Matched Triple-Differences: A Framework for Covariate Adjustment *

Yihong Liu†

August 28, 2026

Abstract

This paper studies how covariate adjustment should be conducted in triple-differences designs. The leading empirical approach is linear regression adjustment, which augments the triple-differences regression with covariates. Under the conditional parallel gap assumption, we first derive a covariate balancing condition that a weighting estimator should satisfy to consistently estimate the average treatment effect on the treated (ATT). Expressing estimators from commonly used linear regression adjustments as weighting estimators reveals that they generally fail to satisfy this balancing condition and are therefore inconsistent for the ATT. To address this limitation, we propose a *matched triple-differences* framework that achieves the required covariate balance through matching. Within this framework, we construct a consistent matched triple-differences estimator, establish its asymptotic normality, and provide a consistent variance estimator. The framework accommodates a general class of matching procedures, including nearest-neighbor matching and kernel matching.

Keywords: Triple-Differences; Difference-in-Differences-Differences; Parallel Gaps; Matching

JEL Classification: C14, C21, C23

*I thank Douglas Steigerwald, Max Farrell, and Gonzalo Vazquez-Bare for valuable comments and suggestions.

†Department of Economics, UC Santa Barbara.

1 Introduction

In many empirical settings, a policy applies only to an eligible subpopulation within the areas in which it operates. A unit is treated only if it satisfies two criteria: (i) it belongs to a group exposed to the policy, such as a county in which the program operates, and (ii) it belongs to the subpopulation eligible for the policy, such as households growing a particular crop. A leading example is Cai (2016), who studies an insurance program implemented in part of Jiangxi province, China, and targeted at tobacco-growing households, so that a household is treated only if it lives in a program area and grows tobacco. For such settings, triple-differences (DDD) is a more credible design than difference-in-differences (DiD): its identifying assumption, parallel gaps, accommodates group-specific and subpopulation-specific trends that would violate the parallel trends assumption of a DiD design. In practice, researchers often find parallel gaps more credible after conditioning on observed covariates, which leads to a conditional parallel gaps assumption. Under conditional parallel gaps, the estimation method commonly used in empirical work is linear regression adjustment, which includes the covariates linearly in the triple-differences regression.

This paper studies how covariate adjustment should be conducted in triple-differences designs to consistently estimate the average treatment effect on the treated (ATT) under the conditional parallel gap assumption. Our analysis begins with deriving the covariate balancing condition that a weighting estimator must satisfy to consistently estimate the ATT. We then recast the estimators from commonly used linear regression adjustments as weighting estimators and show that they are generally inconsistent for the ATT because they fail to attain this balancing condition. To address this limitation, we propose a *matched triple-differences* framework that attains the required balance through matching. Finally, we establish the consistency of the resulting matched triple-differences estimator and develop a valid inference procedure.

Throughout, we consider a triple-differences design with multiple periods and simultaneous adoption of treatment, and the parameters of interest are the ATTs in each post-treatment period. Our starting point is a weighting representation of the identification result. Under conditional parallel gaps, the ATT can be written as a weighting estimand in which the treated subgroup is left unweighted and each of the three untreated subgroups is reweighted by a density ratio that aligns its covariate distribution with that of the treated subgroup. This representation pins down a balancing condition: a weighting estimator consistent for the ATT must balance the covariate distribution of each untreated subgroup against that of the treated subgroup, and it must attain this balance by reweighting the three untreated subgroups while leaving the treated subgroup unweighted.

We use this condition to diagnose a prevalent estimation method in empirical work, linear regression adjustment. Empirical practice largely follows one of two specifications: in the first, the covariates enter the regression additively; in the second, they are interacted with time indicators. We Casting each regression estimator as a weighting estimator and derive the weighting estimand it induces and compare that estimand with the balancing condition. Our results show that neither specification attains the balancing condition, and both therefore yield inconsistent estimators of the ATT. When the covariates enter additively with time-invariant coefficients, they are differenced away and the estimand reduces to the unconditional triple-differences estimand. This specification is consistent for the ATT only when parallel gaps holds unconditionally, the case in which covariate adjustment was unnecessary to begin with. When the covariates are interacted with a saturated set of time indicators, all four subgroups are reweighted by weights proportional to linear projection residuals. This specification falls short of the balancing condition in two respects: it balances only covariate means, and only between the treated subgroup and a pooled untreated group that merges the three untreated subgroups, rather than the covariate distribution of each untreated subgroup separately; and it attains this balance in the wrong way, because it also reweights the treated subgroup.

Guided by the balancing condition, we propose the matched triple-differences framework for estimation and inference. Estimation proceeds in two steps. The first step conducts three pairwise matchings, each between the treated subgroup and one of the three untreated subgroups. The second step estimates a weighted, fully interacted triple-differences regression on the matched sample, using the weights generated in the matching step. The resulting matched triple-differences estimator satisfies the balancing condition by construction: matching reweights each untreated subgroup toward the covariate distribution of the treated subgroup and leaves the treated subgroup unweighted. We establish the asymptotic properties of the estimator for a general class of matching procedures, defined by a set of four conditions on the matching procedure. The class includes nearest-neighbor matching with either a fixed or a diverging number of neighbors, as well as kernel matching. Finally, we propose a unified variance estimator that is consistent for every matching procedure under our framework to conduct valid inference.

1.1 Related Literature and Contribution

This paper contributes to two literature. The first is the literature on triple-differences designs. First proposed by Gruber (1994), triple-differences is now a textbook method in empirical work; Olden and Møen (2022) formalize the triple-differences estimator in a two-period setting where covariates play no role in identification. An important recent question is how to incorporate covariates when identification relies on conditional parallel gaps. In contemporaneous work, Ortiz-Villavicencio and Sant'Anna (2025) also study identification, estimation, and inference in triple-differences under conditional parallel gaps. Through Monte Carlo simulations, they document that the linear regression adjustments similar to those studied here can produce biased estimates of the ATT. They also develop regression adjustment, inverse probability weighting, and doubly robust triple-differences estimators that rely on parametric models for the propensity score and the conditional mean of the outcome. We differ from and complement their paper in two ways. First, our diagnosis of empirical practice is analytical. By casting each linear regression estimator as a weighting estimator, we derive the estimand it induces and trace its inconsistency to a mismatch between the covariate balance it attains and the balance required by the identification argument. This explains why the common specifications fail and complements the simulation evidence in Ortiz-Villavicencio and Sant'Anna (2025). Second, our matching-based framework delivers valid estimation and inference under conditional parallel gaps without parametric models assumption for either the propensity score or the conditional mean; it is fully nonparametric and easy to implement.

This paper also contributes to the large literature on matching estimators (Abadie and Imbens, 2006) in two ways. First, it adds to the line of work that combines matching with panel data. Heckman et al. (1997) propose a difference-in-differences matching estimator for a two-period setting based on kernel matching and local linear regression; Liu and Vazquez-Bare (2026) study nearest-neighbor matching combined with difference-in-differences designs in settings with multiple periods and staggered adoption; and Imai et al. (2023) develop matching methods for panel data that match units on their treatment histories. We extend this line of work to the triple-differences setting. Second, it contributes to the recent literature on asymptotic theory and variance estimation for general classes of matching procedures (Lin and Han, 2025). Lin and Han (2025) develop a general semi-parametric efficiency theory for regression-adjusted matching estimators built on linear smoothers in the cross sectional setting. Our paper differs from and extends this literature in two respects: it is tailored to the panel data structure and to the matching procedure of triple-differences designs, and it proposes a different variance estimator, which extends the approach of Abadie and Imbens (2006) to a general class of matching procedures.

The remainder of the paper is organized as follows. Section 2 describes the setup and notation and establishes identification of the parameters of interest. Section 3 derives the balancing condition implied by the identification argument and uses it to diagnose the linear regression adjustments commonly used in empirical practice. Section 4 introduces the matched triple-differences framework, establishes the consistency and asymptotic normality of the resulting estimators, and develops the variance estimator for inference. Section 5 presents simulation studies, Section 6 illustrates the methods through a reanalysis of Cai (2016), and Section 7 concludes.

2 Setup, Estimand, and Identification

2.1 Econometrics Set Up

Consider a balanced panel data of n observations over T periods, $t = 1, \ldots, T$. The treatment is implemented simultaneously in a single period t^*, with $2 \leq t^* \leq T$, so that $t < t^*$ are the pre-treatment periods and $t \geq t^*$ are the post treatment periods. Each unit i is characterized by two time-invariant binary indicators: an exposure indicator G_i and an eligibility indicator S_i. Specifically,

- the *exposure* indicator $G_i \in \{0, 1\}$ indicates whether the unit i belongs to a group (e.g., a county) that is exposed to the treatment in the post treatment period.
- the *eligibility* indicator $S_i \in \{0, 1\}$ indicates whether unit i belongs to a subpopulation (e.g., households growing a specific crop) that is eligible for the treatment.

These two indicators partition the population into four cross-sectional subgroups displayed in Figure 1. The key feature of the triple-differences is that only units satisfying both criteria are treated: a unit must belong to the exposed group and to the eligible subpopulation. The treatment is an absorbing state, so once a unit gets treated, it remains to be treated throughout period T. Let $D_i \in \{0, 1\}$ indicate whether unit i is treated in the post treatment periods, and let $D_{it} \in \{0, 1\}$ indicate whether whether unit i is treated in period t.

$$D_i = G_i S_i, \quad D_{it} = G_i S_i \mathbb{1}(t \geq t^*)$$

In Figure 1, the treated subgroup is the top-left cell, $(G_i = 1, S_i = 1)$; we refer to the remaining three cells as the *untreated subgroups*.

Throughout the paper, I will use Cai (2016) to fix ideas. Cai (2016) uses household-level panel data from 12 tobacco-producing counties in Jiangxi Province, China, to study how an agricultural insurance program affects household borrowing and saving behavior. In 2003, one county collaborated with the People's Insurance Company of China (PICC) to design and launch a tobacco production insurance program and only the tobacco-growing households in that county were offered the insurance program. In our notation, the exposure indicator G_i

indicates whether household i is located in the county that launched the program (hereafter, the exposed county), the eligibility indicator S_i indicates whether household i grows tobacco and $t^* = 2003$. Only the tobacco-growing households located in the exposed county are treated in the post treatment periods.

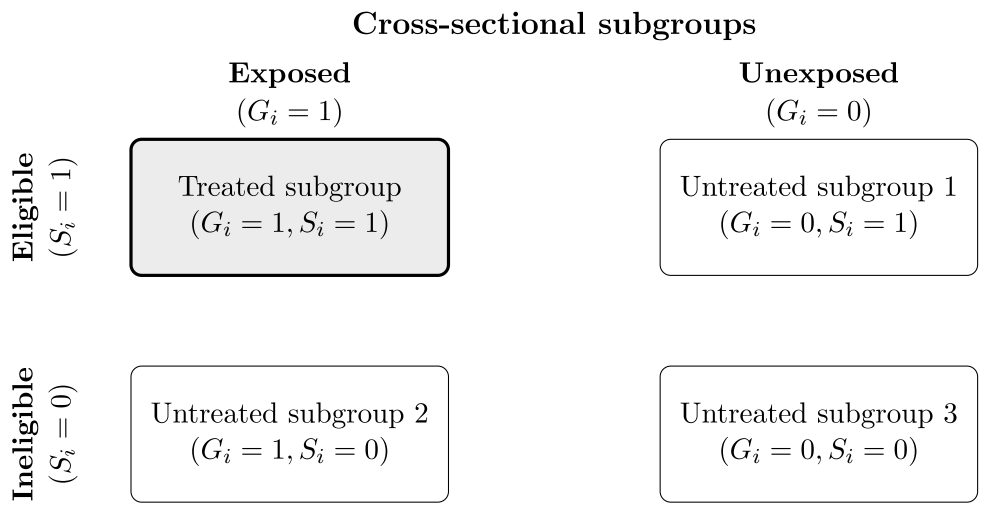

Figure 1: Four cross-sectional subgroups

We adopt the potential-outcome framework in which the potential outcome depends on the entire treatment sequence. Let $Y_{it}(d_{i1}, \ldots, d_{iT})$ denote the potential outcome in period t under treatment sequence $(d_{i1}, \ldots, d_{iT})$, and let $\mathbf{0}_s$ and $\mathbf{1}_s$ be s-dimensional vector of zeros and ones. Because the adoption is simultaneous and the treatment absorbing, only two treatment paths can occur in our setting: treated units follow $(\mathbf{0}_{t^*-1}, \mathbf{1}_{T-t^*+1})$ and untreated units follow $\mathbf{0}_T$. We therefore abbreviate

$$Y_{it}(1) \equiv Y_{it}(\mathbf{0}_{t^*-1}, \mathbf{1}_{T-t^*+1}), \quad Y_{it}(0) \equiv Y_{it}(\mathbf{0}_T)$$

We assume the SUTVA and consistency to link the observed and potential outcomes; as is common in causal inference literature in panel data (de Chaisemartin and D'Haultfœuille, 2022;Roth et al., 2023). As a result,

$$Y_{it} = Y_{it}(1)D_i + Y_{it}(0)(1 - D_i)$$

In many empirical settings, researchers view their identification assumption for triple-differences as more credible after conditioning on observed covariates. In our running example, Cai (2016) controls for household's characteristics such as the household head's age and education. We let $X_i \in \mathbb{R}^q$ denote a vector of time-invariant, pre-treatment covariates of dimension q. Finally, we assume that we observe a random sample of size n from the population distribution.

Assumption 1. (Random Sampling) $\{Y_{i1}, Y_{i2}, \ldots, Y_{iT}, G_i, S_i, X_i\}_{i=1}^{n}$ is a random sample from the population $(Y_1, Y_2, \ldots, Y_T, G, S, X)$

2.2 Estimand and Identification Assumption

Our parameter of interest is the average treatment effect on the treated subgroup (ATT) l periods after the initial treatment period,

$$\tau_l = \mathbb{E}[Y_{t^*+l}(1) - Y_{t^*+l}(0) \mid D = 1] = \mathbb{E}[Y_{t^*+l}(1) - Y_{t^*+l}(0) \mid G = 1,\ S = 1]\,, \quad l = 0, 1, \ldots, T - t^*$$

In our empirical example, τ_l corresponds to the average effect of the insurance program on the borrowing and saving behavior of tobacco-growing household in the exposed county, l years after program's adoption.

To identify τ_l, we make the following three assumptions:

Assumption 2. (no anticipation) For all $t = 1, \ldots, T$, $Y_t(d_1, \ldots, d_T) = Y_t(d_1, \ldots, d_t)$

The no-anticipation assumption requires that future treatments do not affect current outcomes: the potential outcome in period t depends only on the treatment path up to period t.

Assumption 3. (Strong Overlap) There exists $\epsilon > 0$ such that,

$$\Pr(G = g, S = s \mid X = x) > \epsilon$$

for every $(g, c) \in \{0, 1\}^2$ and almost every $x \in \mathcal{X}$, where $\mathcal{X}$ denotes the support of X.

The strong overlap assumption requires that, at almost every covariate value x, each of the four subgroups occurs with probability bounded away from zero: within each covariate stratum, the population contains units from all four subgroups of Figure 1.

The key identifying assumption of the triple-differences replaces the familiar parallel trends condition from the difference-in-differences (DiD) designs with a parallel gaps assumption, following the recent triple-differences literature (Olden and Møen, 2022; Ortiz-Villavicencio and Sant'Anna, 2025).

