{4(\overline{\sigma}^2)^2}{n^2M^4}+\frac{4(\overline{\sigma}^2)^2}{n^2M^4}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=s_g)\mathbb{1}(t_{i'}^*=s_g)\\
&\sum_{m=1}^{M}\sum_{r>m}\sum_{m'=1}^{M}\sum_{r'>m'}\mathbb{1}(\{j_m(i,s_g),j_r(i,s_g)\}=\{j_{m'}(i',s_g),j_{r'}(i',s_g)\})\\
&\leq\frac{4(\overline{\sigma}^2)^2}{n^2M^4}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=s_g)\mathbb{1}(t_{i'}^*=s_g)\\
&\sum_{m=1}^{M}\sum_{r=1}^{M}\sum_{m'=1}^{M}\sum_{r'=1}^{M}\mathbb{1}\left(j_m(i,s_g)=j_{m'}(i',s_g)\right)+\mathbb{1}\left(j_m(i,s_g)=j_{r'}(i',s_g)\right)+o_{\mathbb{P}}(1)\\
&=\frac{8(\overline{\sigma}^2)^2}{n^2M^2}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=s_g)\mathbb{1}(t_{i'}^*=s_g)\sum_{m=1}^{M}\sum_{m'=1}^{M}\mathbb{1}\left(j_m(i,s_g)=j_{m'}(i',s_g)\right)+o_{\mathbb{P}}(1)\\
&=\frac{8(\overline{\sigma}^2)^2}{n^2M^2}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\sum_{m=1}^{M}\sum_{i'\neq i}\mathbb{1}(t_{i'}^*=s_g)\sum_{m'=1}^{M}\mathbb{1}\left(j_m(i,s_g)=j_{m'}(i',s_g)\right)+o_{\mathbb{P}}(1)\\
&\leq\frac{8(\overline{\sigma}^2)^2}{n^2M^2}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\sum_{m=1}^{M}K_M(j_m(i,s_g),s_g)+o_{\mathbb{P}}(1)\\
&=\frac{8(\overline{\sigma}^2)^2}{nM^2}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^2(i,s_g)+o_{\mathbb{P}}(1)=o_{\mathbb{P}}(1)
\end{aligned}
$$

The first inequality follows from the fact the fourth moment of Y is uniformly bounded and the fact that $\mathbb{E}\left[\epsilon_{j_m(i,s_g)t}\epsilon_{j_r(i,s_g)t}\epsilon_{j_{m'}(i',s_g)t}\epsilon_{j_{r'}(i',s_g)t}\middle|\mathbf{t}^*,\mathbf{X}\right]$ is nonzero if and only if $\{j_m(i,s_g),j_r(i,s_g)\}=\{j_{m'}(i',s_g),j_{r'}(i',s_g)\}$. The second inequality follows from the fact that the event $\{j_m(i,s_g),j_r(i,s_g)\}=$

$\{j_{m'}(i', s_g), j_{r'}(i', s_g)\}$ is included in the event $\{j_m(i, s_g) = j_{m'}(i', s_g)\} \cup \{j_m(i, s_g) = j_{r'}(i', s_g)\}$. The fourth equality follows from the fact the two terms in the summation are symmetric. The third inequality follows from the definition of $K_M(j_m(i, s_g), s_g)$. The last equality follows from the fact that these two expressions are two equivalent ways to express the same summation. The last equality follows from Lemma B5. Finally, by conditional Markov's inequality and bounded convergence,

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)\frac{2}{M^2}\sum_{m=1}^{M}\sum_{r>m} \epsilon_{j_m(i,s_g)t}\epsilon_{j_r(i,s_g)t} = o_{\mathbb{P}}(1)$$

For the last term,

$$\begin{aligned}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)\frac{2}{M}\sum_{m=1}^{M}\epsilon_{it}\epsilon_{j_m(i,s_g)t}\middle|\, \mathbf{t}^*, \mathbf{X}\right] &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)\frac{2}{M}\sum_{m=1}^{M}\mathbb{E}\left[\epsilon_{it}\middle|\, \mathbf{t}^*, \mathbf{X}\right]\mathbb{E}\left[\epsilon_{j_m(i,s_g)t}\middle|\, \mathbf{t}^*, \mathbf{X}\right] \\ &= 0\end{aligned}$$

and the conditional variance is:

$$\begin{aligned}\mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)\frac{2}{M}\sum_{m=1}^{M}\epsilon_{it}\epsilon_{j_m(i,s_g)t}\middle|\, \mathbf{t}^*, \mathbf{X}\right] &= \frac{4}{n^2M^2}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)\sum_{m=1}^{M} Var\left[\epsilon_{it}\epsilon_{j_m(i,s_g)t}\middle|\, \mathbf{t}^*, \mathbf{X}\right] \\ &= \frac{4}{n^2M^2}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)\sum_{m=1}^{M}\mathbb{E}\left[\epsilon_{it}^2\epsilon_{j_m(i,s_g)t}^2\middle|\, \mathbf{t}^*, \mathbf{X}\right] \\ &\leq \frac{4(\overline{\sigma}^2)^2}{nM^4} = o_{\mathbb{P}}(1).\end{aligned}$$

which implies

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)\frac{2}{M}\sum_{m=1}^{M}\epsilon_{it}\epsilon_{j_m(i,s_g)t} = o_{\mathbb{P}}(1).$$

Therefore,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)\left(\epsilon_{it} - \frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_m(i,s_g)t}\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)\sigma_t^2(s_g, X_i) + \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(t_i^* = \infty)\frac{K_M(i, s_g)}{M^2}\sigma_t^2(0, X_i) + o_{\mathbb{P}}(1).\end{aligned}$$

Third term: for notation simplicity, let

$$\Phi_{M,i} = \mu_t(s_g, X_i) - \frac{1}{M}\sum_{m=1}^{M} \mu_t(0, X_{j_m(i,s_g)}) - (\mu_t(s_g) - \mathbb{E}[\mu_t(0,X)|t^* = s_g])\,.$$

so we can rewrite the third term as:

$$\frac{2}{n}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)\Phi_{M,i}\left(\epsilon_{it} - \frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_m(i,s_g)t}\right).$$

First notice that

$$\begin{aligned}
&\mathbb{E}\left[\frac{2}{n}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)\Phi_{M,i}\left(\epsilon_{it} - \frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_m(i,s_g)t}\right)\middle|\, \mathbf{t}^*, \mathbf{X}\right] \\
&= \frac{2}{n}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s_g)\Phi_{M,i}\mathbb{E}\left[\epsilon_{it} - \frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_m(i,s_g)t}\middle|\, \mathbf{t}^*, \mathbf{X}\right] = 0
\end{aligned}$$

In addition,

$$\begin{aligned}
|\Phi_{M,i}| &= \left|\mu_t(s_g, X_i) - \frac{1}{M}\sum_{m=1}^{M} \mu_t(0, X_{j_m(i,s_g)}) - (\mu_t(s_g) - \mathbb{E}[\mu_t(0,X)|t^* = s_g])\right| \\
&\leq |\mu_t(s_g, X_i)| + \frac{1}{M}\sum_{m=1}^{M}\left|\mu_t(0, X_{j_m(i,s_g)})\right| + |\mu_t(s_g)| + |\mathbb{E}[\mu_t(0,X)|t^* = s_g]| \\
&\leq C
\end{aligned}$$

The first inequality follows from triangle inequality. The second inequality follows from the Lipschitz continuity of the conditional mean of the outcome and the support of the covariate is

bounded. This shows that $|\Phi_{M,i}| \leq C$ for all i. Now consider the conditional variance:

$$
\begin{aligned}
&\mathbb{V}\left[\frac{2}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\Phi_{M,i}\left(\epsilon_{it}-\frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_m(i,s_g)t}\right)\middle|\,\mathbf{t}^*,\mathbf{X}\right]\\
&=\frac{4}{n^2}\mathbb{E}\left[\left(\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\Phi_{M,i}\left(\epsilon_{it}-\frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_m(i,s_g)t}\right)\right)^2\middle|\,\mathbf{t}^*,\mathbf{X}\right]\\
&=\frac{4}{n^2}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{i'=1}^{n}\mathbb{1}(t_i^*=s_g)\mathbb{1}(t_{i'}^*=s_g)\Phi_{M,i}\Phi_{M,i'}\left(\epsilon_{it}-\frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_m(i,s_g)t}\right)\left(\epsilon_{i't}-\frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_{m'}(i',s_g)t}\right)\middle|\,\mathbf{t}^*,\mathbf{X}\right]\\
&=\frac{4}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\mathbb{1}(t_i^*=s_g)\mathbb{1}(t_{i'}^*=s_g)\Phi_{M,i}\Phi_{M,i'}\mathbb{E}\left[\left(\epsilon_{it}-\frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_m(i,s_g)t}\right)\left(\epsilon_{i't}-\frac{1}{M}\sum_{m'=1}^{M}\epsilon_{j_{m'}(i',s_g)t}\right)\middle|\,\mathbf{t}^*,\mathbf{X}\right]\\
&\leq\left|\frac{4}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\mathbb{1}(t_i^*=s_g)\mathbb{1}(t_{i'}^*=s_g)\Phi_{M,i}\Phi_{M,i'}\mathbb{E}\left[\epsilon_{it}\epsilon_{i't}\middle|\,\mathbf{t}^*,\mathbf{X}\right]\right|\\
&+\left|\frac{8}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\mathbb{1}(t_i^*=s_g)\mathbb{1}(t_{i'}^*=s_g)\Phi_{M,i}\Phi_{M,i'}\mathbb{E}\left[\epsilon_{it}\left(\frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_{m'}(i',s_g)t}\right)\middle|\,\mathbf{t}^*,\mathbf{X}\right]\right|\\
&+\left|\frac{4}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\mathbb{1}(t_i^*=s_g)\mathbb{1}(t_{i'}^*=s_g)\Phi_{M,i}\Phi_{M,i'}\mathbb{E}\left[\left(\frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_m(i,s_g)t}\right)\left(\frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_{m'}(i',s_g)t}\right)\middle|\,\mathbf{t}^*,\mathbf{X}\right]\right|
\end{aligned}
$$

For the first term,

$$
\begin{aligned}
&\left|\frac{4}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\mathbb{1}(t_i^*=s_g)\mathbb{1}(t_{i'}^*=s_g)\Phi_{M,i}\Phi_{M,i'}\mathbb{E}\left[\epsilon_{it}\epsilon_{i't}\middle|\,\mathbf{t}^*,\mathbf{X}\right]\right|\\
&=\left|\frac{4}{n^2}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\Phi_{M,i}^2\mathbb{E}\left[\epsilon_{it}^2\middle|\,\mathbf{t}^*,\mathbf{X}\right]\right|\leq\frac{4}{n^2}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)C^2\overline{\sigma}^2\leq\frac{4C^2\overline{\sigma}^2}{n}=o(1).
\end{aligned}
$$

The second term is zero because ϵ_{it} is independent of $\epsilon_{j_m(i',s_g)t}$ after conditioning on $\mathbf{t}^*,\mathbf{X}$. For the

third term,

$$
\begin{aligned}
&\left|\frac{4}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\mathbb{1}(t_i^*=s_g)\mathbb{1}(t_{i'}^*=s_g)\Phi_{M,i}\Phi_{M,i'}\mathbb{E}\left[\left(\frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_m(i,s_g)t}\right)\left(\frac{1}{M}\sum_{m'=1}^{M}\epsilon_{j_{m'}(i',s_g)t}\right)\middle|\mathbf{t}^*,\mathbf{X}\right]\right| \\
&\leq\left|\frac{4}{n^2M^2}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\Phi_{M,i}^2\sum_{m=1}^{M}\mathbb{E}\left[\epsilon_{j_m(i,s_g)t}^2\middle|\mathbf{t}^*,\mathbf{X}\right]\right| \\
&+\left|\frac{4}{n^2M^2}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=s_g)\mathbb{1}(t_{i'}^*=s_g)\Phi_{M,i}\Phi_{M,i'}\sum_{m=1}^{M}\sum_{m'=1}^{M}\mathbb{E}\left[\epsilon_{j_m(i,s_g)t}\epsilon_{j_{m'}(i',s_g)t}\middle|\mathbf{t}^*,\mathbf{X}\right]\right| \\
&\leq\left|\frac{4C^2}{n^2M^2}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\sum_{m=1}^{M}\mathbb{E}\left[\epsilon_{j_m(i,s_g)t}^2\middle|\mathbf{t}^*,\mathbf{X}\right]\right| \\
&+\left|\frac{4C^2}{n^2M^2}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=s_g)\mathbb{1}(t_{i'}^*=s_g)\sum_{m=1}^{M}\sum_{m'=1}^{M}\mathbb{E}\left[\epsilon_{j_m(i,s_g)t}\epsilon_{j_{m'}(i',s_g)t}\middle|\mathbf{t}^*,\mathbf{X}\right]\right| \\
&\leq\frac{4C^2\overline{\sigma}^2}{nM}+\frac{4C^2\overline{\sigma}^2}{n^2M^2}\sum_{i=1}^{n}\sum_{i'\neq i}\mathbb{1}(t_i^*=s_g)\mathbb{1}(t_{i'}^*=s_g)\sum_{m=1}^{M}\sum_{m'=1}^{M}\mathbb{1}(j_m(i,s_g)=j_{m'}(i',s_g)) \\
&\leq\frac{4C^2\overline{\sigma}^2}{nM^2}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^2(i,s_g)+o_{\mathbb{P}}(1)=o_{\mathbb{P}}(1)
\end{aligned}
$$

The last inequality follows from the previous arguments, and thus

$$
\frac{2}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\Phi_{M,i}\left(\epsilon_{it}-\frac{1}{M}\sum_{m=1}^{M}\epsilon_{j_m(i,s_g)t}\right)=o_{\mathbb{P}}(1).
$$

Therefore,

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\left(Y_{it}-\widehat{Y}_{it}-\left(\widetilde{Y}_t^{s_g}-\widetilde{Y}_t^{\infty,s_g}\right)\right)^2 \\
&=p_{s_g}\mathbb{V}\left[\mu_t(s_g,X)-\mu_t(0,X)|t^*=s_g\right]+\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\sigma_t^2(s_g,X_i) \\
&+\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)\frac{K_M(i,s_g)}{M^2}\sigma_t^2(0,X_i)+o_{\mathbb{P}}(1) \\
&=p_{s_g}\mathbb{V}\left[\mu_t(s_g,X)-\mu_t(0,X)|t^*=s_g\right]+\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s_g)\sigma_t^2(s_g,X_i) \\
&+\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)\frac{K_M(i,s_g)}{M^2}\widehat{\sigma}_t^2(0,X_i)+o_{\mathbb{P}}(1).
\end{aligned}
$$

The last equality follows from the proof of the first statement by setting $\alpha = 1$. As a result,

$$\begin{aligned}
\widehat{V}_{tt}^{s_g} &= \frac{n^2}{N_{s_g}^2}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^* = s_g)\left(Y_{it} - \widehat{Y}_{it} - \left(\widetilde{Y}_t^{s_g} - \widetilde{Y}_t^{\infty,s_g}\right)\right)^2 - \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^* = \infty)\frac{K_M(i,s_g)}{M^2}\widehat{\sigma}_t^2(0,X_i)\right) \\
&= \frac{n^2}{N_{s_g}^2}\left(p_{s_g} Var\left(\mu_t(s_g,X) - \mu_t(0,X)|t^* = s_g\right) + \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^* = s_g)\sigma_t^2(s_g,X_i) + o_{\mathbb{P}}(1)\right) \\
&\to_{\mathbb{P}} \frac{1}{p_{s_g}}\left(Var\left(\mu_t(s_g,X) - \mu_t(0,X)|t^* = s_g\right) + \mathbb{E}[\sigma_t^2(s_g,X)|t^* = s_g]\right) \\
&= V_{tt}^{s_g}.
\end{aligned}$$

This completes the proof. □

Proof of Proposition 3

From Theorem 2,

$$\begin{aligned}
\Omega_{t,\tilde{t}}^{s} = \frac{1}{p_s}\mathbb{V}\left[\mathbb{E}[\tau_{it}|t_i^* = s, X_i]|t_i^* = s\right] + \frac{1}{p_s}\mathbb{E}\left[\mathbb{V}[Y_{it} - Y_{i\tilde{t}}|t_i^* = s, X_i]|t_i^* = s\right] \\
+ \frac{1}{p_s^2 M^2}\mathbb{E}\left[\left(M + \alpha(M,q)\frac{e_s(X_i)}{e_\infty(X_i)}\right)e_s(X_i)\mathbb{V}[Y_{it} - Y_{i\tilde{t}}|t_i^* = \infty, X_i]\right]
\end{aligned}$$

but

$$\begin{aligned}
\frac{1}{p_s}\mathbb{V}\left[\mathbb{E}[\tau_{it}|t_i^* = s, X_i]|t_i^* = s\right] &= \frac{1}{p_s}\mathbb{E}\left[\left(\mathbb{E}[\tau_{it}|t_i^* = s, X_i] - \mathbb{E}[\tau_{it}|t_i^* = s]\right)^2\middle|\, t_i^* = s\right] \\
&= \frac{1}{p_s^2}\mathbb{E}\left[\left(\mathbb{E}[\tau_{it}|t_i^* = s, X_i] - \mathbb{E}[\tau_{it}|t_i^* = s]\right)^2 e_s(X_i)\right]
\end{aligned}$$

by the law of iterated expectations. Similarly,

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

and the result follows from straightforward algebra. Finally, to see that $\alpha(M,q) \geq M^2$, notice that by Theorem 4.1 in Chen and Han (2026), using their notation,

$$n\mathbb{E}[\nu_1\left(\mathcal{A}_M(x)|X_1 = x\right)] \to M\frac{f_1(x)}{f_0(x)}, \quad n^2\mathbb{E}\left[\nu_1\left(\mathcal{A}_M(x)^2|X_1 = x\right)\right] \to \alpha(M,q)\left(\frac{f_1(x)}{f_0(x)}\right)^2$$

from which

$$0 \leq \mathbb{V}[n\nu_1\left(\mathcal{A}_M(x)\right)|X_1 = x] \to \left[\alpha(M,q) - M^2\right]\left(\frac{f_1(x)}{f_0(x)}\right)^2$$

and since $f_1(x)/f_0(x) > 0$ under the overlap condition, it follows that $\alpha(M,q) \geq M^2$. □