Assumption 4. (Conditional Parallel Gaps)

$$\begin{aligned}&\mathbb{E}[\Delta Y_t(0) \mid G = 1,\ S = 1,\ X] - \mathbb{E}[\Delta Y_t(0) \mid G = 0,\ S = 1,\ X] \\ &= \mathbb{E}[\Delta Y_t(0) \mid G = 1,\ S = 0,\ X] - \mathbb{E}[\Delta Y_t(0) \mid G = 0,\ S = 0,\ X],\ a.s.\end{aligned}$$

where $\Delta Y_t(0) \equiv Y_{it}(0) - Y_{it-1}(0)$

The conditional parallel gaps condition requires that, conditional on the covariates X and in absence of treatment, the gap in the average outcome trends between the exposed and unexposed units is the same in the eligible subpopulation as in the ineligible subpopulation. The assumption therefore allows the DiD comparison among eligible units to violate conditional parallel trends, it only requires that the violation coincide with the corresponding violation among ineligible units. Figure 2 illustrates the idea: at a covariate value x, the exposed-unexposed trend gap among eligible units, $\delta_1(x)$ may be nonzero, but it must equal the corresponding gap among ineligible units, $\delta_0(x)$.

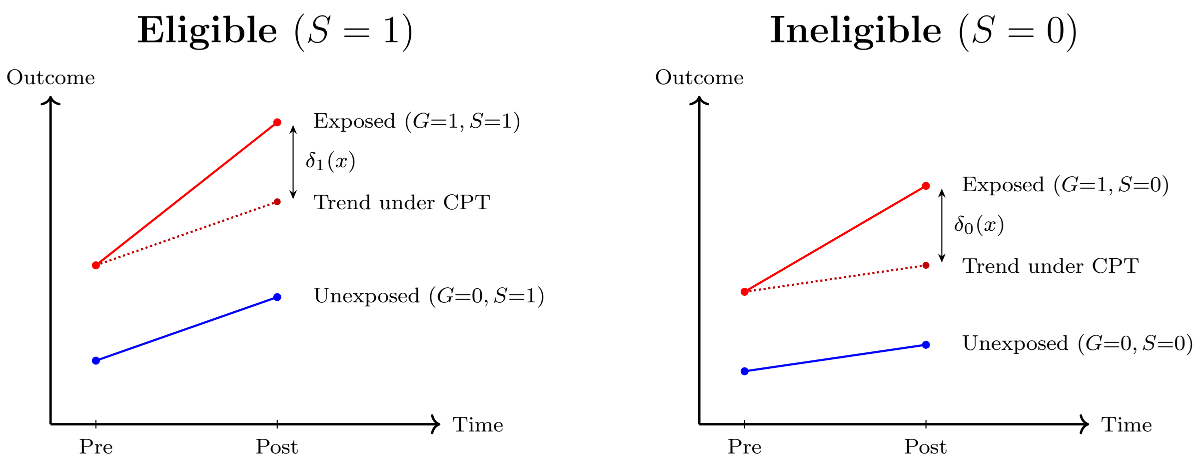

Figure 2: Conditional parallel gaps

Note: This figure illustrates the conditional parallel-gaps assumption at a fixed covariate value $X = x$. In each panel, the dashed line is the trend the exposed units would follow if conditional parallel trends (CPT) held. In the left panel, $\delta_1(x)$ is the deviation from conditional parallel trends in untreated outcomes between eligible exposed units, $(G = 1, S = 1)$, and eligible unexposed units, $(G = 0, S = 1)$. In the right panel, $\delta_0(x)$ is the corresponding deviation between ineligible exposed units, $(G = 1, S = 0)$, and ineligible unexposed units, $(G = 0, S = 0)$. The assumption allows conditional parallel trends to fail within the eligible comparison, but requires the failure to be the same as in the ineligible comparison: $\delta_1(x) = \delta_0(x)$.

We emphasize that the conditional parallel gap is not a weaker version of the conditional parallel trends that would justify a DiD design among eligible units,

$$\mathbb{E}[\Delta Y(0) \mid G = 1,\ S = 1,\ X] = \mathbb{E}[\Delta Y(0) \mid G = 0,\ S = 1,\ X] \quad a.s.$$

Neither condition implies the other. A triple-differences design should therefore not be interpreted as a mechanical robustness check on a DiD design: the two designs rest on distinct identifying assumptions, and their credibility depends on the empirical context. Introducing the second comparison dimension improves credibility when the ineligible units are subject to the same parallel-trends violation as the eligible units, but it can introduce bias when the parallel gaps condition itself is implausible.

In our empirical example, the conditional parallel gap assumption allows the untreated financial behavior for tobacco-growing households in the exposed county to evolve differently from that of tobacco-growing households in the unexposed counties. It requires that, conditional on the household characteristics X_i, this differential evolution equal the corresponding differential for non-tobacco households across the exposed and unexposed counties.

The following theorem gives the identification result for the average treatment effect on the treated subgroup.

Theorem 1. Under Assumptions 2, 3, and 4, the average treatment effect on the treated subgroup is identified for every $l = 0, 1, \ldots, T - t^*$,

$$\begin{aligned}\tau_l = {} & \mathbb{E}[\Delta Y_l \mid G = 1, S = 1] - \mathbb{E}[\mathbb{E}[\Delta Y_l \mid G = 0, S = 1, X] \mid G = 1, S = 1] \\ & - \mathbb{E}[\mathbb{E}[\Delta Y_l \mid G = 1, S = 0, X] \mid G = 1, S = 1] + \mathbb{E}[\mathbb{E}[\Delta Y_l \mid G = 0, S = 0, X] \mid G = 1, S = 1]\,,\end{aligned}$$

where $\Delta Y_l \equiv Y_{t^*+l} - Y_{t^*-1}$ and $f_{gs}(x)$ denotes the conditional density of X given $(G, S) = (g, s)$.

Remark 2.1. Assumption 3 is stronger than what identification of τ_t requires: it would suffice to impose overlap almost everywhere on the support of X conditional on the treated group, rather than on the whole support $\mathcal{X}$. We maintain the stronger version for simplicity.

3 Diagnosis of Empirical Practice

Empirical researchers incorporate covariates into triple-differences designs mainly through *parametric adjustment*: augmenting the triple-differences regression with linear functions of the covariates. In this section, we diagnose the two leading empirical implementations by casting each linear regression estimator as a weighting estimator and analyzing the induced weighting estimand.[1] Our results show that common empirical practice generally fails to consistently estimate the ATT, τ_l, because the induced weights do not correctly align the covariate distributions of the untreated subgroups with that of the treated subgroup.

We first derive the balancing condition and the weight profile that a weighting estimator requires to consistently estimate τ_l.

[1] Chattopadhyay and R Zubizarreta (2023) and Caetano and Callaway (2024) use a similar approach to study the causal interpretation of regression coefficients under linear covariate adjustment in cross-sectional and DiD settings, respectively.

Lemma 1. Under Assumptions 2, 3, and 4, the average treatment effect on the treated subgroup can be rewritten as the following weighting estimand,

$$\tau_l = \mathbb{E}[\Delta Y_l \mid G = 1, S = 1] - \mathbb{E}\left[\frac{f_{11}(X)}{f_{01}(X)}\Delta Y_l \mid G = 0, S = 1\right] \\ - \mathbb{E}\left[\frac{f_{11}(X)}{f_{10}(X)}\Delta Y_l \mid G = 1, S = 0\right] + \mathbb{E}\left[\frac{f_{11}(X)}{f_{00}(X)}\Delta Y_l \mid G = 0, S = 0\right],$$

Lemma 1 gives a weighting representation of the identification result: the treated subgroup is left unweighted, and each untreated subgroup (g, s) is reweighted by the density ratio $f_{11}(X)/f_{gs}(X)$. The density ratio is a balancing weight: it aligns the covariate distribution of the untreated subgroup (g, s) with that of the treated subgroup.

From Lemma 1, we obtain a balancing condition for consistent estimation of the ATT in the triple-differences design: the weights induced by a weighting estimator should (i) balance the covariate distribution of *each* untreated subgroup against that of the treated subgroup, and (ii) attain this balance while leaving the treated subgroup unweighted. Conditions (i) and (ii) together pin down the balancing weight profile

$$w_{11}(X) = 1, \qquad w_{gs}(X) = \frac{f_{11}(X)}{f_{gs}(X)} \quad \text{for } (g, s) \in \{(0, 1), (1, 0), (0, 0)\}. \tag{1}$$

The balancing condition provides a two-part diagnostic for analyzing two two commonly used parametric adjustment methods in empirical literature. For each method, we cast the induced estimand as a weighting estimand and determine (i) which features of the covariate distributions the induced weight profile balances, and (ii) how it attains this balance compared with the balancing weight profile (1).

Parametric adjustment in the empirical literature largely follows two specifications. In the first specification, the covariates X_i enter the fully saturated triple-differences regression additively, with time-invariant coefficients:

$$Y_{it} = \lambda_{G_i S_i} + \lambda_{G_i t} + \lambda_{S_i t} + \sum_{l \neq -1} \tau_l \, D_i \mathbb{1}(t - t^* = l) + X_i'\theta + u_{it}. \tag{2}$$

The second specification interacts the covariates with a saturated set of time indicators, allowing for covariate-specific time trends:

$$Y_{it} = \lambda_{G_i S_i} + \lambda_{G_i t} + \lambda_{S_i t} + \sum_{l \neq -1} \tau_l \, D_i \mathbb{1}(t - t^* = l) + \sum_{r=1}^{T} X_i'\theta_r \mathbb{1}(t = r) + u_{it}. \tag{3}$$

By the Frisch-Waugh-Lovell Theorem, each OLS estimators converges to a weighting estimand that implicitly reweights the four subgroups, and Lemma 2 characterizes their induced weighting estimands. Let $L(D \mid G, S, X)$ denote the population linear projection of D onto $(1, G, S, X)$, $\kappa \equiv \mathbb{E}[(D - L(D \mid G, S, X))^2]$ denote the mean squared projection residual, and let $p_{gs} = P(G = g, S = s)$ denote the subgroup proportion.

Lemma 2. Under Assumptions 1, 3, and 6, the weighted estimand induced by the regression specifications (2) and (3) are:

$$
\begin{aligned}
\tau_l^1 &= \mathbb{E}[\Delta Y_l \mid G=1, S=1] - \mathbb{E}[\Delta Y_l \mid G=0, S=1] - \mathbb{E}[\Delta Y_l \mid G=1, S=0] + \mathbb{E}[\Delta Y_l \mid G=0, S=0], \\
\tau_l^2 &= \mathbb{E}\left[\frac{p_{11}\,(1 - L(D \mid G, S, X))}{\kappa}\,\Delta Y_l \,\middle|\, G=1, S=1\right] - \mathbb{E}\left[\frac{p_{01}\,L(D \mid G, S, X)}{\kappa}\,\Delta Y_l \,\middle|\, G=0, S=1\right] \\
&\quad - \mathbb{E}\left[\frac{p_{10}\,L(D \mid G, S, X)}{\kappa}\,\Delta Y_l \,\middle|\, G=1, S=0\right] + \mathbb{E}\left[\frac{-p_{00}\,L(D \mid G, S, X)}{\kappa}\,\Delta Y_l \,\middle|\, G=0, S=0\right].
\end{aligned}
$$

The first specification induces the constant weight profile $\widetilde{w}_{gs}(X) = 1$ for all four subgroups, so τ_l^1 is exactly the unconditional triple-differences estimand. The covariates drop out because they enter (2) additively with time-invariant coefficient and are therefore differenced away. Consequently, this weighting profile does not balance the covariate distribution and specification (2) recovers τ_l only when the parallel-gaps condition holds *unconditionally* in which covariate adjustment was unnecessary.

Specification (3) does adjust the covariates, through weights proportional to the linear projection residual $D - L(D \mid G, S, X)$. These weights are the triple-differences analogues of the implied regression weights studied by Chattopadhyay and R Zubizarreta (2023) in cross-sectional settings and by Caetano and Callaway (2024) in DiD settings. In those two settings, residual-based regression weights have an appealing property: they balance the covariate means of the treated and untreated groups. However, in triple-differences setting, the residual regression weighting profile induced by the specification (3) fails on both aspects of our diagnostic.

First, it balances a wrong object. Lemma 4 in the appendix shows that the weights in τ_l^2 continue to balance covariate means, between the treated subgroup ($D = 1$) and the pooled untreated group ($D = 0$) that merges the three untreated subgroups. In a triple-differences design this is the wrong object because the untreated group now pools three distinct untreated subgroups, while the balancing weight (1) requires aligning the covariate distribution of *each* of them with the treated subgroup separately. Balancing the mean of the pooled untreated group does not imply balancing mean and distribution of the three subgroup. Second, it attains this pooled balance in the wrong manner. The balancing weight profile leaves the treated subgroup unweighted and aligns each untreated subgroup with it; the weights in τ_l^2 instead reweight all four subgroups, including the treated subgroup ($\widetilde{w}_{11}(X) = p_{11}(1 - L(D \mid G, S, X))/\kappa \neq 1$).

Remark 3.1. Specification (3) can be enriched by further interacting the covariates with the subgroup indicators G_i and S_i. Such interactions can attain mean balance between each untreated subgroup and the treated subgroup, partially addressing the first issue above. They still do not satisfy the covariate balancing condition: (i) mean balance between the four subgroups does not imply the balance of covariate distribution, and (ii) the implied weights continue to reweight the treated subgroup. Hence the enriched specification does not consistently estimate τ_l in general.

4 Matched Triple-Differences: Estimation and Inference

Section 3 established a balancing condition for consistent estimation of the ATT: the weights induced by a weighting estimator should align the covariate distribution of each untreated subgroup with that of the treated subgroup through the balancing weight profile (1). Weighting estimator that fail this condition are generally inconsistent for the ATT without additional assumptions. Guided by this condition, we propose a matched triple-differences framework that attains the required balance directly, by matching the treated subgroup to each of the three untreated subgroups. The resulting matched triple-differences estimator is consistent for the ATT, and we provide a valid inference procedure. Our framework is general: it accommodates the matching methods most commonly used in practice, including nearest-neighbor matching and kernel matching.

4.1 Estimation

Estimation in our framework can be viewed as a two-step procedure: a matching step, followed by a weighted regression on the matched sample. In the first step, we conduct three pairwise matchings on the observed covariates X. Let

$$\mathcal{I}_{gs} \equiv \{i : G_i = g,\, S_i = s\}\,, \qquad (g, s) \in \{0, 1\}^2,$$

denote the four subgroups. Each pairwise matching involves the treated subgroup $\mathcal{I}_{11}$ and one of the three untreated subgroups $\mathcal{I}_{gs}$, $(g, s) \neq (1, 1)$: it assigns a non-negative *pairwise weight* $w_{ij}(\mathcal{I}_{11}, \mathcal{I}_{gs})$ to each treated–untreated pair $(i, j) \in \mathcal{I}_{11} \times \mathcal{I}_{gs}$. The pairwise weights are computed by a prespecified matching algorithm (e.g., nearest-neighbor or kernel matching) from the subgroup indicators (G, S) and the observed covariates X of the two subgroups. The *matched set* of treated unit i collects the untreated units that receive positive weight,

$$\mathcal{M}_{gs}(i) \equiv \{j \in \mathcal{I}_{gs} : w_{ij}(\mathcal{I}_{11}, \mathcal{I}_{gs}) > 0\},$$

so the pairwise weight w_{ij} measures the contribution of untreated unit j to the match for treated unit i. An untreated unit may contribute to the matched sets of several treated units; summing the pairwise weights it receives across treated units yields its *matching weight*,

$$w_j(\mathcal{I}_{11}, \mathcal{I}_{gs}) \equiv \sum_{i \in \mathcal{I}_{11}} w_{ij}(\mathcal{I}_{11}, \mathcal{I}_{gs}), \qquad j \in \mathcal{I}_{gs},$$

which measures unit j's total contribution to the matched sample. The two objects play distinct roles: the pairwise weights $\{w_{ij}\}_{j \in \mathcal{I}_{gs}}$ describe how the matched set of a single treated unit is constructed, while the matching weight w_j records how intensively untreated unit j is used overall. To avoid clutter, we suppress the dependence of the weights on the pair of subgroups and write w_{ij} and w_j when no confusion arises.

Altogether, the first step assigns to each unit i in the sample the weight

$$w_i = \begin{cases} 1 & i \in \mathcal{I}_{11}, \\ w_i(\mathcal{I}_{11}, \mathcal{I}_{01}) & i \in \mathcal{I}_{01}, \\ w_i(\mathcal{I}_{11}, \mathcal{I}_{10}) & i \in \mathcal{I}_{10}, \\ w_i(\mathcal{I}_{11}, \mathcal{I}_{00}) & i \in \mathcal{I}_{00}. \end{cases}$$

The weight structure mirrors the balancing weight profile (1): the treated subgroup is unweighted, and each untreated subgroup is reweighted toward the treated subgroup through the matching weights. The three pairwise matchings and the corresponding weights are visualized in Figure 3.

In the second step, we run a weighted least-squares regression on the matched sample with weights w_i:

$$\sqrt{w_i} Y_{it} = \sqrt{w_i} \left(\lambda_{G_i S_i} + \lambda_{G_i t} + \lambda_{S_i t} + \sum_{l \neq -1} \tau_l G_i S_i \mathbb{1}(t - t^* = l) + u_{it} \right). \tag{4}$$

The coefficient on the triple interaction term at event time l is our matched triple-differences estimator, which takes the form

$$\widehat{\tau}_l = \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \Delta Y_{il} - \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{01}} w_i \Delta Y_{il} - \left(\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{10}} w_i \Delta Y_{il} - \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{00}} w_i \Delta Y_{il} \right).$$

In the empirical example of Section 2, the researcher would construct the matched triple-differences estimator as follows. First, each tobacco household i in the exposed counties ($i \in \mathcal{I}_{11}$) is matched to similar tobacco households in the unexposed counties ($\mathcal{I}_{01}$) based on household-head characteristics X. Second, each such household is matched, in the same fashion, to the non-tobacco households in the exposed counties ($\mathcal{I}_{10}$). Third, it is matched to the non-tobacco households in the unexposed counties ($\mathcal{I}_{00}$). The weighted regression (4) on the resulting matched sample then yields the matched triple-differences estimator.

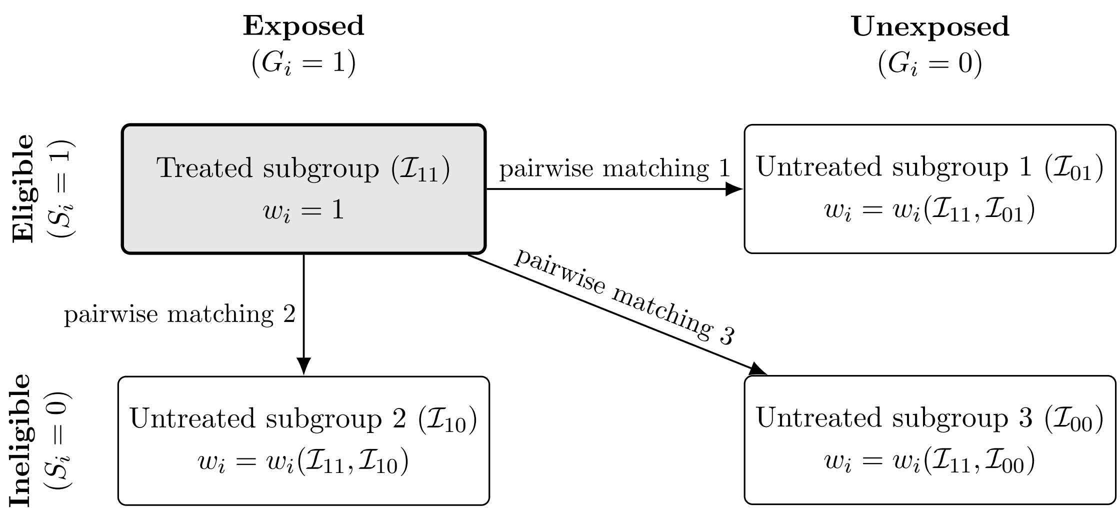

Figure 3: Matched Triple-Differences

Note. The matched triple-differences performs three separate pairwise matchings. In each pairwise matching, every unit in the treated subgroup ($\mathcal{I}_{11}$) is matched to similar units in one untreated subgroup, based on observed covariates X_i.

4.2 Asymptotic Theory

We now show that the matched triple-differences estimator is consistent and asymptotically normal under mild regularity conditions on the matching procedure together with standard regularity conditions on the data.

Assumption 5 (Regularity of the matching procedure)**.** The pairwise matching procedure between the treated subgroup $\mathcal{I}_{11}$ and untreated subgroup $\mathcal{I}_{gs}$ satisfies the following conditions.

(i) The pairwise weights $\{w_{ij} : i \in \mathcal{I}_{11},\ j \in \mathcal{I}_{gs}\}$ are constructed using only the subgroup indicators and observed covariates of the treated and untreated groups, and they do not depend on the order in which the observations are supplied: permuting the observations within $\mathcal{I}_{11}$ and $\mathcal{I}_{gs}$ permutes the corresponding pairwise weights accordingly. Moreover, for each $i \in \mathcal{I}_{11}$,

$$w_{ij} \geq 0 \quad \text{for every } j \in \mathcal{I}_{gs}, \qquad \sum_{j \in \mathcal{I}_{gs}} w_{ij} = 1.$$

(ii)

$$\mathbb{E}\left[\sum_{j \in \mathcal{I}_{gs}} w_{ij} \, \|X_j - X_i\| \,\middle|\, i \in \mathcal{I}_{11}\right] = o(1).$$

(iii) For some $\delta > 0$,

$$\mathbb{E}\big[w_j^{4+\delta} \,\big|\, j \in \mathcal{I}_{gs}\big] = O(1).$$

(iv) Let $w_{ij}^{(1)}$ and $w_j^{(1)}$ denote the pairwise matching weight calculated after replacing the first observation $(Y_{11}, \ldots, Y_{1T}, G_1, S_1, X_1)$ with an independent copy $(Y_{1'1}, \ldots, Y_{1'T}, G_{1'}, S_{1'}, X_{1'})$.

$$\mathbb{E}\left[\left(\sum_{j \in \mathcal{I}_{gs} \setminus \{1\}} \big|\, w_j - w_j^{(1)} \big|\right)^4\right] = O(1).$$

Condition (i) restricts the information available to the matching procedure: the pairwise weights depend only on the subgroup indicators and observed covariates, and they are invariant to the order in which observations are supplied. The non-negativity and adding-up requirements imply that the matched counterpart of each treated unit is a convex combination of untreated units.

Condition (ii) imposes locality: the covariate discrepancy between a treated unit and its matched untreated units, averaged using the pairwise weights, vanishes as the sample grows. In other words, each treated unit is asymptotically matched to untreated units with nearly identical covariate values.

Condition (iii) restricts how intensively an individual untreated unit may be reused across matched sets. Because the matching weight w_j measures unit j's total contribution to the matched sample, the moment bound rules out matched samples that rely on only a handful of heavily reused untreated units.

Condition (iv) is a stability requirement on the matching procedure: it rules out procedures in which a single observation can exert a large influence on the overall allocation of the matching weights. The condition is technical but not innocuous, as it is required to prove the *unconditional* asymptotic normality of our estimator. Absent such a condition, asymptotic normality is guaranteed only conditionally on the subgroup indicators and the covariates. In Appendix B, we provide an example of a matching procedure that satisfies conditions (i)–(iii) but not (iv), for which the resulting matched triple-differences estimator fails to be asymptotically normal unconditionally.

Assumption 5 nests several commonly used matching procedures. In particular, Lemma 3 in the appendix shows that two popular procedures satisfy it: nearest-neighbor matching with replacement (with either a fixed or a diverging number of neighbors) and kernel matching. The assumption does exclude some procedures used in the empirical literature. Nearest-neighbor matching *without* replacement violates condition (i), because its pairwise weights depend on the order in which the observations are supplied; local linear matching violates condition (i) as well, because its pairwise weights can be negative.

We also impose the following regularity conditions that are standard in the matching literature (Abadie and Imbens, 2006).

Assumption 6 (Regularity conditions)**.** For every $(g, s) \in \{0, 1\}^2$, $d \in \{0, 1\}$, and $t = 1, \dots, T$, and every event time l:

(i) the conditional density f_{gs} of X given $(G, S) = (g, s)$ is continuous, has convex and compact support, and is bounded and bounded away from zero on its support;

(ii) $\mathbb{E}[Y_t(d) \mid G = g, S = s, X = x]$ is Lipschitz continuous in x;

(iii) $\mathrm{Var}[\Delta Y_l(d) \mid G = g, S = s, X = x]$ is Lipschitz continuous in x and uniformly bounded, that is $0 < \underline{\sigma}^2 \leq \mathrm{Var}[\Delta Y_l(d) \mid G = g, S = s, X = x] \leq \overline{\sigma}^2$;

(iv) $\mathbb{E}[Y_t(d)^4 \mid G = g, S = s, X = x]$ is uniformly bounded.

For notation simplicity, we make the following definition. $\mu_l^{gs}(x) = \mathbb{E}[\Delta Y_l | G = g, S = s, X = x]$, and $\sigma^2_{gs,l}(x) = Var(\Delta Y_l | G = g, S = s, X = x)$.

Theorem 2. Under assumption 1, 2, 3, 4, 5, and 6,

$$\widehat{\tau}_l - \tau_l \xrightarrow{p} 0, \quad (V_1 + V_2)^{-1/2} \sqrt{n_{11}} \, (\widehat{\tau}_l - \tau_l - B) \xrightarrow{d} \mathcal{N}(0, 1)$$

where,

$$V_1 = \mathbb{E}\left[\left(\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i) - \tau_l \right)^2 \middle| G_i = 1, S_i = 1 \right]$$

$$V_2 = \frac{1}{n_{11}} \left(\sum_{i \in \mathcal{I}_{11}} \sigma_{11,l}^2(X_i) + \sum_{i \in \mathcal{I}_{01}} w_i^2 \sigma_{01,l}^2(X_i) + \sum_{i \in \mathcal{I}_{10}} w_i^2 \sigma_{10,l}^2(X_i) + \sum_{i \in \mathcal{I}_{00}} w_i^2 \sigma_{00,l}^2(X_i) \right)$$

$$B = \frac{1}{n_{11}} \sum_{i=1}^{n} \left[\mathbb{1}(i \in \mathcal{I}_{11}) - \mathbb{1}(i \in \mathcal{I}_{01}) w_i \right] \mu_l^{01}(X_i) + \frac{1}{n_{11}} \sum_{i=1}^{n} \left[\mathbb{1}(i \in \mathcal{I}_{11}) - \mathbb{1}(i \in \mathcal{I}_{10}) w_i \right] \mu_l^{10}(X_i)$$
$$- \frac{1}{n_{11}} \sum_{i=1}^{n} \left[\mathbb{1}(i \in \mathcal{I}_{11}) - \mathbb{1}(i \in \mathcal{I}_{00}) w_i \right] \mu_l^{00}(X_i)$$

Theorem 2 establishes consistency and asymptotic normality for the matched triple-differences estimators that fall within our framework. It extends classical large-sample results for matching estimators, such as those of Abadie and Imbens (2006), to the triple-differences setting and to a general class of matching procedures. The variance has two components: V_1 captures the variance of the effect heterogeneity and V_2 captures the average conditional outcome variances.

Theorem 2 also shows that our matched triple-differences estimator carries a finite-sample bias term B, which collects the biases from the three pairwise matchings. The bias term does not affect consistency, but it can affect the inference: $\sqrt{n_{11}}B$ need not be asymptotically negligible, so the matched triple-differences estimator may need bias correction to conduct valid inference. This issue parallels the one first pointed out by Abadie and Imbens (2006) for cross-sectional matching estimators. In our setting, bias correction can be carried out by applying the procedures of Abadie and Imbens (2011) and Lin and Han (2025) separately to each of the three bias terms. In Appendix C, we present the bias-correction procedure and establish the asymptotic properties of the bias-corrected estimator.

4.3 Variance Estimation

To conduct inference, we now propose a unified variance estimator that applies to all matching procedures within our framework. In view of Theorem 2, two components need to be estimated: V_1, the effect-heterogeneity component, and V_2, the conditional-variance component.

To estimate V_2, since the matching weights are computed from the data, the only unknown components of V_2 are the conditional variance functions $\sigma_{gs,l}^2(\cdot)$ of the four subgroups. We estimate them with the strategy proposed by Abadie and Imbens (2006): within-subgroup nearest-neighbor matching with a fixed number K of neighbors. Specifically, for each unit

j, let $k_m(j)$ denote the index of the m-th closest unit to j in covariate distance among the units in the same subgroup (G_j, S_j). We then estimate the conditional variance as

$$\widehat{\sigma}^2_{gs,l}(X_j) = \frac{K}{K+1}\left(\Delta Y_{jl} - \frac{1}{K}\sum_{m=1}^{K}\Delta Y_{k_m(j)l}\right)^2,$$

Substituting these estimates into the expression for V_2 yields

$$\widetilde{V}_2 = \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{11}}\widehat{\sigma}^2_{11,l}(X_j) + \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{01}} w_j^2\widehat{\sigma}^2_{01,l}(X_j) + \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{10}} w_j^2\widehat{\sigma}^2_{10,l}(X_j) + \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{00}} w_j^2\widehat{\sigma}^2_{00,l}(X_j).$$

To estimate V_1, we use a sample analog. For each treated unit i, we replace the conditional mean $\mu_l^{11}(X_i)$ with the unit's own outcome ΔY_{il} and each untreated conditional mean $\mu_l^{gs}(X_i)$ with the weighted average of the matched untreated outcomes, $\widehat{\Delta Y_{il}^{gs}} \equiv \sum_{j\in\mathcal{I}_{gs}} w_{ij}\Delta Y_{jl}$ for $(g,s)\neq(1,1)$, and the treatment effect τ_l with the matched triple-differences estimator $\widehat{\tau}_l$. The resulting estimator is

$$\widehat{V}_1 = \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\Delta Y_{il} - \widehat{\Delta Y_{il}^{01}} - \widehat{\Delta Y_{il}^{10}} + \widehat{\Delta Y_{il}^{00}} - \widehat{\tau}_l\right)^2.$$

Note that $\widehat{V}_1$ does not estimate V_1 alone: it captures not only the treatment-effect heterogeneity but also the noise introduced by replacing the conditional means with outcomes and matched averages. To see this, let $\widetilde{w}_j^2 \equiv \sum_{i\in\mathcal{I}_{11}} w_{ij}^2$ denote the sum of the squared pairwise weights received by untreated unit j. Then,

$$\widehat{V}_1 = V_1 + \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_j) + \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{01}}\widetilde{w}_j^2\sigma^2_{01,l}(X_j) + \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{10}}\widetilde{w}_j^2\sigma^2_{10,l}(X_j) + \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{00}}\widetilde{w}_j^2\sigma^2_{00,l}(X_j) + o_p(1).$$

Comparing this display with $\widetilde{V}_2$ shows that $\widehat{V}_1$ already absorbs part of $\widetilde{V}_2$. Removing the overlap part from $\widetilde{V}_2$ delivers our final variance estimator, $\widehat{V}_1 + \widehat{V}_2$, where

$$\widehat{V}_1 = \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\Delta Y_{il} - \widehat{\Delta Y_{il}^{01}} - \widehat{\Delta Y_{il}^{10}} + \widehat{\Delta Y_{il}^{00}} - \widehat{\tau}_l\right)^2,$$
$$\widehat{V}_2 = \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{01}}\left(w_j^2 - \widetilde{w}_j^2\right)\widehat{\sigma}^2_{01,l}(X_j) + \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{10}}\left(w_j^2 - \widetilde{w}_j^2\right)\widehat{\sigma}^2_{10,l}(X_j) + \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{00}}\left(w_j^2 - \widetilde{w}_j^2\right)\widehat{\sigma}^2_{00,l}(X_j).$$

Theorem 3. Under Assumptions 1, 2, 3, 4, 5, and 6,

$$\widehat{V}_1 + \widehat{V}_2 - (V_1 + V_2) \xrightarrow{p} 0.$$

Theorem 3 shows that our variance estimator is consistent under the same assumptions as Theorem 2. Our variance estimator extends that of Abadie and Imbens (2006) in two directions. First, it is tailored to the data structure and estimation procedure of the triple-differences design. Second, the Abadie–Imbens estimator and its consistency proof are specific to nearest-neighbor matching, ours is valid for the general class of matching procedures characterized by Assumption 5.