Proof of Proposition 4

The regression residual is:

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

where

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

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

$$\hat{\mu} \to_{\mathbb{P}} \mu := (\mathbb{E}[\Delta Y_i | t_i^* = s], \mathbb{E}\left[\mathbb{E}[\Delta Y_i | t_i^* = \infty, X_i] | t_i^* = s\right])'.$$

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

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

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

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

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:

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

Note that

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

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

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

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

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

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

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

Collecting the results,

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

Now, letting $\mu_\Delta(s,x) = \mathbb{E}[\Delta Y_i | t_i^* = s, X_i = x]$,

$$\begin{aligned}\mathbb{E}[\varepsilon_i^2 | t_i^* = s] &= \mathbb{E}[(\Delta Y_i - \mathbb{E}[\Delta Y_i | t_i^* = s])^2 | t_i^* = s] = \mathbb{V}[\Delta\bar{Y}_i^s | t_i^* = s] \\ &= \mathbb{V}[\mu_\Delta(s, X_i) | t_i^* = s] + \mathbb{E}[\sigma_\Delta^2(s, X_i) | t_i^* = s] \\ &= \mathbb{V}[\mathbb{E}[\tau_{it} | t_i^* = s, X_i] | t_i^* = s] + \mathbb{V}[\mathbb{E}[\Delta Y_i(0) | t_i^* = s, X_i] | t_i^* = s] \\ &\quad + 2\mathbb{C}\mathrm{ov}(\mathbb{E}[\tau_{it} | t_i^* = s, X_i], \mathbb{E}[\Delta Y_i(0) | t_i^* = s, X_i] | t_i^* = s) + \mathbb{E}[\sigma_\Delta^2(s, X_i) | t_i^* = s]\end{aligned}$$

and

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

Next, use that under conditional parallel trends,

$$\mu_\Delta(0,X_i) = \mathbb{E}[\Delta Y_i(0)|t_i^*=s,X_i], \quad \mathbb{E}[\mu_\Delta(0,X_i)|t_i^*=s] = \mathbb{E}[\Delta Y_i(0)|t_i^*=s]$$

Collecting the results,

$$\begin{aligned}n\hat{V}_{\mathsf{ncr}} \to_{\mathbb{P}} \Omega^s_{t,\tilde{t}} \\ &+ \frac{1}{p_s}\mathbb{V}[\mathbb{E}[\Delta Y_i(0)|t_i^*=s,X_i]|t_i^*=s] \\ &+ \frac{2}{p_s}\mathbb{C}\mathrm{ov}(\mathbb{E}[\tau_{it}|t_i^*=s,X_i],\mathbb{E}[\Delta Y_i(0)|t_i^*=s,X_i]|t_i^*=s) \\ &+ \frac{1}{p_s^2M^2}\mathbb{E}\left[\left(M+\alpha(M,q)\frac{e_s(X_i)}{e_\infty(X_i)}\right)e_s(X_i)(\mathbb{E}[\Delta Y_i(0)|t_i^*=s,X_i]-\mathbb{E}[\Delta Y_i(0)|t_i^*=s])^2\right].\end{aligned}$$

as required. □

Proof of Proposition 5

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

$$\begin{aligned}\mathbb{E}[\tau_{it}|t_i^*=s] &= \mathbb{E}[\Delta Y_i|t_i^*=s] - \sum_{x\in\mathcal{X}}\mu_\Delta(0,x)\mathbb{P}[X_i=x|t_i^*=s] \\ &= \mathbb{E}[\Delta Y_i|t_i^*=s] - \sum_{x\in\mathcal{X}}\mu_\Delta(0,x)\mathbb{P}[t_i^*=s|X_i=x]\frac{p(x)}{p_s} \\ &= \mathbb{E}[\Delta Y_i|t_i^*=s] - \sum_{x\in\mathcal{X}}\mu_\Delta(0,x)e_s(x)\frac{p(x)}{p_s}\end{aligned}$$

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:

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

The estimator can be rewritten as:

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

where in this case:

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

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

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

Thus,

$$\begin{aligned}\hat{\tau}_{\mathsf{disc}} - \mathbb{E}[\tau_{it}|t_i^* = s] = \frac{1}{N_s}\sum_i \mathbb{1}(t_i^* = s)(\mu_\Delta(s, X_i) - \mu_\Delta(0, X_i) - \mathbb{E}[\tau_{it}|t_i^* = s]) \\ + \frac{1}{N_s}\sum_i \left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty)\frac{e_s(X_i)}{e_\infty(X_i)}\right)(\Delta Y_i - \mu_\Delta(t_i^*, X_i)) \\ - \sum_{x\in\mathcal{X}}\left(\frac{\hat{e}_s(x)}{\hat{e}_\infty(x)} - \frac{e_s(x)}{e_\infty(x)}\right)\frac{1}{N_s}\sum_i \mathbb{1}(t_i^* = \infty)\mathbb{1}(X_i = x)(\Delta Y_i - \mu_\Delta(0, x)).\end{aligned}$$

Now,

$$\begin{aligned}\frac{1}{N_s}\sum_i \mathbb{1}(t_i^* = \infty)\mathbb{1}(X_i = x)(\Delta Y_i - \mu_\Delta(0, x)) &\to_{\mathbb{P}} \frac{e_\infty(x)p(x)}{p_s}\mathbb{E}\left[\Delta Y_i - \mu_\Delta(0, x)\middle|\, t_i^* = \infty, X_i = x\right] \\ &= 0\end{aligned}$$

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

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

as long as $\hat{e}_s(x) > 0$ for every cell with $N_s(x) > 0$. Therefore,

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

and

$$\sqrt{n}(\hat{\tau}_{\mathsf{disc}} - \mathbb{E}[\tau_{it}|t_i^* = s]) = \frac{1}{\sqrt{n}}\sum_i \psi_i(s) + o_{\mathbb{P}}(1) \to_{\mathcal{D}} \mathcal{N}(0, \mathbb{V}[\psi_i(s)])$$

where

$$\begin{aligned}\psi_i(s) = &\frac{\mathbb{1}(t_i^* = s)}{p_s}(\mu_\Delta(s, X_i) - \mu_\Delta(0, X_i) - \mathbb{E}[\tau_{it}|t_i^* = s]) \\ &+ \frac{1}{p_s}\left(\mathbb{1}(t_i^* = s) - \mathbb{1}(t_i^* = \infty)\frac{e_s(X_i)}{e_\infty(X_i)}\right)(\Delta Y_i - \mu_\Delta(t_i^*, X_i)).\end{aligned}$$

The variance can be expressed as:

$$\begin{aligned}\mathbb{V}[\psi_i(s)] = &\frac{1}{p_s}\mathbb{E}\left[(\mathbb{E}[\tau_{it}|t_i^* = s, X_i] - \mathbb{E}[\tau_{it}|t_i^* = s])^2 \middle| t_i^* = s\right] + \frac{1}{p_s}\mathbb{E}[\sigma_\Delta^2(s, X_i)|t_i^* = s] \\ &+ \frac{1}{p_s^2}\mathbb{E}\left[\frac{e_s(X_i)^2}{e_\infty(X_i)}\sigma_\Delta^2(0, X_i)\right]\end{aligned}$$

as required. □

Proof of Proposition 6

The estimator of interest can be written as the estimator from a regression of ΔY_i 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 $\hat{\beta} = (\hat{\alpha}, \hat{\tau}_{\mathsf{nr}})'$ where $\hat{\alpha}$ is the intercept from the regression. Thus,

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

where in this case $\beta = (\mathbb{E}[\mu_\Delta(0, X_i)|t_i^* = s], \mathbb{E}[\tau_{it}|t_i^* = s])'$,

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

and $J = J_1 + J_2$ where

$$\begin{aligned}J_1 &= \frac{1}{1+M}\mathbb{V}\left[\begin{matrix}\mu_\Delta(s, X_i) + M\mu_\Delta(0, X_i)\\ \mu_\Delta(s, X_i)\end{matrix}\middle| t_i^* = s\right], \\ J_2 &= \frac{1}{1+M}\mathbb{E}\left[\begin{matrix}\sigma_\Delta^2(s, X_i) + M\sigma_\Delta^2(0, X_i) & \sigma_\Delta^2(s, X_i)\\ \sigma_\Delta^2(s, X_i) & \sigma_\Delta^2(s, X_i)\end{matrix}\middle| t_i^* = s\right].\end{aligned}$$