5 Simulation

In this section, I use Monte Carlo simulations to illustrate: (i) the asymptotic bias induced by the two matching workflows commonly used in the applied literature, and (ii) the consistency of the proposed matched DDD estimator and its variance estimator.

The simulation design has two periods, two binary group indicators, and a scalar covariate. As in the main setup, only units in subgroup $(G_i, C_i) = (1, 1)$ are treated in the post-treatment period. I begin by specifying the group assignment probabilities for the four cross-sectional groups $(g, c) \in \{0, 1\}^2$. For simplicity, I assume that $G_i \perp\!\!\!\perp C_i \mid X_i$. Thus, for any $(g, c) \in \{0, 1\}^2$,

$$P(G_i = g, C_i = c \mid X_i) = P(G_i = g \mid X_i)P(C_i = c \mid X_i),$$

with

$$P(G_i = 1 \mid X_i) = \frac{\exp(\gamma_G X_i)}{1 + \exp(\gamma_G X_i)}, \qquad P(C_i = 1 \mid X_i) = \frac{\exp(\gamma_C X_i)}{1 + \exp(\gamma_C X_i)}.$$

Throughout the simulations, I set $\gamma_G = 1$, $\gamma_C = 2$, and draw the covariate from $X_i \sim \text{Uniform}(-0.5, 0.5)$. The total sample size is $1,000$.

Before introducing the specific data-generating processes, it is useful to describe a broad class of untreated potential-outcome models that satisfy the conditional parallel gap assumption. In particular, the model takes the form:

$$\mathbb{E}[Y_{it}(0) \mid X_i = x, G_i = g, C_i = c] = \alpha(x, g, c) + m_t(x, g) + n_t(x, c).$$

This specification allows for a time-invariant component that depends flexibly on (X_i, G_i, C_i), as well as time-varying components that depend on (X_i, G_i) and (X_i, C_i). Under this structure,

$$\mathbb{E}[\Delta Y_i(0) \mid X_i = x, G_i = g, C_i = c] = \big[m_1(x, g) - m_0(x, g)\big] + \big[n_1(x, c) - n_0(x, c)\big].$$

Therefore,

$$\begin{aligned}
&\mathbb{E}[\Delta Y_i(0) \mid X_i = x, G_i = 1, C_i = 1] - \mathbb{E}[\Delta Y_i(0) \mid X_i = x, G_i = 0, C_i = 1] \\
&= \big[m_1(x,1) - m_1(x,0)\big] - \big[m_0(x,1) - m_0(x,0)\big],
\end{aligned}$$

and similarly,

$$\begin{aligned}
&\mathbb{E}[\Delta Y_i(0) \mid X_i = x, G_i = 1, C_i = 0] - \mathbb{E}[\Delta Y_i(0) \mid X_i = x, G_i = 0, C_i = 0] \\
&= \big[m_1(x,1) - m_1(x,0)\big] - \big[m_0(x,1) - m_0(x,0)\big].
\end{aligned}$$

Hence the conditional parallel gap assumption holds. I consider two simulation designs that satisfy this condition.

In the first design, the untreated potential outcomes are given by

$$Y_{it}(0) = \alpha_i + 5X_i * t\,\mathbb{1}(G_i = 1) + 5X_i * t\,\mathbb{1}(C_i = 1) + u_{it},$$

where $\alpha_i \sim \mathcal{N}(0,1)$ and $u_{it} \sim \mathcal{N}(0,1)$. I assume that (α_i, u_{it}) are mutually independent and independent of (X_i, G_i, C_i). The treated potential outcomes are

$$Y_{it}(1) = 1 + \alpha_i + 5X_i * t\,\mathbb{1}(G_i = 1) + 5X_i * t\,\mathbb{1}(C_i = 1) + v_{it}, \qquad v_{it} \sim \mathcal{N}(0,1).$$

This design implies a homogeneous treatment effect, and the ATT equals 1.

In the second design, the untreated potential outcomes are

$$Y_{it}(0) = \alpha_i - 10X_i * t\,\mathbb{1}(G_i = 1) - 10X_i * t\,\mathbb{1}(C_i = 1) + u_{it},$$

where again $\alpha_i \sim \mathcal{N}(0,1)$, $u_{it} \sim \mathcal{N}(0,1)$, and (α_i, u_{it}) are mutually independent and independent of (X_i, G_i, C_i). The treated potential outcomes are

$$Y_{it}(1) = 1 + \alpha_i + 10X_i * t\,\mathbb{1}(G_i = 1) + 10X_i * t\,\mathbb{1}(C_i = 1) + v_{it}, \qquad v_{it} \sim \mathcal{N}(0,1).$$

This design generates a heterogeneous treatment effect, and the ATT is approximately 5.64.

For each design, I compare three estimators: the matched DDD estimators induced by the two matching workflows commonly used in applied work and the workflow proposed in this paper. For the first two estimators, I report the Monte Carlo bias, the Monte Carlo standard deviation, the average unit-level cluster-robust standard error from the second-stage regression that ignores first-stage matching uncertainty, and the empirical coverage rate of the corresponding 95% confidence interval. For the proposed estimator, I report the same quantities and additionally report the average analytical standard error based on the variance formula in Section 5, together with the empirical coverage rate of the corresponding 95% confidence interval.

Table 1: Design 1: Homogeneous Treatment Effect

	(1)	(2)	(3)
Bias	0.781	0.779	0.001
MC se	0.317	0.270	0.290
Naive CR se	0.334	0.334	0.383
Coverage (naive)	34.9	31.8	99.0
CR se	-	-	0.285
Coverage	-	-	94.7

Notes: Column (1), (2) correspond to the estimators induced by matching workflow 1 and 2 in section 3. Column (3) corresponds to the estimator induced by the proposed matching workflow in section 4. “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.

Table 2: Design 2: Heterogeneous Treatment Effect

	(1)	(2)	(3)
Bias	-1.56	-1.56	-0.002
MC se	0.571	0.577	0.720
Naive CR se	0.523	0.523	0.587
Coverage (naive)	17.4	17.5	88.1
CR se	-	-	0.725
Coverage	-	-	95.0

Notes: Column (1), (2) correspond to the estimators induced by matching workflow 1 and 2 in section 3. Column (3) corresponds to the estimator induced by the proposed matching workflow in section 4. “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.

The results are reported in Tables 1 and 2. Columns (1) and (2) correspond to the estimators induced by matching workflows 1 and 2 in Section 3, whereas column (3) corresponds to the estimator induced by the proposed matching workflow in Section 4. In both designs, the estimators produced by workflows 1 and 2 exhibit substantial bias. In Design 1, the bias is approximately 0.78 for both estimators, which is about 78% of the true ATT. In Design 2, the bias is approximately -1.56 for both estimators, which is about 28% of the true ATT. By contrast, the proposed matched DDD estimator in column (3) has negligible bias in both designs, which is consistent with theorem 2.

The simulation results also show that inference based on the naive cluster-robust standard error from the second-stage regression, which ignores the matching step, is generally invalid. Focusing on the proposed estimator in column (3), in Design 1 with a homogeneous treatment effect, the Monte Carlo standard error is 0.290, whereas the average naive cluster-robust standard error is 0.383. Thus, the naive standard error overestimates the true standard error and yields a naive 95% confidence interval with a coverage rate of 99.0%. In contrast, the average value of the proposed standard error estimate is 0.285, which is very close to the Monte Carlo benchmark, and the resulting coverage rate is 94.7%.

In Design 2 with a heterogeneous treatment effect, the pattern reverses. The Monte Carlo standard error is 0.720, while the average naive cluster-robust standard error is only 0.587, leading to an underestimation of about 18.5% of the true standard error. As a result, the naive 95% confidence interval only has a coverage rate of 88.1%. By comparison, the proposed variance estimator again performs well: its average value is 0.725 and the corresponding coverage rate is 95.0%. This pattern is consistent with the findings in Liu and Vazquez-Bare (2026) for matched DiD designs, where ignoring the matching step can either overestimate or underestimate the true variance, which depends on the data generating process. The simulation evidence here suggests that the same issue may carry over to the matched triple-differences setting.

6 Empirical Application

In this section, I revisit Cai (2016) by applying the method proposed in Sections 4 and 5. Cai (2016) uses household-level panel data from 12 tobacco-producing counties in Jiangxi Province, China, to study how the introduction of an agricultural insurance program affects rural household borrowing and saving behavior.

Among these counties, some households primarily rely on tobacco cultivation as their main source of income (hereafter referred to as tobacco households). Tobacco production is particularly exposed to weather shocks and natural disasters because tobacco has a longer growing season, about six months, than other local crops such as rice, which has a growing season of about three months. In 2003, one of the 12 counties, Guangchang County, collaborated with the People's Insurance Company of China (PICC) to design and launch a tobacco production insurance program to mitigate these risks. If an insurance-covered weather disaster causes at least a 30 percent reduction in a household's tobacco yield, the household is eligible for an insurance payout. Among the 12 counties, only tobacco households in Guangchang County were offered the tobacco insurance program. Since purchase was compulsory, all tobacco households in Guangchang County were covered. In contrast, non-tobacco households in Guangchang County and households in the other counties were not eligible to purchase the insurance.

The policy rollout fits the econometric framework developed above. A household is treated only if it satisfies two conditions: it is located in the treated county and is a tobacco household. Cai (2016) notes that tobacco and non-tobacco households differ in borrowing and saving behavior across treated and untreated counties, while appearing to share similar region-specific trends. This setting motivates the use of a triple-differences (DDD) design as her main empirical strategy. The DDD design differences out the common region-specific trends and isolates the effect of the insurance program on tobacco-growing households in the treated county.

Cai (2016) uses a household-level panel dataset containing 5,746 households from the 12 tobacco-producing counties. The panel covers the years 2000 to 2008. The compulsory insurance policy was implemented in 2003, which yields three pre-treatment years, 2000–2002, and six post-treatment years, 2003–2008. The dataset combines administrative financial and savings records from Rural Credit Cooperative (a main rural bank) with information on household demographics, land allocations, and crop revenues. It includes more than 3,000 tobacco-growing households, approximately 1,200 of which are located in the exposed county, and about 2,200 non-tobacco households. Cai (2016) examines outcomes such as tobacco production area, loan size, loan limits, interest rates, savings rates, and flexible-term savings.

In this draft, I focus on the effect of insurance provision on rural household saving behavior. For each household, I first average the outcome separately over the pre-treatment and post-treatment periods, thereby transforming the original household-level panel into a two-period setting with four cross-sectional subgroups. I then analyze the outcomes of interest using a regression specification similar to that in Cai (2016), where she uses a parametric approach by directly including covariates into the triple-differences regression[2]:

$$\begin{aligned}
Y_{it} &= \beta_0 + \beta_1 Tobacco_i + \beta_2 Treat_i + \beta_3 Post_t + \beta_4 Tobacco_i \times Treat_i + \beta_5 Tobacco_i \times Post_t \\
&\quad + \beta_6 Treat_i \times Post_t + \beta_7 Tobacco_i \times Treat_i \times Post_t + X_i'\theta + u_{it}.
\end{aligned}$$

The outcomes Y_{it} include net saving, defined as the annual increase in total savings; the

[2]Directly including time-invariant covariates in this regression does not create a collinearity problem, because the included dummy variables and their interactions saturate only the cells defined by the four cross-sectional subgroups and the two time periods; they do not absorb all unit-level cross-sectional variation.

saving rate, defined as the ratio of net saving to current household income; and flexible-term saving, defined as the ratio of net savings in checking accounts to total net savings. The variable $Tobacco_i$ is an indicator for whether household i is a tobacco-growing household. The variable $Treat_i$ is an indicator for whether household i is located in the exposed county. The variable $Post_t$ indicates whether the observation belongs to the post-treatment period. Finally, X_i denotes a vector of time-invariant household characteristics, including the age and education level of the household head and household size.

I then implement the matched DDD estimator proposed in Section 4 using the same set of covariates X_i. I use nearest-neighbor matching with one neighbor. Because the covariates are discrete, ties arise in the matching procedure. When ties occur, I use the average outcome among all tied matched controls rather than randomly selecting one control unit. This choice makes the estimator invariant to arbitrary random tie-breaking. For the matched DDD estimator, I report two standard error estimates: the naive cluster-robust standard error obtained after matching, which does not account for the first-step matching procedure, and the valid standard error based on the variance estimator proposed in Theorem 3.

Across Columns (1)–(3) of Table 3, the matched DDD estimator delivers point estimates with larger absolute magnitudes than the regression specification that directly adjusts for covariates. For the saving rate, both approaches suggest that insurance provision reduces household saving. The regression estimate is statistically significant at the 10 percent level, whereas the matched DDD estimate is slightly larger in magnitude but falls just outside the 10% significance level. For net saving, the regression approach yields an estimate very close to zero, while the matched DDD estimator suggests a negative effect of insurance provision. However, both estimates are not statistically significant. For flexible-term saving, both approaches indicate a positive and statistically significant effect of insurance provision, with the matched DDD estimate again larger in magnitude than the regression estimate.

Comparing the two matched DDD rows (rows 2 and 3) in Table 3, the point estimates are identical by construction, but the inference differs because the naive standard error does not account for the matching step. In this application, the valid standard error is slightly smaller than the naive standard error for net saving and flexible-term saving, and nearly identical for the saving rate. As a result, the valid confidence intervals are slightly narrower in Columns (2) and (3).

Table 3: Estimated Effects of Insurance Provision on Saving

	(1) Saving Rate	(2) Net Saving	(3) Flexible Saving
Regression			
Coefficient	-0.0177	0.000	0.086
Std. err.	0.010	0.018	0.011
p-value	0.077	0.985	<0.001
95% CI	[-0.037 ; 0.002]	[-0.035 ; 0.034]	[0.064 ; 0.107]
Observations	7,912	7,934	7,264
Naive matched DDD			
Coefficient	-0.022	-0.029	0.106
Std. err.	0.014	0.024	0.016
p-value	0.100	0.214	<0.001
95% CI	[-0.049 ; 0.004]	[-0.076 ; 0.017]	[0.076 ; 0.137]
Observations	7,912	7,934	7,264
Matched DDD			
Coefficient	-0.022	-0.029	0.106
Std. err.	0.014	0.022	0.014
p-value	0.107	0.192	<0.001
95% CI	[-0.049 ; 0.005]	[-0.073 ; 0.015]	[0.079 ; 0.133]
Observations	7,912	7,934	7,264

Notes: This table reports estimates of the effect of insurance provision on saving outcomes. "Regression" reports estimates from a DDD regression that directly includes covariates. "Naive matched DDD" reports the matched DDD point estimates with standard errors that do not account for the first-step matching procedure. "Matched DDD" reports the matched DDD estimates with valid standard errors that account for the matching step. All standard errors are clustered at the household level.

7 Conclusion

This paper studies how to combine matching methods with triple-differences designs. I first analyze two workflows commonly used in applied studies: (i) conducting a single pairwise matching step between a pooled treated group and a pooled untreated group, and (ii) conducting pairwise matching within two separate DiD comparisons. I show that, even in the baseline setting with two periods and four cross-sectional subgroups, both workflows generally produce estimators that converge to the target estimand plus an asymptotic bias term. The source of this bias is a mismatch between the covariate distribution required by the identification result and the covariate distributions induced by the matching workflows. Under the conditional parallel gap assumption, identification requires each of the three untreated subgroups to be balanced with respect to the covariate distribution of the eligible exposed group. The two empirical workflows fail to satisfy this requirement because they generally induce matched untreated groups whose covariate distribution differs from that of the eligible exposed group.

To address this limitation, I propose an alternative matching workflow and a matched DDD estimator that is consistent for the target estimand. I establish the estimator's asymptotic normality and derive a closed-form representation of its asymptotic variance. Although the proposed estimator can be written as a linear combination of three pairwise matched DiD estimators, the variance formula in Abadie and Imbens (2006) cannot be applied directly because of the additional dependence structure between those three pairwise matched DiD estimators. I therefore propose a consistent variance estimator tailored to the matched triple-differences setting that accounts for the first-step matching uncertainty. Simulation evidence and a reanalysis of the data from Cai (2016) illustrate the validity and practical relevance of the proposed methods.

Recent literature on triple-differences designs has proposed several approaches for incorporating covariates when the parallel gap assumption holds conditionally. Matching-based methods are also useful in this setting because they are transparent, intuitive, easy to implement, and fully nonparametric. The results in this paper provide empirical researchers with a principled way to combine matching methods with triple-differences designs and to conduct valid inference for the parameter of interest.

References

A. Abadie and G. W. Imbens. Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74(1):235–267, 2006.

A. Abadie and G. W. Imbens. Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics*, 29(1):1–11, 2011.

C. Caetano and B. Callaway. Difference-in-differences when parallel trends holds conditional on covariates. *Working Paper*, 2024.

J. Cai. Insurance impact on production and financial decisions. *American Economic Journal: Economic Policy*, 8(2):44–88, 2016.

A. Chattopadhyay and J. R Zubizarreta. On the implied weights of linear regression for causal inference. *Biometrika*, 110(3):615–629, 2023.

C. de Chaisemartin and X. D'Haultfœuille. Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey. *The Econometrics Journal*, 06 2022.

J. Gruber. The incidence of mandated maternity benefits. *American Economic Review*, 84 (3):622–641, 1994.

J. J. Heckman, H. Ichimura, and P. E. Todd. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, 64(4):605–654, 10 1997.

J. J. Heckman, H. Ichimura, A. Smith, Jeffrey, and T. Petra. Characterizing selection bias using experimental data. *Econometrica*, 66(5):1017–1098, 1998.

K. Imai, I. S. Kim, and E. H. Wang. Matching methods for causal inference with time-series cross-sectional data. *American Journal of Political Science*, 67(3):587–605, 2023.

Z. Lin and F. Han. On regression-adjusted imputation estimators of average treatment effects. *Journal of Econometrics*, 251:106080, 2025.

Z. Lin, P. Ding, and F. Han. Estimation based on nearest neighbor matching: From density ratio to average treatment effect. *Econometrica*, 91(6):2187–2217, 2023.

Y. Liu and G. Vazquez-Bare. Post-matching two-way fixed effects estimation. *Working Paper*, 2026.

A. Olden and J. Møen. The triple difference estimator. *The Econometrics Journal*, 25(3): 531–553, 2022.

M. Ortiz-Villavicencio and P. H. Sant'Anna. Better understanding triple differences estimators. *arXiv*, pages 1–60, 2025.

J. Roth, P. H. Sant'Anna, A. Bilinski, and J. Poe. What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244, 2023.

S. Viel, L. Truquet, and I. Yamane. Convergence rate for nearest neighbour matching: geometry of the domain and higher-order regularity. *Working Paper*, 2025.

Appendix

A Examples of Matching Procedure

A.1 Nearest Neighbor Matching with Fixed Neighbors

In nearest-neighbor matching with fixed neighbors (Abadie and Imbens, 2006), each treated unit is matched to its nearest M (fixed) neighbors in the untreated subgroup. For each $i \in \mathcal{I}_{11}$, the matched set $\mathcal{M}_{gs}(i)$ contains the M comparison units whose covariate values are closest to X_i. Writing it formally, $\mathcal{M}_{gs}(i) = \{j_1(i,g,s), \ldots, j_M(i,g,s)\}$, where $j_m(i,g,s)$ is the index of the m-th nearest neighbor of unit i among the untreated group (g,s)

$$\sum_{l:(G_l,S_l)=(g,s)} \mathbb{1}\{\|X_l - X_i\| \leq \|X_j - X_i\|\} = m.$$

The corresponding pairwise weight w_{ij}^{fix} equals $\frac{1}{M}$ if $j \in \mathcal{M}_{gs}(i)$ and equals 0 if $j \notin \mathcal{M}_{gs}(i)$:

$$w_{ij}^{fix} = \frac{1}{M} \mathbb{1}\{j \in \mathcal{M}_{gs}(i)\}, \qquad j \in \mathcal{I}_{gs}.$$

The matching weight of untreated unit j is therefore

$$w_j^{fix} = \frac{1}{M} \sum_{i \in \mathcal{I}_{11}} \mathbb{1}\{j \in \mathcal{M}_{gs}(i)\},$$

that is, the number of times unit j is used as a match, divided by M.

A.2 Nearest Neighbor Matching with Diverging Neighbors

Nearest neighbor matching with diverging neighbors (Lin et al. (2023)) has the same matching algorithm and the corresponding pairwise weight $w_{ij}^{diverge}$ and matching weight $w_j^{diverge}$ as the fixed neighbors. The only differences is that the number of the neighbors M is diverging with the untreated group sample size n_{gs}. To be specific M satisfies the the following assumption.

Assumption 7. As $n \to \infty$,

$$\frac{Mlog(n_{gs})}{n_{gs}} \to 0, \quad M \to \infty$$

A.3 Kernel Matching

Kernel matching (Heckman et al. (1997), Heckman et al. (1998)) each reference unit is matched to a kernel-weighted combination of comparison units, with the weights determined by the proximity of their covariate values. Writing formally, let $K : \mathbb{R}^q \to [0, \infty)$ be a multivariate kernel function and let $h > 0$ be the bandwidth that depends on the the sample size of the untreated group n_{gs}. The pairwise weight correspond to kernel matching is:

$$w_{ij}^{kernel} = \frac{K\left(\frac{X_j - X_i}{h}\right)}{\sum_{l \in \mathcal{I}_{gs}} K\left(\frac{X_l - X_i}{h}\right)}, \qquad j \in \mathcal{I}_{gs}.$$

For each $i \in \mathcal{I}_{11}$, the matched set consists of the comparison units that receive positive kernel weight:

$$\mathcal{M}_{gs}(i) = \left\{ j \in \mathcal{I}_{gs} : K\left(\frac{X_j - X_i}{h}\right) > 0 \right\}.$$

The matching weight of comparison unit j is therefore

$$w_j^{kernel} = \sum_{i \in \mathcal{I}_{11}} \frac{K\left(\frac{X_j - X_i}{h}\right)}{\sum_{l \in \mathcal{I}_{gs}} K\left(\frac{X_l - X_i}{h}\right)},$$

We impose the following assumption on the Kernel function and the bandwidth, which are common in kernel matching literature:

Assumption 8. (i) The kernel function has a compact support. There exists $0 < r < R < \infty$ such that the kernel function K satisfies: $0 \le K(u) \le \overline{K}$, $\underline{K} \le K(u) \le \overline{K}$, if $\|u\| \le r$, and $K(u) = 0$ if $\|u\| > R$, where $\underline{K}$ and $\overline{K}$ are positive real numbers.
(ii) As $n \to \infty$,

$$h \to 0, \quad n_{gs} h^q \to \infty$$

Lemma 3. Under assumption 6(i), we have the following results:
(a) w_{ij}^{fix} and w_j^{fix} satisfy assumption 5.
(b) If assumption 7 holds, then $w_{ij}^{diverge}$ and $w_j^{diverge}$ satisfies assumption 5.
(c) If assumption 8 holds, then w_{ij}^{kernel} and w_j^{kernel} satisfy assumption 5.

B Necessity of Condition (iv) in Assumption 5

In this section, we construct a matching procedure that satisfies conditions (i)–(iii) of Assumption 5 but violates condition (iv), and we provide simulation evidence that the resulting matched triple-differences estimator is not asymptotically normal unconditionally.

Consider the following *switch matching rule*. Let $T_n = \frac{\overline{X}_{11}}{\sqrt{\widehat{V}(\overline{X}_{11})/n_{11}}}$ denote the standardized sample mean of X in the treated cell. If $|T_n| < z_{0.75}$, we construct the matched triple-differences estimator using one-nearest-neighbor matching, following the procedure in Section 4; if $|T_n| \geq z_{0.75}$, we use fifty-nearest-neighbor matching instead. Here $z_{0.75}$ denotes the 75th percentile of the standard normal distribution. Because T_n is asymptotically standard normal, each of the two matching procedures is used in roughly half of all repeated samples.

By construction, the switch rule satisfies condition (i). The rule switches between two nearest-neighbor procedures, each of which satisfies conditions (ii) and (iii) by Lemma 3; the law of iterated expectations then implies that the switch rule satisfies conditions (ii) and (iii) as well. Condition (iv), however, is violated: one can show that, for some constant $C > 0$,

$$\mathbb{E}\left[\left(\sum_{j \in \mathcal{I}_{gs} \setminus \{1\}} \left| w_j - w_j^{(1)} \right|\right)^4\right] \geq C n^{7/2}.$$

Next, we provide simulation evidence that the centered and normalized matched triple-differences estimator $(V_1 + V_2)^{-1/2} \sqrt{n_{11}}\,(\widehat{\tau}_l - \tau_l - B)$ constructed under the switch rule is not asymptotically normal. The data-generating process follows Design 2 in Section 5, with two modifications. First, the logit models for G and S both have an intercept of -1.5 and slopes $\gamma_G = -\gamma_S = 1$, so that the treated cell contains a smaller share of observations:

$$P(G_i = 1 \mid X_i) = \frac{\exp(-1.5 + \gamma_G X_i)}{1 + \exp(-1.5 + \gamma_G X_i)}, \qquad P(S_i = 1 \mid X_i) = \frac{\exp(-2 + \gamma_S X_i)}{1 + \exp(-2 + \gamma_S X_i)}.$$

Second, the coefficient on the covariate-specific trend is reduced from 10 to 2, so that effect heterogeneity is smaller:

$$\begin{aligned} Y_{it}(0) &= \alpha_i - 2X_i t\, \mathbb{1}(G_i = 1) - 2X_i t\, \mathbb{1}(S_i = 1) + u_{it}, \\ Y_{it}(1) &= \alpha_i + 2X_i t\, \mathbb{1}(G_i = 1) + 2X_i t\, \mathbb{1}(S_i = 1) + v_{it}. \end{aligned}$$

The simulation uses a sample size of 10,000 and 10,000 Monte Carlo replications. We plot the empirical density of the centered and normalized estimator against the standard normal density and report the Kolmogorov–Smirnov test for normality.

Figure 4 shows a sizable discrepancy between the empirical density of the centered and normalized matched triple-differences estimator and the standard normal density. The Kolmogorov–Smirnov statistic is 0.0424, with a p-value below 0.001. Together, these results provide strong evidence that, under the switch matching rule, the matched triple-differences

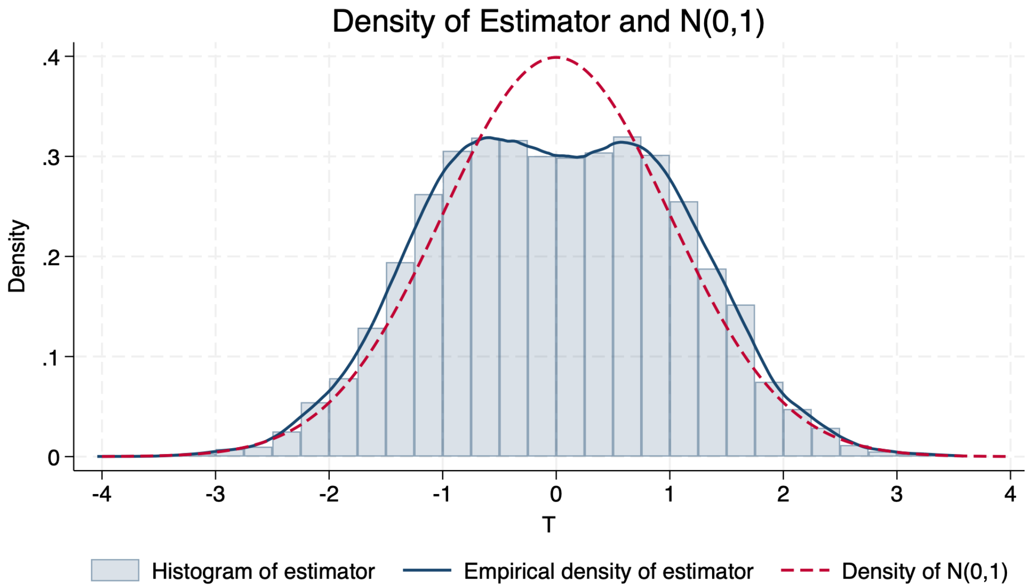

Figure 4: Density of the centered and normalized estimator under the switch rule versus $N(0,1)$

estimator is not asymptotically normal unconditionally.

C Bias Correction

In this section, we present the bias-correction procedure and establish the asymptotic theory for the bias-corrected matched triple-differences estimator.