Then, by straightforward algebra,

$$\Omega_{\mathsf{nr}} = H^{-1}JH^{-1}_{(2,2)} = (1+M)\mathbb{E}\left[\left(\mathbb{E}[\tau_{it}|t_i^*=s,X_i]-\mathbb{E}[\tau_{it}|t_i^*=s]\right)^2\middle|\, t_i^*=s\right] + (1+M)\mathbb{E}\left[\sigma^2_\Delta(s,X_i)+\frac{\sigma^2_\Delta(0,X_i)}{M}\middle|\, t_i^*=s\right]$$

as required. □

D Proofs of Additional Results

Proof of Lemma B1

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

$$\begin{aligned}
\bar{D}_i &= \sum_{s<\infty} \mathbb{1}(t_i^*=s)\frac{\sum_t \lambda_t \mathbb{1}(t\geq s)}{\sum_t \lambda_t} = \sum_{s<\infty}\mathbb{1}(t_i^*=s)\Lambda_s \\
\tilde{D}_t &= \sum_{s<\infty} \mathbb{1}(t\geq s)\frac{\sum_i w_i \mathbb{1}(t_i^*=s)}{\sum_i w_i} = \sum_{s<\infty}\mathbb{1}(t\geq s)\hat{p}_s^w \\
\bar{D} &= \sum_{s<\infty}\hat{p}_s^w\Lambda_s
\end{aligned}$$

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

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

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

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

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^{\mathsf{post},s}(\lambda) - \bar{Y}_i^{\mathsf{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^{\mathsf{post},s}(\lambda) - \bar{Y}_i^{\mathsf{pre},s}(\lambda)) \left(\sum_t \lambda_t \right) \\
&- \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^{\mathsf{post},s}(\lambda) - \bar{Y}_i^{\mathsf{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^{\mathsf{post},s}(\lambda) - \bar{Y}_i^{\mathsf{pre},s}(\lambda))}{\sum_i w_i \mathbb{1}(t_i^* = s)} \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) \frac{\sum_i w_i \mathbb{1}(t_i^* \neq s) (\bar{Y}_i^{\mathsf{post},s}(\lambda) - \bar{Y}_i^{\mathsf{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^{\mathsf{post},s}(\lambda) - \bar{Y}_i^{\mathsf{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^{\mathsf{post},s}(\lambda) - \bar{Y}_i^{\mathsf{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^{\mathsf{post},s}(\lambda) - \bar{Y}_i^{\mathsf{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^{\mathsf{post},s}(\lambda) - \bar{Y}_i^{\mathsf{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:

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

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:

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

and thus

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

from which

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

where

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

Therefore,

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

Collecting the results, we have that:

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

where

$$\hat{\tau}_{\mathsf{num}}(w, \lambda) = \sum_{s<\infty} \sum_{s' \neq s} \sum_{t \geq s} \sum_{t' < s} \lambda_t \lambda_{t'} \hat{p}_s^w \hat{p}_{s'}^w \hat{\tau}_{t,t'}^{s,s'}$$

$$\hat{\tau}_{\mathsf{den}}(w, \lambda) = \left(\sum_t \lambda_t\right)^2 \sum_{s<\infty} \hat{p}_s^w \left\{\hat{p}_\infty^w \Lambda_s (1 - \Lambda_s) + \sum_{s' < s} \hat{p}_{s'}^w \Lambda_s (\Lambda_{s'} - \Lambda_s) + \sum_{s < s' < \infty} \hat{p}_{s'}^w (1 - \Lambda_s)(\Lambda_s - \Lambda_{s'})\right\}.$$

which completes the proof. □

Proof of Lemma B2

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

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

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,

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

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

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

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

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

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 the triangle inequality and Lipschitz continuity,

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

Hence, $\mathbb{E}[|T_{3,n}|] \le C\mathbb{E}[\left\|X_i - X_{j_M(i,s')}\right\| \, | t_i^* = s] = O(n^{-1/q})$ by identical distributions and Lemma 2 in Abadie and Imbens (2006). Thus, by Markov's inequality, $T_{3,n} = o_{\mathbb{P}}(1)$.

Therefore,

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

and

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

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

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

Now,

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

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 (2026) 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

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

For the first term, note that:

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

By Assumption 5(1) and Lemma 5.3 in Chen and Han (2026), $N_{s'}\mathbb{E}[\nu_s(\mathcal{A}_M(X_i))|t_i^* = s', X_i]m(s',X_i)$ 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,

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

because $N_s/N_{s'} \to_{\mathbb{P}} p_s/p_{s'}$ 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:

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

and thus

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

by Theorem 4.1 and Lemma 5.3 in Chen and Han (2026) 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

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

Similarly, for the second term,

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

By Theorem 4.1 and Lemma 5.3 in Chen and Han (2026),

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

On the other hand, by previous arguments

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

Thus

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

and therefore

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

Thus,

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

Finally, note that:

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

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

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

and thus

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

which completes the proof. □

Proof of Lemma B3

Let $\mathbf{X}_{s''} = (X_i)_{i:t_i^*=s''}$ be the subset of covariates in cohort $t_i^* = s''$. We have that

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

and

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

The last equality follows by the fact that $K_M(i,s)$ are independent of $K_M(i,s')$ after conditioning on $\mathbf{t}^*, \mathbf{X}_{s''}, t_i^* = s''$ Now,

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

where $\nu_s(\cdot)$, $\nu_{s'}(\cdot)$ are the conditional distributions of $X|t_i^*=s$, $X|t_i^*=s'$ and $\mathcal{A}_M(x)$ is the catchment area. This implies that

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

and thus

$$\mathbb{E}\left[K_M(i,s)K_M(i,s')m(s'',X_i)|t_i^*=s''\right] = \mathbb{E}\left[N_s\nu_s(\mathcal{A}_M(X_i))N_{s'}\nu_{s'}(\mathcal{A}_M(X_i))m(s'',X_i)|t_i^*=s''\right].$$

Notice that:

$$
\begin{aligned}
&\mathbb{E}\left[N_s\nu_s(\mathcal{A}_M(X_i))N_{s'}\nu_{s'}(\mathcal{A}_M(X_i))m(s'',X_i)|t_i^*=s''\right] \\
&=\mathbb{E}\left[\frac{N_sN_{s'}}{N_{s''}^2}N_{s''}^2\mathbb{E}[\nu_s(\mathcal{A}_M(X_i))\nu_{s'}(\mathcal{A}_M(X_i))|\mathbf{t}^*,t_i^*=s'',X_i]m(s'',X_i)|t_i^*=s''\right] \\
&=\mathbb{E}\left[\frac{N_sN_{s'}}{N_{s''}^2}N_{s''}^2\mathbb{E}[\nu_s(\mathcal{A}_M(X_i))\nu_{s'}(\mathcal{A}_M(X_i))|N_{s''},t_i^*=s'',X_i]m(s'',X_i)|t_i^*=s''\right] \\
&=\frac{p_sp_{s'}}{p_{s''}^2}\mathbb{E}\left[N_{s''}^2\mathbb{E}[\nu_s(\mathcal{A}_M(X_i))\nu_{s'}(\mathcal{A}_M(X_i))|N_{s''},t_i^*=s'',X_i]m(s'',X_i)|t_i^*=s''\right] \\
&+\mathbb{E}\left[\left(\frac{N_sN_{s'}}{N_{s''}^2}-\frac{p_sp_{s'}}{p_{s''}^2}\right)N_{s''}^2\mathbb{E}[\nu_s(\mathcal{A}_M(X_i))\nu_{s'}(\mathcal{A}_M(X_i))|N_{s''},t_i^*=s'',X_i]m(s'',X_i)|t_i^*=s''\right].
\end{aligned}
\tag{7}
$$

The second equality follows from the fact that, under the i.i.d. sampling assumption, after conditioning on $N_{s''}, t_i^*=s'', X_i$, any extra information from the set $\mathbf{t}^*, t_i^*=s'', X_i$ will not affect the conditional expectation of the catchment area.

To derive the probability limit of the first term in the last equality of (7), we first define the mixture distribution with measure $\overline{\nu}$ and density $\overline{f}$,

$$
\overline{\nu}=\frac{\nu_s+\nu_{s'}}{2},\quad \overline{f}(X_i)=\frac{f_s(X_i)+f_{s'}(X_i)}{2}
$$

Theorem 4.1 in Chen and Han (2026) implies that for almost all $x\in\mathcal{X}$

$$
\begin{aligned}
&\lim_{n\to\infty}N_{s''}^2\mathbb{E}[\nu_s(\mathcal{A}_M(X_i))\nu_{s'}(\mathcal{A}_M(X_i))|N_{s''},t_i^*=s'',X_i=x] \\
&=\lim_{n\to\infty}2N_{s''}^2\mathbb{E}\left[\left(\frac{\nu_s(\mathcal{A}_M(X_i))+\nu_{s'}(\mathcal{A}_M(X_i))}{2}\right)^2\middle|N_{s''},t_i^*=s'',X_i=x\right] \\
&-\lim_{n\to\infty}N_{s''}^2\mathbb{E}\left[\frac{\nu_s^2(\mathcal{A}_M(X_i))}{2}\middle|N_{s''},t_i^*=s'',X_i=x\right] \\
&-\lim_{n\to\infty}N_{s''}^2\mathbb{E}\left[\frac{\nu_{s'}^2(\mathcal{A}_M(X_i))}{2}\middle|N_{s''},t_i^*=s'',X_i=x\right] \\
&=\lim_{n\to\infty}N_{s''}^2 2\mathbb{E}\left[\overline{\nu}^2(\mathcal{A}_M(X_i))\middle|N_{s''},t_i^*=s'',X_i=x\right] \\
&-\lim_{n\to\infty}N_{s''}^2\mathbb{E}\left[\frac{\nu_s^2(\mathcal{A}_M(X_i))}{2}\middle|N_{s''},t_i^*=s'',X_i=x\right] \\
&-\lim_{n\to\infty}N_{s''}^2\mathbb{E}\left[\frac{\nu_{s'}^2(\mathcal{A}_M(X_i))}{2}\middle|N_{s''},t_i^*=s'',X_i=x\right] \\
&=2\alpha(M,q)\left[\left(\frac{f_s(x)+f_{s'}(x)}{2f_{s''}(x)}\right)^2-\frac{1}{4}\left(\frac{f_s(x)}{f_{s''}(x)}\right)^2-\frac{1}{4}\left(\frac{f_{s'}(x)}{f_{s''}(x)}\right)^2\right] \\
&=\alpha(M,q)\frac{f_s(x)f_{s'}(x)}{f_{s''}^2(x)}
\end{aligned}
$$