Recall that the bias B of the matched triple-differences estimator is

$$B = \frac{1}{n_{11}} \sum_{i=1}^{n} \left[\mathbb{1}(i \in \mathcal{I}_{11}) - \mathbb{1}(i \in \mathcal{I}_{01}) w_i \right] \mu_l^{01}(X_i) + \frac{1}{n_{11}} \sum_{i=1}^{n} \left[\mathbb{1}(i \in \mathcal{I}_{11}) - \mathbb{1}(i \in \mathcal{I}_{10}) w_i \right] \mu_l^{10}(X_i)$$
$$- \frac{1}{n_{11}} \sum_{i=1}^{n} \left[\mathbb{1}(i \in \mathcal{I}_{11}) - \mathbb{1}(i \in \mathcal{I}_{00}) w_i \right] \mu_l^{00}(X_i).$$

To conduct the bias correction, we estimate the three unknown conditional mean functions μ_l^{gs}, $(g,s) \in \{(0,1),(1,0),(0,0)\}$, non-parametrically by $\widehat{\mu}_l^{gs}$, using off-the-shelf methods such as series or spline estimators. The bias-corrected matched triple-differences estimator then takes the form $\widehat{\tau}_l - \widehat{B}$, where

$$\widehat{B} = \frac{1}{n_{11}} \sum_{i=1}^{n} \left[\mathbb{1}(i \in \mathcal{I}_{11}) - \mathbb{1}(i \in \mathcal{I}_{01}) w_i\right] \widehat{\mu}_l^{01}(X_i) + \frac{1}{n_{11}} \sum_{i=1}^{n} \left[\mathbb{1}(i \in \mathcal{I}_{11}) - \mathbb{1}(i \in \mathcal{I}_{10}) w_i\right] \widehat{\mu}_l^{10}(X_i)$$
$$- \frac{1}{n_{11}} \sum_{i=1}^{n} \left[\mathbb{1}(i \in \mathcal{I}_{11}) - \mathbb{1}(i \in \mathcal{I}_{00}) w_i\right] \widehat{\mu}_l^{00}(X_i).$$

Assumption 9 (Bias correction)**.** (i) The support of X is a Cartesian product of compact intervals.

(ii) For every $(g, s) \in \{(0,1),(1,0),(0,0)\}$ and every event time l, μ_l^{gs} is p times continuously differentiable.

(iii) For every $(g, s) \in \{(0,1),(1,0),(0,0)\}$ and every event time l, the estimator $\widehat{\mu}_l^{gs}$ satisfies

$$\max_{m \in \Lambda_p} \|\partial^m \widehat{\mu}_l^{gs}\|_\infty = O_p(1),$$
$$\max_{m \in \Lambda_r} \|\partial^m \widehat{\mu}_l^{gs} - \partial^m \mu_l^{gs}\|_\infty = O_p(n^{-\gamma_r}) \quad \text{for every } r \in \{1, \ldots, p-1\},$$

where, for $r \in \{1, \ldots, p\}$, Λ_r denotes the set of all q-dimensional vectors of non-negative integers $m = (m_1, \ldots, m_q)$ with $\sum_{i=1}^q m_i = r$, and $\gamma_1, \ldots, \gamma_{p-1}$ are positive constants.

(iv) For every $(g, s) \in \{(0,1),(1,0),(0,0)\}$, the pairwise weights in addition satisfy

$$\mathbb{E}\left[\sum_{j \in \mathcal{I}_{gs}} w_{ij} \|X_j - X_i\|^p \,\middle|\, i \in \mathcal{I}_{11}\right] = o(n^{-1/2}),$$
$$\mathbb{E}\left[\sum_{j \in \mathcal{I}_{gs}} w_{ij} \|X_j - X_i\|^r \,\middle|\, i \in \mathcal{I}_{11}\right] = o(n^{-1/2+\gamma_r}) \quad \text{for every } r \in \{1, \ldots, p-1\}.$$

Assumption 9 is similar to Assumption 3.5′ of Lin and Han (2025). Condition (iii) can be verified for standard nonparametric regression methods, such as series and spline estimators, provided the conditional means are smooth as required by condition (ii). Condition (iv) is a locality requirement, similar to condition (ii) of Assumption 5, but it requires the weighted moments of the covariate discrepancy to shrink fast enough for the bias-corrected estimator to attain $\sqrt{n}$-consistency.

Theorem 4. Under Assumptions 1, 2, 3, 4, 5, 6, and 9,

$$\sqrt{n_{11}}(\widehat{B} - B) \xrightarrow{p} 0, \quad (V_1 + V_2)^{-1/2} \sqrt{n_{11}} \left(\widehat{\tau}_l - \tau_l - \widehat{B}\right) \xrightarrow{d} \mathcal{N}(0,1).$$

Theorem 4 shows that the bias can be estimated with an error of order smaller than $n_{11}^{-1/2}$. Consequently, the bias-corrected matched triple-differences estimator is $\sqrt{n_{11}}$-consistent and asymptotically normal, and the correction leaves the asymptotic variance unchanged, so the variance estimator in section 4 still applies.

D Auxiliary Lemma

Lemma 4. The induced weights by regression specification (2) satisfy,

$$\mathbb{E}\left[\frac{(1-p_{11})L(D|G,S,X)}{\mathbb{E}[(D-L(D|G,S,X))^2]}X\,\middle|\,D=0\right]=\mathbb{E}\left[\frac{p_{01}L(D|G,S,X)}{\mathbb{E}[(D-L(D|G,S,X))^2]}X\,\middle|\,G=0,S=1\right]$$
$$+\,\mathbb{E}\left[\frac{p_{10}L(D|G,S,X)}{\mathbb{E}[(D-L(D|G,S,X))^2]}X\,\middle|\,G=1,S=0\right]+\mathbb{E}\left[\frac{p_{00}L(D|G,S,X)}{\mathbb{E}[(D-L(D|G,S,X))^2]}X\,\middle|\,G=0,S=0\right]$$
$$\mathbb{E}\left[\frac{p_{11}\left(1-L(D|G,S,X)\right)}{\mathbb{E}[(D-L(D|G,S,X))^2]}X\,\middle|\,G=1,S=1\right]=\mathbb{E}\left[\frac{(1-p_{11})L(D|G,S,X)}{\mathbb{E}[(D-L(D|G,S,X))^2]}X\,\middle|\,D=0\right]$$

Lemma 5. Consider a subset of the data with two subgroups, the treated subgroup $(1,1)$ and one of the comparison group (g,s). And we conduct the pairwise matching procedure described in section 4 satisfying assumption 5. Let $(Z_i)_{i=1}^n$ be a sequence of i.i.d random variables with $\mathbb{E}[Z_i|G_i=g,S_i=s]=\mu_Z(g,s)$, $\mathbb{E}[Z_i|G_i=g,S_i=s,X_i=x]=\mu_Z(g,s,x)$ and $\mathbb{V}[Z_i|G_i=g,S_i=s,X_i=x]=\sigma_Z^2(g,s,x)$ with both functions being Lipschitz-continuous and with $\mathbb{E}[Z_i^4|G_i=g,S_i=s,X_i=x]$ uniformly bounded over x. Then

$$T_n=\frac{1}{N_{11}}\sum_{i\in\mathcal{I}_{11}}Z_i-\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{gs}}w_iZ_i\xrightarrow{p}\mu_Z(1,1)-\mathbb{E}[\mu_Z(g,s,X)|G=1,S=1]$$

Lemma 6. Under assumption 1, 5, and 6

$$\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)-\mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right]=o_p(1)$$
$$\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)-\mathbb{E}\left[\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right]=o_p(1)$$

and for $(g,s) \in \{(0,1),(1,0),(0,0)\}$

$$\frac{1}{n}\sum_{i\in\mathcal{I}_{gs}} w_i^2\sigma_{gs,l}^2(X_i) - \mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{gs}} w_i^2\sigma_{gs,l}^2(X_i)\right] = o_p(1)$$
$$\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{gs}} w_i^2\sigma_{gs,l}^2(X_i) - \mathbb{E}\left[\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{gs}} w_i^2\sigma_{gs,l}^2(X_i)\right] = o_p(1)$$

Lemma 7. Under assumption 5, for any $(g,s) \in \{(0,1),(1,0),(0,0)\}$ the pairwise weight w_{ij}, $w_{ij'}$ and the the matching weight w_j, $w_{j'}$ induced by the pairwise matching between the treated subgroup $(1,1)$, and the untreated subgroup (g,s) satisfies:

(i)

$$\mathbb{E}\left[\frac{1}{n^2}\sum_{j\in\mathcal{I}_{gs}} w_j^2\right] = O(n^{-1}), \quad \frac{1}{n^2}\sum_{j\in\mathcal{I}_{gs}} w_j^2 = O_p(n^{-1})$$
$$\sum_{i\in\mathcal{I}_{11}} w_{ij}w_{ij'} \le \max_{j\in\mathcal{I}_{gs}} w_j, \quad \max_{j\in\mathcal{I}_{gs}} w_j \le \left(\sum_{j\in\mathcal{I}_{gs}} w_j^2\right) = O_p(n^{1/2})$$

(ii)
$$\mathbb{E}\left[\left(\sum_{j\in\mathcal{I}_{gs\setminus\{1\}}}\left|w_j^2 - \left(w_j^{(1)}\right)^2\right|\right)^2\right] = O(n^{1/2})$$

(iii) For $r < 4+\delta$, we have:

$$\mathbb{E}[w_j^r | j \in \mathcal{I}_{gs}] = O(1), \quad \frac{1}{n_{gs}}\sum_{i\in\mathcal{I}_{gs}} w_j^r = O(1)$$

(iv)

$$\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left\|X_i - X_j\right\| = o_p(1), \quad \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left\|X_i - X_j\right\|\right)^2 = o_p(1)$$

Lemma 8. For any $(g,s) \in \{0,1\}^2$, under assumption 1, we have:

$$\mathbb{E}\left[\left(\frac{n}{n_{gs}} - \frac{1}{p_{gs}}\right)^2\right] = O(n^{-1})$$

Lemma 9. Let T_n be a sequence of random variables such that $\mathbb{E}[T_n|\mathbf{G},\mathbf{S},\mathbf{X}] = 0$ and $Var(T_n|\mathbf{G},\mathbf{S},\mathbf{X}) \xrightarrow{p} 0$, then $T_n \xrightarrow{p} 0$.

Lemma 10. For the nearest neighbor matching with fixed neighbor, nearest neighbor matching with diverging neighbor, and the kernel matching introduced in appendix A. The matching weight w_j generated by those three procedures all satisfies:

$$\sum_{j\in\mathcal{I}_{gs}\setminus\{1\}} \left|w_j - w_j^{(1)}\right| \leq w_1 + w_1^{(1)} + \mathbb{1}(1 \in \mathcal{I}_{11}) + \mathbb{1}(1' \in \mathcal{I}_{11})$$

This implies that the nearest neighbor matching with fixed neighbor, nearest neighbor matching with diverging neighbor, and the kernel matching all satisfies assumption 5, which is:

$$\mathbb{E}\left[\left(\sum_{j\in\mathcal{I}_{gs}\setminus\{1\}} \left|w_j - w_j^{(1)}\right|\right)^4\right] = O(1)$$

Lemma 11. If we use nearest-neighbor matching to match each unit in subgroup (g,s) to units within the same subgroup, then for each fixed m, the sample maximum of the norm of the matching discrepancy for the mth nearest neighbor is $o_p(1)$. Specifically, let $k_m(i)$ denote the index of the mth closest match within the same subgroup for unit i in subgroup (g,s). Then

$$\max_{i\in\mathcal{I}_{gs}} \left\|X_i - X_{k_m(i)}\right\| = o_p(1).$$

Lemma 12. Define $\epsilon_{il}^{gs} = \Delta Y_{il} - \mathbb{E}[\Delta Y_{il}|G_i = g, S_i = s, X_i] = \Delta Y_{il} - \mu_l^{gs}(X_i)$, then for any $(g,s) \in \{0,1\}^2$, under assumption 6, there exists some constant $C > 0$, such that

$$\mathbb{E}\left[(\epsilon_{il}^{gs})^4\middle|\mathbf{G},\mathbf{S},\mathbf{X}, G_i = g, S_i = s\right] < C$$

Lemma 13. The estimator $\widehat{\tau}_l$ can be written as the following two equivalent ways:

$$\begin{aligned}\widehat{\tau}_l &= \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\Delta Y_{il} - \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{01}} w_i\Delta Y_{il} - \left(\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{10}} w_i\Delta Y_{il} - \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{00}} w_i\Delta Y_{il}\right)\\ &= \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\Delta Y_{il} - \widehat{\Delta Y_{il}^{01}} - \widehat{\Delta Y_{il}^{10}} + \widehat{\Delta Y_{il}^{00}}\right),\end{aligned}$$

where, $\widehat{\Delta Y_{il}^{gs}} = \sum_{j\in\mathcal{I}_{gs}} w_{ij}\Delta Y_{jl}$

E Proof of Theorems

Proof. Proof of Theorem 1

Notice that: $Y_t(0) - Y_{t'}(0) = \sum_{g=0}^{t+1-t'} Y_{t-g}(0) - Y_{t-g-1}(0)$, so

$$
\begin{aligned}
&\mathbb{E}[Y_t(0) - Y_{t'}(0)|G = 1, S = 1, X] - \mathbb{E}[Y_t(0) - Y_{t'}(0)|G = 0, S = 1, X]\\
&- [\mathbb{E}[Y_t(0) - Y_{t'}(0)|G = 1, S = 0, X] - \mathbb{E}[Y_t(0) - Y_{t'}(0)|G = 0, S = 0, X]\\
&= \sum_{g=0}^{t+1-t'} (\mathbb{E}[Y_{t-g}(0) - Y_{t-g-1}(0)|G = 1, S = 1, X] - \mathbb{E}[Y_{t-g}(0) - Y_{t-g-1}(0)|G = 0, S = 1, X]\\
&-\mathbb{E}[Y_{t-g}(0) - Y_{t-g-1}(0)|G = 1, S = 0, X] - \mathbb{E}[Y_{t-g}(0) - Y_{t-g-1}(0)|G = 0, S = 0, X])\\
&= 0
\end{aligned}
$$

The last line follows from assumption 4.

$$
\begin{aligned}
\tau_t &= \mathbb{E}[Y_t(1) - Y_t(0)|G = 1, S = 1]\\
&= \mathbb{E}[\mathbb{E}[Y_t(1) - Y_t(0)|G_i = 1, S = 1, X]|G = 1, S = 1]\\
&= \mathbb{E}[\mathbb{E}[Y_t(1) - Y_t(0)|G = 1, S = 1, X]|G = 1, S = 1]\\
&+ \mathbb{E}[\mathbb{E}[Y_t(0) - Y_{t'}(0)|G = 1, S = 1, X] - \mathbb{E}[Y_t(0) - Y_{t'}(0)|G = 0, S = 1, X]|G = 1, S = 1]\\
&- \mathbb{E}[\mathbb{E}[Y_t(0) - Y_{t'}(0)|G = 1, S = 0, X] - \mathbb{E}[Y_t(0) - Y_{t'}(0)|G = 0, S = 0, X]|G = 1, S = 1]\\
&= \mathbb{E}[\mathbb{E}[Y_t(1) - Y_{t'}(0)|G = 1, S = 1, X] - \mathbb{E}[Y_t(0) - Y_{t'}(0)|G = 0, S = 1, X]|G = 1, S = 1]\\
&- \mathbb{E}[\mathbb{E}[Y_t(0) - Y_{t'}(0)|G = 1, S = 0, X] - \mathbb{E}[Y_t(0) - Y_{t'}(0)|G = 0, S = 0, X]|G = 1, S = 1]\\
&= \mathbb{E}[\mathbb{E}[Y_t - Y_{t'}|G = 1, S = 1, X] - \mathbb{E}[Y_t - Y_{t'}|G = 0, S = 1, X]|G = 1, S = 1]\\
&- \mathbb{E}[\mathbb{E}[Y_t - Y_{t'}|G = 1, S = 0, X] - \mathbb{E}[Y_t - Y_{t'}|G = 0, S = 0, X]|G = 1, S = 1]\\
&= \mathbb{E}[Y_t - Y_{t'}|G = 1, S = 1] - \mathbb{E}[\mathbb{E}[Y_t - Y_{t'}|G = 0, C = 1, X]|G = 1, S = 1]\\
&- \mathbb{E}[Y_t - Y_{t'}|G = 1, S = 0, X]|G = 1, S = 1] + \mathbb{E}[\mathbb{E}[Y_t - Y_{t'}|G = 0, S = 0, X_i]|G = 1, S = 1]
\end{aligned}
$$

The first equality follows from law of iterated expectation. The second equality follows from the equation above and law of iterated expectation. The third equality follows by canceling out $\mathbb{E}[Y_t(0)|G = 1, S = 1]$. The fourth inequality follows from the assumption 3. The last equality follows from applying law of iterated expectation to the first term. ■

Proof. Proof of theorem 2.

$$
\begin{aligned}
\widehat{\tau}_l &= \frac{1}{n_{11}} \sum_{i\in\mathcal{I}_{11}} \Delta Y_{il} - \frac{1}{n_{11}} \sum_{i\in\mathcal{I}_{01}} w_i \Delta Y_{il} - \left(\frac{1}{n_{11}} \sum_{i\in\mathcal{I}_{10}} w_i \Delta Y_{il} - \frac{1}{n_{11}} \sum_{i\in\mathcal{I}_{00}} w_i \Delta Y_{il} \right) \\
&= \frac{1}{n_{11}} \sum_{i\in\mathcal{I}_{11}} \Delta Y_{il} - \frac{1}{n_{11}} \sum_{i\in\mathcal{I}_{01}} w_i \Delta Y_{il} + \left(\frac{1}{n_{11}} \sum_{i\in\mathcal{I}_{11}} \Delta Y_{il} - \frac{1}{n_{11}} \sum_{i\in\mathcal{I}_{10}} w_i \Delta Y_{il} \right) \\
&- \left(\frac{1}{n_{11}} \sum_{i\in\mathcal{I}_{11}} \Delta Y_{il} - \frac{1}{n_{11}} \sum_{i\in\mathcal{I}_{00}} w_i \Delta Y_{il} \right) \\
&\xrightarrow{p} \mathbb{E}[\Delta Y_l | G=1, S=1] - \mathbb{E}[\mathbb{E}[\Delta Y_l | G=0, S=1, X] | G=1, S=1] \\
&+ \mathbb{E}[\Delta Y_l | G=1, S=1] - \mathbb{E}[\mathbb{E}[\Delta Y_l | G=1, S=0, X] | G=1, S=1] \\
&- (\mathbb{E}[\Delta Y_l | G=1, S=1] - \mathbb{E}[\mathbb{E}[\Delta Y_l | G=0, S=0, X] | G=1, S=1]) \\
&= \mathbb{E}[\Delta Y_l | G=1, S=1] - \mathbb{E}[\mathbb{E}[\Delta Y_l | G=0, S=1, X] | G=1, S=1] \\
&- \mathbb{E}[\mathbb{E}[\Delta Y_l | G=1, S=0, X] | G=1, S=1] + \mathbb{E}[\mathbb{E}[\Delta Y_l | G=0, S=0, X] | G=1, S=1] \\
&= \tau_l
\end{aligned}
$$

The convergence results follows from lemma 5. The last equality follows from the identification results.

For notation compactness define $\mu_l^{gs}(x) = \mathbb{E}[\Delta Y_l | G=g, S=s, X=x]$, $\epsilon_l^{gs}(X_i) = \Delta Y_{il} - \mu_l^{G_iS_i}(X_i)$, $\sigma^2_{gs,l}(X_i) = Var[\Delta Y_l | G=g, S=s, X=x]$. To establish the asymptotic normality, notice we can write $\widehat{\tau}_l$ as

$$
\begin{aligned}
\widehat{\tau}_l - B &= \frac{1}{n_{11}} \sum_{i=1}^{n} \left[\mathbb{1}(i\in\mathcal{I}_{11}) - \mathbb{1}(i\in\mathcal{I}_{01}) w_i - \mathbb{1}(i\in\mathcal{I}_{10}) w_i + \mathbb{1}(i\in\mathcal{I}_{00}) w_i \right] \Delta Y_{il} \\
&= \frac{1}{n_{11}} \sum_{i=1}^{n} \mathbb{1}(i\in\mathcal{I}_{11}) (\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i)) \\
&+ \frac{1}{n_{11}} \sum_{i=1}^{n} \left[\mathbb{1}(i\in\mathcal{I}_{11}) - \mathbb{1}(i\in\mathcal{I}_{01}) w_i - \mathbb{1}(i\in\mathcal{I}_{10}) w_i + \mathbb{1}(i\in\mathcal{I}_{00}) w_i \right] (\Delta Y_{il} - \mu_l^{G_iS_i}(X_i))
\end{aligned}
$$

where,

$$
\begin{aligned}
B &= \frac{1}{n_{11}} \sum_{i=1}^{n} \left[\mathbb{1}(i\in\mathcal{I}_{11}) - \mathbb{1}(i\in\mathcal{I}_{01}) w_i \right] \mu_l^{01}(X_i) + \frac{1}{n_{11}} \sum_{i=1}^{n} \left[\mathbb{1}(i\in\mathcal{I}_{11}) - \mathbb{1}(i\in\mathcal{I}_{10}) w_i \right] \mu_l^{10}(X_i) \\
&- \frac{1}{n_{11}} \sum_{i=1}^{n} \left[\mathbb{1}(i\in\mathcal{I}_{11}) - \mathbb{1}(i\in\mathcal{I}_{00}) w_i \right] \mu_l^{00}(X_i)
\end{aligned}
$$

Define $V = V_1 + \mathbb{E}[V_2]$, where,

$$V_1 = \mathbb{E}\left[\left(\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i) - \tau_l\right)^2 \middle| G_i = 1, S_i = 1\right]$$

$$V_2 = \frac{1}{n_{11}} \left(\sum_{i-n \in \mathcal{I}_{11}} \sigma_{11,l}^2(X_{i-n}) + \sum_{i-n \in \mathcal{I}_{01}} w_{i-n}^2 \sigma_{01,l}^2(X_{i-n}) + \sum_{i-n \in \mathcal{I}_{10}} w_{i-n}^2 \sigma_{10,l}^2(X_{i-n}) + \sum_{i-n \in \mathcal{I}_{00}} w_{i-n}^2 \sigma_{00,l}^2(X_{i-n}) \right)$$

Then,

$$\begin{aligned}
&\frac{\widehat{\tau}_l - \tau_l - B}{\sqrt{V/n_{11}}} \\
&= \frac{1}{\sqrt{n_{11}V}} \sum_{i=1}^{n} \mathbb{1}(i \in \mathcal{I}_{11})(\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i) - \tau_l) \\
&+ \frac{1}{\sqrt{n_{11}V}} \sum_{i=1}^{n} \left[\mathbb{1}(i \in \mathcal{I}_{11}) - \mathbb{1}(i \in \mathcal{I}_{01})w_i - \mathbb{1}(i \in \mathcal{I}_{10})w_i + \mathbb{1}(i \in \mathcal{I}_{00})w_i\right] \epsilon_l^{G_i S_i}(X_i)
\end{aligned}$$

Define

$$Z_{n,i} = \begin{cases} \frac{1}{\sqrt{n_{11}V}} \mathbb{1}\{i \in \mathcal{I}_{11}\} \left(\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i) - \tau_l\right), & \text{if } 1 \le i \le n, \\ \frac{1}{\sqrt{n_{11}V}} \Big[\mathbb{1}\{i-n \in \mathcal{I}_{11}\} - \mathbb{1}\{i-n \in \mathcal{I}_{01}\}w_{i-n} & \\ \qquad - \mathbb{1}\{i-n \in \mathcal{I}_{10}\}w_{i-n} + \mathbb{1}\{i-n \in \mathcal{I}_{00}\}w_{i-n}\Big] \epsilon_l^{G_{i-n}S_{i-n}}(X_{i-n}), & \text{if } n+1 \le i \le 2n. \end{cases}$$

so that

$$\frac{\widehat{\tau}_l - \tau_l - B}{\sqrt{V/n_{11}}} = \sum_{i=1}^{2n} Z_{n,i}.$$

Define the σ-fields:

$$\mathcal{F}_{n,i} = \begin{cases} \sigma(\mathbf{G}, \mathbf{S}, X_1, \ldots, X_i) & \text{if } 1 \le i \le n \\ \sigma(\mathbf{G}, \mathbf{S}, \mathbf{X}, \Delta Y_{1l}, \ldots, \Delta Y_{i-n,l}) & \text{if } n+1 \le i \le 2n. \end{cases}$$

Then, the sequence

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

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

$$
\begin{aligned}
&\mathbb{E}[Z_{n,i}^2|\mathcal{F}_{n,i-1}] \\
&= \mathbb{E}\left[\left(\frac{1}{\sqrt{n_{11}V}}\mathbb{1}(i \in \mathcal{I}_{11})\left(\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i) - \tau_l\right)\right)^2 \middle| \mathbf{G}, \mathbf{S}, X_1, \ldots, X_{i-1}\right] \\
&= \frac{\mathbb{1}(i \in \mathcal{I}_{11})}{n_{11}V}\mathbb{E}\left[\left(\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i) - \tau_l\right)^2 \middle| G_i = 1, S_i = 1\right]
\end{aligned}
$$

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

$$
\begin{aligned}
\mathbb{E}[Z_{n,i}^2|\mathcal{F}_{n,i-1}] &= \mathbb{E}\Bigg[\Bigg(\frac{1}{\sqrt{n_{11}V}}\left[\mathbb{1}\{i-n \in \mathcal{I}_{11}\} - \mathbb{1}\{i-n \in \mathcal{I}_{01}\}w_{i-n}\right. \\
&\left.-\mathbb{1}\{i-n \in \mathcal{I}_{10}\}w_{i-n} + \mathbb{1}\{i-n \in \mathcal{I}_{00}\}w_{i-n}\right]\epsilon_l^{G_{i-n}S_{i-n}}(X_{i-n})\Bigg)^2 \Bigg| \mathcal{F}_{n,i-1}\Bigg] \\
&= \frac{\mathbb{1}(i-n \in \mathcal{I}_{11})}{n_{11}V}\sigma^2(1,1,X_{i-n}) + \frac{\mathbb{1}(i-n \in \mathcal{I}_{01})}{n_{11}V}\cdot w_{i-n}^2\sigma^2(0,1,X_{i-n}) \\
&+ \frac{\mathbb{1}(i-n \in \mathcal{I}_{10})}{n_{11}V}\cdot w_{i-n}^2\sigma^2(1,0,X_{i-n}) + \frac{\mathbb{1}(i-n \in \mathcal{I}_{00})}{n_{11}V}\cdot w_{i-n}^2\sigma^2(0,0,X_{i-n})
\end{aligned}
$$

It follows that:

$$
\begin{aligned}
\sum_{i=1}^{2n}\mathbb{E}[Z_{n,i}^2|\mathcal{F}_{n,i}] &= \frac{1}{V}\Bigg(\mathbb{E}\left[\left(\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i) - \tau_l\right)^2 \middle| G_i = 1, S_i = 1\right] \\
&+\frac{1}{n_{11}}\Bigg(\sum_{i-n \in \mathcal{I}_{11}}\sigma^2(1,1,X_{i-n}) + \sum_{i-n \in \mathcal{I}_{01}} w_{i-n}^2\sigma^2(0,1,X_{i-n}) + \sum_{i-n \in \mathcal{I}_{10}} w_{i-n}^2\sigma^2(1,0,X_{i-n}) \\
&+ \sum_{i-n \in \mathcal{I}_{00}} w_{i-n}^2\sigma^2(0,0,X_{i-n})\Bigg)\Bigg) \\
&\xrightarrow{p} 1
\end{aligned}
$$

The last line follows from lemma 6.

On the other hand, under finite moments, for $1 \leq i \leq n$,

$$\begin{aligned}
\mathbb{E}[Z_{n,i}^4 | \mathcal{F}_{n,i}] &= \mathbb{E}\left[\left(\frac{1}{\sqrt{n_{11}V}} \mathbb{1}(i \in \mathcal{I}_{11}) \left(\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i) - \tau_l \right) \right)^4 \middle| \mathbf{G}, \mathbf{S}, X_1, \ldots, X_{i-1} \right] \\
&= \frac{\mathbb{1}(i \in \mathcal{I}_{11})}{n_{11}^2 V^2} \mathbb{E}\left[\left(\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i) - \tau_l \right)^4 \middle| G_i = 1, S_i = 1 \right] \\
&\leq \frac{125 \mathbb{1}(i \in \mathcal{I}_{11})}{n_{11}^2 V^2} \mathbb{E}\left[\left(\mu_l^{11}(X_i) \right)^4 + \left(\mu_l^{01}(X_i) \right)^4 + \left(\mu_l^{10}(X_i) \right)^4 + \left(\mu_l^{00}(X_i) \right)^4 + \tau_l^4 \mid G_i = 1, S_i = 1 \right] \\
&\leq \frac{125 \mathbb{1}(i \in \mathcal{I}_{11})}{n_{11}^2 \underline{\sigma}^2} C
\end{aligned}$$