Now define

$$
\begin{aligned}
a_{N_{s''}}(x) &= N_{s''}^2 \mathbb{E}\left[\nu_s(\mathcal{A}_M(X_i))\nu_{s'}(\mathcal{A}_M(X_i))\right| N_{s''}, t_i^* = s'', X_i = x] \\
a(x) &= \alpha(M,q)\frac{f_s(x)f_{s'}(x)}{f_{s''}^2(x)} \\
F_n(x) &= \mathbb{E}\left[N_{s''}^2\nu_s(\mathcal{A}_M(X_i))\nu_{s'}(\mathcal{A}_M(X_i))\right| t_i^* = s'', X_i = x\big] \\
&= \mathbb{E}[a_{N_{s''}}|t_i^* = s'', X_i = x]
\end{aligned}
$$

By the derivation above, for almost all $x \in \mathcal{X}$, $\lim_{n\to\infty} a_{N_{s''}}(x) = a(x)$. Now, for any $\epsilon > 0$, there exists a constant C such that for all $N_{s''} > C$, $\left|a_{N_{s''}}(x) - a(x)\right| \leq \epsilon$

$$
\begin{aligned}
|F_n(x) - a| &= \left|\mathbb{E}[a_{N_{s''}}(x)|t_i^* = s'', X_i = x] - a\right| \\
&\leq \left|\mathbb{E}\left[a_{N_{s''}}(x) - a(x)|N_{s''} > C, t_i^* = s'', X_i = x\right] P(N_{s''} > C)\right| \\
&+ \left|\mathbb{E}\left[a_{N_{s''}}(x) - a(x)|N_{s''} \leq C, t_i^* = s'', X_i = x\right] P(N_{s''} \leq C)\right| \\
&\leq \epsilon + P(N_{s''} \leq C) \max_{N_{s''}<C} \left|a_{N_{s''}}(x) - a(x)\right| \\
&\leq \epsilon + C^2 P(N_{s''} < C)
\end{aligned}
$$

The last inequality follows because when $N_{s''} < C$ and the fact that the measure of the catchment area is bounded by 1. Notice that as $n \to \infty$, $N_{s''} \to \infty$ almost surely, therefore,

$$
\limsup_{n\to\infty} |F_n(x) - a| \leq \lim_{n\to\infty} \epsilon + C^2 P(N_{s''} < C) = \epsilon
$$

This shows that for almost all $x \in \mathcal{X}$

$$
F_n(x) \to \alpha(M,q)\frac{f_s(x)f_{s'}(x)}{f_{s''}^2(x)}
$$

Under the assumptions in the paper, $a(x) = \alpha(M,q)\frac{f_s(x)f_{s'}(x)}{f_{s''}^2(x)}$ is uniformly bounded for all x in $\mathcal{X}$. Lastly,

$$
\begin{aligned}
&\frac{N_{s''}^2}{N_s N_{s'}}\mathbb{E}\left[K_M(i,s)K_M(i,s')|N_s, N_{s'}, N_{s''}, t_i^* = s'', X_i = x\right] \\
&= \frac{N_{s''}^2}{N_s N_{s'}}\mathbb{E}\left[\mathbb{E}\left[K_M(i,s)K_M(i,s')|\mathbf{t}^*, \mathbf{X}_{s''}, t_i^* = s'', X_i = x\right]\right. \\
&\left.|N_s, N_{s'}, N_{s''}, t_i^* = s'', X_i = x\right] \\
&= \frac{N_{s''}^2}{N_s N_{s'}}\mathbb{E}\left[N_s\nu_s(\mathcal{A}_M(X_i))N_{s'}\nu_{s'}(\mathcal{A}_M(X_i))|N_s, N_{s'}, N_{s''}, t_i^* = s'', X_i = x\right] \\
&= N_{s''}^2\mathbb{E}\left[\nu_s(\mathcal{A}_M(X_i))\nu_{s'}(\mathcal{A}_M(X_i))|N_s, N_{s'}, N_{s''}, t_i^* = s'', X_i = x\right] \\
&= N_{s''}^2\mathbb{E}\left[\nu_s(\mathcal{A}_M(X_i))\nu_{s'}(\mathcal{A}_M(X_i))|N_{s''}, t_i^* = s'', X_i = x\right]
\end{aligned}
$$

The last equality follows from the fact that after conditioning on the number of unit in cohort s'', unit i being in cohort s'' and its covariate value, then by definition, the catchment area is not a function of the number of units in other cohorts.
Therefore, by law of iterated expectation,

$$\begin{aligned}
F_n(x) &= \mathbb{E}\left[N_{s''}^2 \nu_s(\mathcal{A}_M(X_i))\nu_{s'}(\mathcal{A}_M(X_i))\middle| t_i^* = s'', X_i = x\right] \\
&= \mathbb{E}\left[N_{s''}^2 \mathbb{E}\left[\nu_s(\mathcal{A}_M(X_i))\nu_{s'}(\mathcal{A}_M(X_i))|N_{s''}, t_i^* = s'', X_i = x\right]\middle| t_i^* = s'', X_i = x\right] \\
&= \mathbb{E}\left[\frac{N_{s''}^2}{N_s N_{s'}}\mathbb{E}\left[K_M(i,s)K_M(i,s')|N_s, N_{s'}, N_{s''}, t_i^* = s'', X_i = x\right]\middle| t_i^* = s'', X_i = x\right] \\
&= \mathbb{E}\left[\frac{N_{s''}^2}{N_s N_{s'}}K_M(i,s)K_M(i,s')\middle| t_i^* = s'', X_i = x\right] \\
&\leq \left(\mathbb{E}\left[\frac{N_{s''}^4}{N_s^4}\middle| t_i^* = s'', X_i = x\right]\right)^{1/4}\left(\mathbb{E}\left[\frac{N_{s''}^4}{N_{s'}^4}\middle| t_i^* = s'', X_i = x\right]\right)^{1/4} \\
&\left(\mathbb{E}\left[K_M^4(i,s)\middle| t_i^* = s'', X_i = x\right]\right)^{1/4}\left(\mathbb{E}\left[K_M^4(i,s')\middle| t_i^* = s'', X_i = x\right]\right)^{1/4} \\
&= \left(\mathbb{E}\left[\frac{N_{s''}^4}{N_s^4}\middle| t_i^* = s''\right]\right)^{1/4}\left(\mathbb{E}\left[\frac{N_{s''}^4}{N_{s'}^4}\middle| t_i^* = s''\right]\right)^{1/4} \\
&\left(\mathbb{E}\left[K_M^4(i,s)\middle| t_i^* = s'', X_i = x\right]\right)^{1/4}\left(\mathbb{E}\left[K_M^4(i,s')\middle| t_i^* = s'', X_i = x\right]\right)^{1/4}
\end{aligned}$$

The inequality follows from the Hölder's inequality. By Lemma S.3 in Abadie and Imbens (2016) and proof of Lemma 3 in Abadie and Imbens (2006) (page 264), the product above is uniformly bounded in n and x and thus $F_n(x)$ is uniformly bounded in n and x. Apply the dominated convergence theorem, for the first term in the last equality of (7), we have:

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

For the second term in the last equality of (7),

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

$$
\begin{aligned}
&\leq \frac{C}{p_{s''}}\mathbb{E}\left[\left(\frac{N_sN_{s'}}{N_{s''}^2}-\frac{p_sp_{s'}}{p_{s''}^2}\right)^2\right]^{1/2}\mathbb{E}[N_{s''}^4\left(\mathbb{E}[\nu_s(\mathcal{A}_M(X_i))\nu_{s'}(\mathcal{A}_M(X_i))|N_{s''},t_i^*=s'',X_i]\right)^2]^{1/2}\\
&=\frac{C}{p_{s''}}\mathbb{E}\left[\left(\frac{N_sN_{s'}}{N_{s''}^2}-\frac{p_sp_{s'}}{p_{s''}^2}\right)^2\right]^{1/2}\mathbb{E}\left[\left(\frac{N_{s''}^2}{N_sN_{s'}}\mathbb{E}\left[K_M(i,s)K_M(i,s')\right|N_s,N_{s'},N_{s''},t_i^*=s'',X_i\right]\right)^2\right]^{1/2}\\
&\leq\frac{C}{p_{s''}}\mathbb{E}\left[\left(\frac{N_sN_{s'}}{N_{s''}^2}-\frac{p_sp_{s'}}{p_{s''}^2}\right)^2\right]^{1/2}\mathbb{E}\left[\frac{N_{s''}^4}{N_s^2N_{s'}^2}\mathbb{E}\left[K_M^2(i,s)K_M^2(i,s')\right|N_s,N_{s'},N_{s''},t_i^*=s'',X_i\right]\right]^{1/2}\\
&\leq\frac{C}{p_{s''}}\mathbb{E}\left[\left(\frac{N_sN_{s'}}{N_{s''}^2}-\frac{p_sp_{s'}}{p_{s''}^2}\right)^2\right]^{1/2}\mathbb{E}\left[\left(\frac{N_{s''}}{N_s}\right)^8\right]^{1/8}\mathbb{E}\left[\left(\frac{N_{s''}}{N_{s'}}\right)^8\right]^{1/8}\mathbb{E}\left[K_M^8(i,s)\right]^{1/8}\mathbb{E}\left[K_M^8(i,s')\right]^{1/8}
\end{aligned}
$$

The first inequality follows from the fact that $m(s'',X_i)$ is uniformly bounded. The second to last inequality follows from the conditional Jensen's inequality and the last inequality follows from law of iterated expectation and Hölder's inequality. Now,

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{N_sN_{s'}}{N_{s''}^2}-\frac{p_sp_{s'}}{p_{s''}^2}\right)^2\right]&=\mathbb{E}\left[\left(\frac{N_sN_{s'}}{N_{s''}^2}-\frac{p_sp_{s'}}{p_{s''}^2}\right)^2\mathbb{1}\left(\left|\frac{N_sN_{s'}}{N_{s''}^2}-\frac{p_sp_{s'}}{p_{s''}^2}\right|>\delta\right)\right]\\
&+\mathbb{E}\left[\left(\frac{N_sN_{s'}}{N_{s''}^2}-\frac{p_sp_{s'}}{p_{s''}^2}\right)^2\mathbb{1}\left(\left|\frac{N_sN_{s'}}{N_{s''}^2}-\frac{p_sp_{s'}}{p_{s''}^2}\right|\leq\delta\right)\right]\\
&\leq\mathbb{E}\left[\left(\frac{N_sN_{s'}}{N_{s''}^2}-\frac{p_sp_{s'}}{p_{s''}^2}\right)^2\mathbb{1}\left(\left|\frac{N_sN_{s'}}{N_{s''}^2}-\frac{p_sp_{s'}}{p_{s''}^2}\right|>\delta\right)\right]+\delta\\
&\leq\mathbb{E}\left[\left(\frac{N_sN_{s'}}{N_{s''}^2}-\frac{p_sp_{s'}}{p_{s''}^2}\right)^4\right]^{1/2}\mathbb{P}\left[\left|\frac{N_sN_{s'}}{N_{s''}^2}-\frac{p_sp_{s'}}{p_{s''}^2}\right|>\delta\right]^{1/2}+\delta\\
&\to\delta.
\end{aligned}
$$

The last inequality follow by the fact that $\frac{N_sN_{s'}}{N_{s''}^2}$ plim $\frac{p_sp_{s'}}{p_{s''}^2}$ and the first term is uniformly bounded because:

$$
\mathbb{E}\left[\left(\frac{N_sN_{s'}}{N_{s''}^2}-\frac{p_sp_{s'}}{p_{s''}^2}\right)^4\right]\leq 8\mathbb{E}\left[\left|\frac{N_sN_{s'}}{N_{s''}^2}\right|^4+\left|\frac{p_sp_{s'}}{p_{s''}^2}\right|^4\right]\leq 8\left(\mathbb{E}\left[\left|\frac{N_s}{N_{s''}}\right|^8\right]^{1/2}\mathbb{E}\left[\left|\frac{N_{s'}}{N_{s''}}\right|^8\right]^{1/2}+\left|\frac{p_sp_{s'}}{p_{s''}^2}\right|^4\right).
$$

This is uniformly bounded in n by Lemma S3 in Abadie and Imbens (2016). Because δ is arbitrary, it follows that:

$$
\mathbb{E}\left[\left(\frac{N_sN_{s'}}{N_{s''}^2}-\frac{p_sp_{s'}}{p_{s''}^2}\right)^2\right]=o(1)
$$

and thus

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

Also note that:

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

so we have:

$$\begin{aligned}
\mathbb{E}\left[K_M(i,s)K_M(i,s')m(s'',X_i)|t_i^*=s''\right] &\to \alpha(M,q)\frac{p_s p_{s'}}{p_{s''}^2}\mathbb{E}\left[\frac{f_s(X_i)f_{s'}(X_i)}{f_{s''}^2(X_i)}m(s'',X_i)\middle| t_i^*=s''\right] \\
&= \alpha(M,q)\mathbb{E}\left[\frac{e_s(X_i)e_{s'}(X_i)}{e_{s''}^2(X_i)}m(s'',X_i)\middle| t_i^*=s''\right] \\
&= \frac{\alpha(M,q)}{p_{s''}}\mathbb{E}\left[\frac{e_s(X_i)e_{s'}(X_i)}{e_{s''}^2(X_i)}m(s'',X_i)\mathbb{1}(t_i^*=s'')\right] \\
&= \frac{\alpha(M,q)}{p_{s''}}\mathbb{E}\left[\frac{e_s(X_i)e_{s'}(X_i)}{e_{s''}(X_i)}m(s'',X_i)\right]
\end{aligned}$$

Thus

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

Lastly, by law of iterated expectation,

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \mathbb{1}(t_i^*=s'')K_M(i,s)K_M(i,s')(Z_i - m(s'',X_i))\right] = 0$$

and

$$\begin{aligned}
&\mathbb{V}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{1}(t_i^*=s'')K_M(i,s)K_M(i,s')(Z_i - m(s'',X_i))\right) \\
&= \mathbb{E}\left[\mathbb{V}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{1}(t_i^*=s'')K_M(i,s)K_M(i,s')(Z_i - m(s'',X_i))\middle|\mathbf{t}^*,\mathbf{X}\right)\right] \\
&= \mathbb{E}\left[\frac{1}{n^2}\sum_{i=1}^n \mathbb{1}(t_i^*=s'')K_M^2(i,s)K_M^2(i,s')\sigma_Z^2(s'',X_i)\right] \\
&= \frac{1}{n}\mathbb{E}\left[\mathbb{1}(t_i^*=s'')K_M^2(i,s)K_M^2(i,s')\sigma_Z^2(s'',X_i)\right] \\
&\leq \frac{C}{n}\mathbb{E}\left[K_M^2(i,s)K_M^2(i,s')\right] \\
&\leq \frac{C}{n}\mathbb{E}\left[K_M^4(i,s)\right]^{1/2}\mathbb{E}\left[K_M^4(i,s')\right]^{1/2} \\
&= o_{\mathbb{P}}(1).
\end{aligned}$$

Thus,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s'')K_M(i,s)K_M(i,s')(Z_i-m(s'',X_i))\to_{\mathbb{P}}0.$$

Finally, define

$$R_n=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=s'')K_M(i,s)K_M(i,s')m(s'',X_i).$$

Consider replacing the first observation $(Y_{11},\ldots,Y_{1T},t_1^*,X_1)$ of the original sample $(Y_{i1},\ldots,Y_{iT},t_i^*,X_i)_{i=1}^n$ with an independent copy $(Y_{1'1},\ldots,Y_{1'T},t_{1'}^*,X_{1'})$ and conducting the same two nearest neighbor matchings as in Lemma B3, each with M neighbors, on the subset of the new sample with the same three cohorts s, s', and s''. Let $K_M^{(1)}(i,s)$ and $K_M^{(1)}(i,s')$ denote the number of times unit i is used as a match in the nearest neighbor matchings with cohort s and s', respectively, based on the new sample. Define $R_n^{(1)}$ as the counterpart of R_n based on the new sample.

Then

$$\begin{aligned}
\left|R_n-R_n^{(1)}\right| &= \frac{1}{n}\Big|\mathbb{1}(t_1^*=s'')K_M(1,s)K_M(1,s')m(s'',X_1)-\mathbb{1}(t_{1'}^*=s'')K_M^{(1)}(1',s)K_M^{(1)}(1',s')m(s'',X_{1'})\\
&\quad+\sum_{i=2}^{n}\mathbb{1}(t_i^*=s'')\left(K_M(i,s)K_M(i,s')-K_M^{(1)}(i,s)K_M^{(1)}(i,s')\right)m(s'',X_i)\Big|\\
&\le \frac{C}{n}\Bigg(\left|\mathbb{1}(t_1^*=s'')K_M(1,s)K_M(1,s')\right|+\left|\mathbb{1}(t_{1'}^*=s'')K_M^{(1)}(1',s)K_M^{(1)}(1',s')\right|\\
&\quad+\sum_{i=2}^{n}\mathbb{1}(t_i^*=s'')\left|\left(K_M(i,s)K_M(i,s')-K_M^{(1)}(i,s)K_M^{(1)}(i,s')\right)\right|\Bigg)\\
&\le \frac{C}{n}\Bigg(\mathbb{1}(t_1^*=s'')\left|K_M^2(1,s)+K_M^2(1,s')\right|+\mathbb{1}(t_{1'}^*=s'')\left|\left(K_M^{(1)}(1',s)\right)^2+\left(K_M^{(1)}(1',s')\right)^2\right|\\
&\quad+\sum_{i=2}^{n}\mathbb{1}(t_i^*=s'')\left|\left(K_M(i,s)K_M(i,s')-K_M^{(1)}(i,s)K_M^{(1)}(i,s')\right)\right|\Bigg) \qquad (8)\\
&\le \frac{C}{n}\Bigg(\mathbb{1}(t_1^*=s'')\left|K_M^2(1,s)+K_M^2(1,s')\right|+\mathbb{1}(t_{1'}^*=s'')\left|\left(K_M^{(1)}(1',s)\right)^2+\left(K_M^{(1)}(1',s')\right)^2\right|\\
&\quad+\sum_{m=1}^{M}\left(K_M(j_m(1,s''),s)+K_M(j_m(1,s''),s')+K_M^{(1)}(j_m(1,s''),s)+K_M^{(1)}(j_m(1,s''),s)\right)\\
&\quad+\left[\max_{t_i^*=s''}K_M(i,s)\right]K_M(1,s')+\left[\max_{t_i^*=s''}\widetilde{K}_M^{(1)}(i,s')\right]K_M(1,s)\\
&\quad\left[\max_{t_i^*=s''}K_M^{(1)}(i,s)\right]K_M^{(1)}(1,s')+\left[\max_{t_i^*=s''}\widetilde{K}_M^{(1)}(i,s')\right]K_M^{(1)}(1,s)\Bigg)
\end{aligned}$$

Now, by the Efron-Stein inequality,

$$
\begin{aligned}
&\mathbb{V}[R_n] \le \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}\left[\left(R_n - R_n^{(1)}\right)^2\right] \\
&= \frac{n}{2}\mathbb{E}\left[\left(R_n - R_n^{(1)}\right)^2\right] \\
&\le \frac{C^2}{2n}\mathbb{E}\left[\left(\mathbb{1}(t_1^* = s'')\left|K_M^2(1,s) + K_M^2(1,s')\right| + \mathbb{1}(t_{1'}^* = s'')\left|\left(K_M^{(1)}(1',s)\right)^2 + \left(K_M^{(1)}(1',s')\right)^2\right|\right.\right. \\
&+ \sum_{m=1}^{M}\left(K_M(j_m(1,s''),s) + K_M(j_m(1,s''),s') + K_M^{(1)}(j_m(1,s''),s) + K_M^{(1)}(j_m(1,s''),s)\right) \\
&+ \left[\max_{t_i^*=s''} K_M(i,s)\right]K_M(1,s') + \left[\max_{t_i^*=s''}\widetilde{K}_M^{(1)}(i,s')\right]K_M(1,s) \\
&\left.\left.+ \left[\max_{t_i^*=s''} K_M^{(1)}(i,s)\right]K_M^{(1)}(1,s') + \left[\max_{t_i^*=s''}\widetilde{K}_M^{(1)}(i,s')\right]K_M^{(1)}(1,s)\right)^2\right] \qquad (9) \\
&\le \frac{7C^2}{2n}\left(\mathbb{E}\left[\mathbb{1}(t_1^* = s'')\left|K_M^2(1,s) + K_M^2(1,s')\right|^2\right] + \mathbb{E}\left[\mathbb{1}(t_{1'}^* = s'')\left|\left(K_M^{(1)}(1',s)\right)^2 + \left(K_M^{(1)}(1',s')\right)^2\right|^2\right]\right. \\
&+\mathbb{E}\left[\left(\sum_{m=1}^{M}\left(K_M(j_m(1,s''),s) + K_M(j_m(1,s''),s') + K_M^{(1)}(j_m(1,s''),s) + K_M^{(1)}(j_m(1,s''),s)\right)\right)^2\right] \\
&+\mathbb{E}\left[\left[\max_{t_i^*=s''} K_M^2(i,s)\right]K_M^2(1,s')\right] + \mathbb{E}\left[\left[\max_{t_i^*=s''}\left(\widetilde{K}_M^{(1)}(i,s')\right)^2\right]K_M^2(1,s)\right] \\
&\left.+\mathbb{E}\left[\left[\max_{t_i^*=s''}\left(K_M^{(1)}(i,s)\right)^2\right]\left(K_M^{(1)}(1,s')\right)^2\right] + \mathbb{E}\left[\left[\max_{t_i^*=s''}\left(\widetilde{K}_M^{(1)}(i,s')\right)^2\right]\left(K_M^{(1)}(1,s)\right)^2\right]\right)
\end{aligned}
$$

For the first term in the last equality of (9),

$$
\begin{aligned}
\mathbb{E}\left[\mathbb{1}(t_1^* = s'')\left|K_M^2(1,s) + K_M^2(1,s')\right|^2\right] &\le \mathbb{E}\left[\left|K_M^2(1,s) + K_M^2(1,s')\right|^2\right] \\
&\le \mathbb{E}\left[K_M^4(1,s)\right] + \mathbb{E}\left[K_M^4(1,s)\right] \\
&= O(1)
\end{aligned}
$$

Similarly, we can also show that the second term is also $O(1)$.

For the third term in the last equality of (9),

$$
\begin{aligned}
&\mathbb{E}\left[\left(\sum_{m=1}^{M}\left(K_M(j_m(1,s''),s) + K_M(j_m(1,s''),s') + K_M^{(1)}(j_m(1,s''),s) + K_M^{(1)}(j_m(1,s''),s)\right)\right)^2\right] \\
&\le 4M\sum_{m=1}^{M}\mathbb{E}\left[\left(K_M^2(j_m(1,s''),s) + K_M^2(j_m(1,s''),s') + \left(K_M^{(1)}(j_m(1,s''),s)\right)^2 + \left(K_M^{(1)}(j_m(1,s''),s)\right)^2\right)\right] \\
&= O(1)
\end{aligned}
$$

For the fourth term in the last equality of (9),

$$\begin{aligned}
\mathbb{E}\left[\left[\max_{t_i^*=s''} K_M^2(i,s)\right] K_M^2(1,s')\right] &\leq \left(\mathbb{E}\left[\max_{t_i^*=s''} K_M^4(i,s)\right]\right)^{1/2} \left(\mathbb{E}\left[K_M^4(i,s')\right]\right)^{1/2} \\
&\leq \left(\mathbb{E}\left[\sum_{t_i^*=s''} K_M^4(i,s)\right]\right)^{1/2} \left(\mathbb{E}\left[K_M^4(i,s')\right]\right)^{1/2} \\
&\leq (nC)^{1/2} \left(\mathbb{E}\left[K_M^4(i,s')\right]\right)^{1/2} \\
&= O(n^{1/2})O(1) \\
&= O(n^{1/2})
\end{aligned}$$

Similarly, we can show that the fifth, sixth, and seventh terms are also $O(n^{1/2})$. Therefore,

$$\begin{aligned}
\mathbb{V}[R_n] &\leq \frac{7C^2}{2n}\left(O(1)+O(1)+O(1)+O(n^{1/2})+O(n^{1/2})+O(n^{1/2})+O(n^{1/2})\right) \\
&= O(n^{-1/2}) = o(1)
\end{aligned}$$

and

$$\mathbb{E}[R_n] = \mathbb{E}[T_n] \to \alpha(M,q)\mathbb{E}\left[\frac{e_s(X_i)e_{s'}(X_i)}{e_{s''}(X_i)} m(s'', X_i)\right]$$

This shows that

$$R_n \to_{\mathbb{P}} \alpha(M,q)\mathbb{E}\left[\frac{e_s(X_i)e_{s'}(X_i)}{e_{s''}(X_i)} m(s'', X_i)\right]$$

Notice that $T_n - R_n \to_{\mathbb{P}} 0$ and thus

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

By Slutsky's theorem,

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(t_i^* = s'') \frac{K_M(i,s)K_M(i,s')}{M^2} Z_i = \frac{1}{M^2} \cdot T_n \to_{\mathbb{P}} \frac{\alpha(M,q)}{M^2}\mathbb{E}\left[\frac{e_s(X_i)e_{s'}(X_i)}{e_{s''}(X_i)} m(s'', X_i)\right]$$

which completes the proof. □

Proof of Lemma B4

Let $\xi > 0$ be any fixed constant and pick a positive integer r (so r depends on ξ) such that $r\xi > 1$, then:

$$\begin{aligned}
P(N_{s'}^{-\xi} \max_{t_i^*=s} N_{s'}^{1/q} \left\|X_i - X_{j_m(i,s')}\right\| > \epsilon | \mathbf{t}^*) &\leq \sum_{t_i^*=s} P(N_{s'}^{-\xi} N_s^{1/q} \left\|X_i - X_{j_m(i,s')}\right\| > \epsilon | \mathbf{t}^*) \\
&= \sum_{i=1}^{n} \mathbb{1}(t_i^* = s) P(N_{s'}^{-\xi} N_s^{1/q} \left\|X_i - X_{j_m(i,s')}\right\| > \epsilon | \mathbf{t}^*) \\
&= N_s P(N_{s'}^{-r\xi} N_{s'}^{r/q} \left\|X_i - X_{j_m(i,s')}\right\|^r > \epsilon^r | \mathbf{t}_{-i}^*, t_i^* = s) \\
&\leq \frac{\mathbb{E}[N_{s'}^{r/q} \left\|X_i - X_{j_m(i,s')}\right\|^r | \mathbf{t}_{-i}^*, t_i^* = s]}{\epsilon^r} \frac{N_s}{N_{s'}^{r\xi}} \\
&\leq \frac{1}{N_{s'}^{r\xi-1}} \frac{A_{12}}{\epsilon^r} \frac{N_s}{N_{s'}}.
\end{aligned}$$

The first inequality follows from the Bonferroni inequality. The second inequality follows from the Markov's inequality. The third inequality follows from lemma 2 in Abadie and Imbens (2006). Notice under our assumption, when $n \to \infty$, $\frac{N_s}{N_{s'}} \xrightarrow{a.s.} \frac{p_s}{p_{s'}}$, and $N_{s'}^{r\xi-1} \xrightarrow{a.s.} \infty$, , so $\frac{1}{N_{s'}^{r\xi-1}} \frac{C}{\epsilon^r} \frac{N_s}{N_{s'}} \xrightarrow{a.s.} 0$, as $n \to \infty$. This shows that:

$$P(N_{s'}^{-\xi} \max_{t_i^*=s} N_{s'}^{1/q} \left\|X_i - X_{j_m(i,s')}\right\| > \epsilon | \mathbf{t}^*) \xrightarrow{a.s.} 0.$$

On the other hand, by law of iterated expectation

$$P(N_{s'}^{-\xi} \max_{t_i^*=s} N_{s'}^{1/q} \left\|X_i - X_{j_m(i,s')}\right\| > \epsilon) = \mathbb{E}\left[P(N_{s'}^{-\xi} \max_{t_i^*=s} N_{s'}^{1/q} \left\|X_i - X_{j_m(i,s')}\right\| > \epsilon | \mathbf{t}^*)\right]$$

and since the conditional expectation is between $[0, 1]$, by the bounded convergence theorem,

$$N_{s'}^{-\xi} \max_{t_i^*=s} N_{s'}^{1/q} \left\|X_i - X_{j_m(i,s')}\right\| = o_{\mathbb{P}}(1).$$

The above equality holds for any fixed $\xi > 0$, specifically, pick $\xi = 1/q$, then $\max_{t_i^*=s} \left\|X_i - X_{j_m(i,s')}\right\| = o_{\mathbb{P}}(1)$. □

Proof of Lemma B5

By Lemma 3 in Abadie and Imbens (2006), there exists a constant C_α such that $\mathbb{E}\left[K_M^\alpha(i, s_g)\right] \leq C_\alpha$, uniformly in n. Therefore, for any $\epsilon > 0$, there exists $C_\epsilon = \frac{C_\alpha}{\epsilon}$, and $N_\epsilon = 1$ such that for all $n \geq N_\epsilon$,

by Markov's inequality,

$$\begin{aligned}\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^\alpha(i,s_g)>C_\epsilon\right] &\leq \frac{\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(t_i^*=\infty)K_M^\alpha(i,s_g)\right]}{C_\alpha/\epsilon}\\ &\leq \frac{\mathbb{E}\left[K_M^\alpha(i,s_g)\right]}{C_\alpha/\epsilon}\\ &\leq \epsilon\end{aligned}$$

which proves the claim. □

Proof of Theorem B1

The numerator of the estimator can be decomposed as:

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

where

$$\begin{aligned}\hat{\tau}_{t,t'}^{s,\infty}(w^{\mathsf{pool}}) = \hat{\tau}_{t,t'}^{s,\mathsf{pool}} &:= \frac{1}{N_s}\sum_i \mathbb{1}(t_i^*=s)\left(Y_{it}-Y_{it'}\right) - \frac{1}{n-N_0}\sum_i \mathbb{1}(t_i^*=\infty)\frac{K_M(i)}{M}\left(Y_{it}-Y_{it'}\right)\\ \hat{\tau}_{t,t'}^{s,s'}(w^{\mathsf{pool}}) = \hat{\tau}_{t,t'}^{s,s'} &:= \frac{1}{N_s}\sum_i \mathbb{1}(t_i^*=s)\left(Y_{it}-Y_{it'}\right) - \frac{1}{N_{s'}}\sum_i \mathbb{1}(t_i^*=s')\left(Y_{it}-Y_{it'}\right), \quad s'\neq\infty.\end{aligned}$$

Because $\hat{\tau}_{t,t'}^{s,s'} = -\hat{\tau}_{t,t'}^{s',s}$, the third and fifth terms in the above decomposition cancel out, yielding:

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

Notice that the third set of comparisons uses a comparison group that is treated in both periods ($s' \leq t'$), which constitutes the so-called forbidden comparisons.

Let $B_{t,t'}^{s,s'} = \mathbb{E}[Y_{it}(0) - Y_{it'}(0)|t_i^* = s] - \mathbb{E}[Y_{it}(0) - Y_{it'}(0)|t_i^* = s']$. By the law of large numbers, $\hat{p}_s \to_{\mathbb{P}} p_s$ and for $t' < s \leq t$, $s' \neq \infty$,

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

On the other hand, by Lemma B2,

$$\begin{aligned}\frac{1}{n - N_0}\sum_i \mathbb{1}(t_i^* = \infty)\frac{K_M(i)}{M}(Y_{it} - Y_{it'}) &= \sum_{s \neq \infty}\frac{N_s}{n - N_0} \cdot \frac{1}{N_s}\sum_i \mathbb{1}(t_i^* = \infty)\frac{K_M(i,s)}{M}(Y_{it} - Y_{it'}) \\ &\to_{\mathbb{P}} \sum_{s \neq \infty}\frac{p_s}{1 - p_\infty}\mathbb{E}\left\{\mathbb{E}[Y_{it} - Y_{it'}|t_i^* = \infty, X_i]\middle| t_i^* = s\right\}\end{aligned}$$

and thus

$$\begin{aligned}\hat{\tau}_{t,t'}^{s,\mathsf{pool}} &\to_{\mathbb{P}} \mathbb{E}[\tau_{it}|t_i^* = s] + \mathbb{E}[Y_{it}(0) - Y_{it'}(0)|t_i^* = s] - \sum_{s \neq \infty}\frac{p_s}{1 - p_\infty}\mathbb{E}\left\{\mathbb{E}[Y_{it} - Y_{it'}|t_i^* = \infty, X_i]\middle| t_i^* = s\right\} \\ &= \mathbb{E}[\tau_{it}|t_i^* = s] + \mathbb{E}\left\{\mathbb{E}[Y_{it}(0) - Y_{it'}(0)|t_i^* = s, X_i]\middle| t_i^* = s\right\} - \frac{p_s}{1 - p_\infty}\mathbb{E}\left\{\mathbb{E}[Y_{it} - Y_{it'}|t_i^* = \infty, X_i]\middle| t_i^* = s\right\} \\ &\quad - \sum_{s' \neq s, \infty}\frac{p_s}{1 - p_\infty}\mathbb{E}\left\{\mathbb{E}[Y_{it} - Y_{it'}|t_i^* = \infty, X_i]\middle| t_i^* = s'\right\} \\ &= \mathbb{E}[\tau_{it}|t_i^* = s] + \left(\frac{1 - p_s - p_\infty}{1 - p_\infty}\right)\mathbb{E}\left\{\mathbb{E}[Y_{it}(0) - Y_{it'}(0)|t_i^* = s, X_i]\middle| t_i^* = s\right\} \\ &\quad - \sum_{s' \neq s, \infty}\frac{p_{s'}}{1 - p_\infty}\mathbb{E}\left\{\mathbb{E}[Y_{it} - Y_{it'}|t_i^* = \infty, X_i]\middle| t_i^* = s'\right\} \\ &= \mathbb{E}[\tau_{it}|t_i^* = s] \\ &\quad + \sum_{s' \neq s, \infty}\frac{p_{s'}}{1 - p_\infty}\left(\mathbb{E}\left\{\mathbb{E}[Y_{it} - Y_{it'}|t_i^* = \infty, X_i]\middle| t_i^* = s\right\} - \mathbb{E}\left\{\mathbb{E}[Y_{it} - Y_{it'}|t_i^* = \infty, X_i]\middle| t_i^* = s'\right\}\right) \\ &= \mathbb{E}[\tau_{it}|t_i^* = s] + \sum_{s' \neq s, \infty}\frac{p_{s'}}{1 - p_\infty}B_{t,t'}^{s,s'}\end{aligned}$$

where the last line uses the conditional parallel trends assumption. Collecting the results,

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

Rearranging,

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

On the other hand, for the denominator,

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

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