The last inequality follows from assumption 6.
For $n + 1 \leq i \leq 2n$,

$$\begin{aligned}
&\mathbb{E}[Z_{n,i}^4 | \mathcal{F}_{n,i-1}] = \mathbb{E}\Bigg[\Bigg(\frac{1}{\sqrt{n_{11}V}} [\mathbb{1}\{i - n \in \mathcal{I}_{11}\} - \mathbb{1}\{i - n \in \mathcal{I}_{01}\} w_{i-n} \\
&\quad -\mathbb{1}\{i - n \in \mathcal{I}_{10}\} w_{i-n} + \mathbb{1}\{i - n \in \mathcal{I}_{00}\} w_{i-n}] \, \epsilon_l^{G_{i-n} S_{i-n}}(X_{i-n}) \Bigg)^4 \Bigg| \mathcal{F}_{n,i-1} \Bigg] \\
&= \frac{\mathbb{1}(i - n \in \mathcal{I}_{11})}{n_{11}^2 V^2} \mathbb{E}\left[\left(\epsilon_l^{11}(X_{i-n}) \right)^4 \middle| G_{i-n} = 1, S_{i-n} = 1, X_{i-n} \right] \\
&+ \frac{\mathbb{1}(i - n \in \mathcal{I}_{01})}{n_{11}^2 V^2} w_{i-n}^4 \mathbb{E}\left[\left(\epsilon_l^{01}(X_{i-n}) \right)^4 \middle| G_{i-n} = 0, S_{i-n} = 1, X_{i-n} \right] \\
&+ \frac{\mathbb{1}(i - n \in \mathcal{I}_{10})}{n_{11}^2 V^2} w_{i-n}^4 \mathbb{E}\left[\left(\epsilon_l^{10}(X_{i-n}) \right)^4 \middle| G_{i-n} = 1, S_{i-n} = 0, X_{i-n} \right] \\
&+ \frac{\mathbb{1}(i - n \in \mathcal{I}_{00})}{n_{11}^2 V^2} w_{i-n}^4 \mathbb{E}\left[\left(\epsilon_l^{00}(X_{i-n}) \right)^4 \middle| G_{i-n} = 0, S_{i-n} = 0, X_{i-n} \right] \\
&\leq \frac{C}{\underline{\sigma}^2} \left(\frac{\mathbb{1}(i - n \in \mathcal{I}_{11})}{n_{11}^2} + \frac{\mathbb{1}(i - n \in \mathcal{I}_{01})}{n^2} w_{i-n}^4 + \frac{\mathbb{1}(i - n \in \mathcal{I}_{10})}{n_{11}^2} w_{i-n}^4 + \frac{\mathbb{1}(i - n \in \mathcal{I}_{00})}{n_{11}^2} w_{i-n}^4 \right)
\end{aligned}$$

for some finite constant A_4. Thus,

$$\begin{aligned}
\sum_{i=1}^{2n} \mathbb{E}[Z_{n,i}^4 | \mathcal{F}_{n,i-1}] &\leq \frac{125 C \mathbb{1}(i \in \mathcal{I}_{11})}{n_{11} \underline{\sigma}^2} \\
&+ \frac{C}{n_{11} \underline{\sigma}^2} \left(\frac{1}{n_{11}} \sum_{i=n+1}^{2n} \mathbb{1}(i \in \mathcal{I}_{11}) \right) + \frac{C}{n_{11} \underline{\sigma}^2} \left(\frac{1}{n_{11}} \sum_{i=n+1}^{2n} \mathbb{1}(i \in \mathcal{I}_{01}) w_{i-n}^4 \right) \\
&\frac{C}{n_{11} \underline{\sigma}^2} \left(\frac{1}{n_{11}} \sum_{i=n+1}^{2n} \mathbb{1}(i \in \mathcal{I}_{10}) w_{i-n}^4 \right) + \frac{C}{n_{11} \underline{\sigma}^2} \left(\frac{1}{n_{11}} \sum_{i=n+1}^{2n} \mathbb{1}(i \in \mathcal{I}_{00}) w_{i-n}^4 \right) \\
&\xrightarrow{p} 0
\end{aligned}$$

by Lemma 7(iii). Thus, by the martingale central limit theorem

$$\sum_{i=1}^{2n} Z_{n,i} \xrightarrow{d} \mathcal{N}(0,1)$$

Therefore, by lemma 6 and the Slutsky theorem,

$$(V_1+V_2)^{-1/2}\sqrt{n_{11}}\,(\widehat{\tau}_l-\tau_l-B)=\frac{\sqrt{V}}{\sqrt{V_1+V_2}}\frac{\widehat{\tau}_l-\tau_l-B}{\sqrt{V/n_{11}}} \xrightarrow{d} \mathcal{N}(0,1)$$

which completes the proof.

■

Proof. Proof of theorem 3

$$\begin{aligned}
\widehat{V}_1 &= \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\Delta Y_{il}-\widehat{\Delta Y_{il}^{01}}-\widehat{\Delta Y_{il}^{10}}+\widehat{\Delta Y_{il}^{00}}-\widehat{\tau}_l\right)^2 \\
&= \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\Delta Y_{il}-\widehat{\Delta Y_{il}^{01}}-\widehat{\Delta Y_{il}^{10}}+\widehat{\Delta Y_{il}^{00}}-\tau_l-(\widehat{\tau}_l-\tau_l)\right)^2 \\
&= \left(\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\Delta Y_{il}-\widehat{\Delta Y_{il}^{01}}-\widehat{\Delta Y_{il}^{10}}+\widehat{\Delta Y_{il}^{00}}-\tau_l\right)^2\right)+(\tau_l-\widehat{\tau}_l)^2 \\
&- \frac{2(\widehat{\tau}_l-\tau_l)}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\Delta Y_{il}-\widehat{\Delta Y_{il}^{01}}-\widehat{\Delta Y_{il}^{10}}+\widehat{\Delta Y_{il}^{00}}-\tau_l\right) \\
&= \left(\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\Delta Y_{il}-\widehat{\Delta Y_{il}^{01}}-\widehat{\Delta Y_{il}^{10}}+\widehat{\Delta Y_{il}^{00}}-\tau_l\right)^2\right)-(\tau_l-\widehat{\tau}_l)^2 \\
&= \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\Delta Y_{il}-\widehat{\Delta Y_{il}^{01}}-\widehat{\Delta Y_{il}^{10}}+\widehat{\Delta Y_{il}^{00}}-\tau_l\right)^2+o_p(1)
\end{aligned}$$

The third equality follows from lemma 13. The last equality follows from the consistency of

$\widehat{\tau}_l$.

$$\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\Delta Y_{il}-\widehat{\Delta Y_{il}^{01}}-\widehat{\Delta Y_{il}^{10}}+\widehat{\Delta Y_{il}^{00}}-\tau_l\right)^2$$

$$=\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\Delta Y_{il}-\sum_{j\in\mathcal{I}_{01}}w_{ij}\Delta Y_{jl}-\sum_{j\in\mathcal{I}_{10}}w_{ij}\Delta Y_{jl}+\sum_{j\in\mathcal{I}_{00}}w_{ij}\Delta Y_{jl}-\tau_l\right)^2$$

$$=\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\mu_l^{11}(X_i)+\epsilon_{il}^{11}-\sum_{j\in\mathcal{I}_{01}}w_{ij}\left(\mu_l^{01}(X_j)+\epsilon_{jl}^{01}\right)-\sum_{j\in\mathcal{I}_{10}}w_{ij}\left(\mu_l^{10}(X_j)+\epsilon_{jl}^{10}\right)+\sum_{j\in\mathcal{I}_{00}}w_{ij}\left(\mu_l^{00}(X_j)+\epsilon_{jl}^{00}\right)\right.$$

$$=\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\mu_l^{11}(X_i)-\sum_{j\in\mathcal{I}_{01}}w_{ij}\mu_l^{01}(X_j)-\sum_{j\in\mathcal{I}_{10}}w_{ij}\mu_l^{10}(X_j)+\sum_{j\in\mathcal{I}_{00}}w_{ij}\mu_l^{00}(X_j)-\tau\right.$$

$$\left.\epsilon_{il}^{11}-\sum_{j\in\mathcal{I}_{01}}w_{ij}\epsilon_{jl}^{01}-\sum_{j\in\mathcal{I}_{10}}w_{ij}\epsilon_{jl}^{10}+\sum_{j\in\mathcal{I}_{00}}w_{ij}\epsilon_{jl}^{00}\right)^2$$

$$=\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\mu_l^{11}(X_i)-\sum_{j\in\mathcal{I}_{01}}w_{ij}\mu_l^{01}(X_j)-\sum_{j\in\mathcal{I}_{10}}w_{ij}\mu_l^{10}(X_j)+\sum_{j\in\mathcal{I}_{00}}w_{ij}\mu_l^{00}(X_j)-\tau_l\right)^2$$

$$+\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\epsilon_{il}^{11}-\sum_{j\in\mathcal{I}_{01}}w_{ij}\epsilon_{jl}^{01}-\sum_{j\in\mathcal{I}_{10}}w_{ij}\epsilon_{jl}^{10}+\sum_{j\in\mathcal{I}_{00}}w_{ij}\epsilon_{jl}^{00}\right)^2$$

$$+\frac{2}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\mu_l^{11}(X_i)-\sum_{j\in\mathcal{I}_{01}}w_{ij}\mu_l^{01}(X_j)-\sum_{j\in\mathcal{I}_{10}}w_{ij}\mu_l^{10}(X_j)+\sum_{j\in\mathcal{I}_{00}}w_{ij}\mu_l^{00}(X_j)-\tau_l\right)$$

$$\left(\epsilon_{il}^{11}-\sum_{j\in\mathcal{I}_{01}}w_{ij}\epsilon_{jl}^{01}-\sum_{j\in\mathcal{I}_{10}}w_{ij}\epsilon_{jl}^{10}+\sum_{j\in\mathcal{I}_{00}}w_{ij}\epsilon_{jl}^{00}\right)$$

First term:

$$\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\mu_l^{11}(X_i) - \sum_{j \in \mathcal{I}_{01}} w_{ij} \mu_l^{01}(X_j) - \sum_{j \in \mathcal{I}_{10}} w_{ij} \mu_l^{10}(X_j) + \sum_{j \in \mathcal{I}_{00}} w_{ij} \mu_l^{00}(X_j) - \tau_l \right)^2$$

$$= \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \big(\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i) - \tau_l$$

$$+ \sum_{j \in \mathcal{I}_{01}} w_{ij} \left(\mu_l^{01}(X_i) - \mu_l^{01}(X_j) \right) + \sum_{j \in \mathcal{I}_{10}} w_{ij} \left(\mu_l^{10}(X_i) - \mu_l^{10}(X_j) \right) - \sum_{j \in \mathcal{I}_{00}} w_{ij} \left(\mu_l^{00}(X_i) - \mu_l^{00}(X_j) \right) \Bigg)^2$$

$$= \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i) - \tau_l \right)^2$$

$$+ \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\sum_{j \in \mathcal{I}_{01}} w_{ij} \left(\mu_l^{01}(X_i) - \mu_l^{01}(X_j) \right) + \sum_{j \in \mathcal{I}_{10}} w_{ij} \left(\mu_l^{10}(X_i) - \mu_l^{10}(X_j) \right) - \sum_{j \in \mathcal{I}_{00}} w_{ij} \left(\mu_l^{00}(X_i) - \mu_l^{00}(X_j) \right) \right)$$

$$+ \frac{2}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i) - \tau_l \right)$$

$$\left(\sum_{j \in \mathcal{I}_{01}} w_{ij} \left(\mu_l^{01}(X_i) - \mu_l^{01}(X_j) \right) + \sum_{j \in \mathcal{I}_{10}} w_{ij} \left(\mu_l^{10}(X_i) - \mu_l^{10}(X_j) \right) - \sum_{j \in \mathcal{I}_{00}} w_{ij} \left(\mu_l^{00}(X_i) - \mu_l^{00}(X_j) \right) \right)$$

Notice that:

$$\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\sum_{j \in \mathcal{I}_{01}} w_{ij} \left(\mu_l^{01}(X_i) - \mu_l^{01}(X_j) \right) + \sum_{j \in \mathcal{I}_{10}} w_{ij} \left(\mu_l^{10}(X_i) - \mu_l^{10}(X_j) \right) - \sum_{j \in \mathcal{I}_{00}} w_{ij} \left(\mu_l^{00}(X_i) - \mu_l^{00}(X_j) \right) \right)^2$$

$$\leq \frac{3}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\sum_{j \in \mathcal{I}_{01}} w_{ij} \left(\mu_l^{01}(X_i) - \mu_l^{01}(X_j) \right) \right)^2 + \left(\sum_{j \in \mathcal{I}_{10}} w_{ij} \left(\mu_l^{10}(X_i) - \mu_l^{10}(X_j) \right) \right)^2 + \left(\sum_{j \in \mathcal{I}_{00}} w_{ij} \left(\mu_l^{00}(X_i) - \mu \right. \right.$$

$$\leq \frac{3}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\sum_{j \in \mathcal{I}_{01}} w_{ij} \left| \left(\mu_l^{01}(X_i) - \mu_l^{01}(X_j) \right) \right| \right)^2 + \left(\sum_{j \in \mathcal{I}_{10}} w_{ij} \left| \left(\mu_l^{10}(X_i) - \mu_l^{10}(X_j) \right) \right| \right)^2 + \left(\sum_{j \in \mathcal{I}_{00}} w_{ij} \left| \left(\mu_l^{00}(X_i) \right. \right. \right.$$

$$\leq \frac{3C}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\sum_{j \in \mathcal{I}_{01}} w_{ij} \|X_i - X_j\| \right)^2 + \left(\sum_{j \in \mathcal{I}_{10}} w_{ij} \|X_i - X_j\| \right)^2 + \left(\sum_{j \in \mathcal{I}_{00}} w_{ij} \|X_i - X_j\| \right)^2$$

$$= o_p(1)$$

The first inequality follows from $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$, the second inequality follows from triangular inequality, the third inequality follows from the Lipschitz continuity of the conditional mean function. The last equality follows from lemma 7(iv). In addition,

$$
\begin{aligned}
&\frac{2}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left| \left(\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i) - \tau_l \right) \right| \\
&\left| \left(\sum_{j \in \mathcal{I}_{01}} w_{ij} \left(\mu_l^{01}(X_i) - \mu_l^{01}(X_j) \right) + \sum_{j \in \mathcal{I}_{10}} w_{ij} \left(\mu_l^{10}(X_i) - \mu_l^{10}(X_j) \right) - \sum_{j \in \mathcal{I}_{00}} w_{ij} \left(\mu_l^{00}(X_i) - \mu_l^{00}(X_j) \right) \right) \right| \\
&\leq \frac{2C}{n_{11}} \sum_{j \in \mathcal{I}_{11}} \sum_{j \in \mathcal{I}_{01}} w_{ij} \left| \left(\mu_l^{01}(X_i) - \mu_l^{01}(X_j) \right) \right| + \sum_{j \in \mathcal{I}_{10}} w_{ij} \left| \left(\mu_l^{10}(X_i) - \mu_l^{10}(X_j) \right) \right| + \sum_{j \in \mathcal{I}_{00}} w_{ij} \left| \left(\mu_l^{00}(X_i) - \mu_l^{00}(X_j) \right) \right| \\
&\leq \frac{2C'}{n_{11}} \sum_{j \in \mathcal{I}_{11}} \sum_{j \in \mathcal{I}_{01}} w_{ij} \left\| X_i - X_j \right\| + \sum_{j \in \mathcal{I}_{10}} w_{ij} \left\| X_i - X_j \right\| + \sum_{j \in \mathcal{I}_{00}} w_{ij} \left\| X_i - X_j \right\| \\
&= o_p(1)
\end{aligned}
$$

This implies that

$$
\begin{aligned}
&\frac{2}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i) - \tau_l \right) \\
&\left(\sum_{j \in \mathcal{I}_{01}} w_{ij} \left(\mu_l^{01}(X_i) - \mu_l^{01}(X_j) \right) + \sum_{j \in \mathcal{I}_{10}} w_{ij} \left(\mu_l^{10}(X_i) - \mu_l^{10}(X_j) \right) - \sum_{j \in \mathcal{I}_{00}} w_{ij} \left(\mu_l^{00}(X_i) - \mu_l^{00}(X_j) \right) \right) = o_p(1)
\end{aligned}
$$

Therefore, combine the results above,

$$
\begin{aligned}
&\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\mu_l^{11}(X_i) - \sum_{j \in \mathcal{I}_{01}} w_{ij} \mu_l^{01}(X_j) - \sum_{j \in \mathcal{I}_{10}} w_{ij} \mu_l^{10}(X_j) + \sum_{j \in \mathcal{I}_{00}} w_{ij} \mu_l^{00}(X_j) - \tau_l \right)^2 \\
&= \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i) - \tau_l \right)^2 + o_p(1)
\end{aligned}
$$

Notice that by WLLN,

$$
\begin{aligned}
&\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i) - \tau_l \right)^2 \\
&\xrightarrow{p} V_1 = \mathbb{E}\left[\left(\mu_l^{11}(X_i) - \mu_l^{01}(X_i) - \mu_l^{10}(X_i) + \mu_l^{00}(X_i) - \tau_l \right)^2 | G_i = 1, S_i = 1 \right]
\end{aligned}
$$

Therefore,

$$
\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\mu_l^{11}(X_i) - \sum_{j \in \mathcal{I}_{01}} w_{ij} \mu_l^{01}(X_j) - \sum_{j \in \mathcal{I}_{10}} w_{ij} \mu_l^{10}(X_j) + \sum_{j \in \mathcal{I}_{00}} w_{ij} \mu_l^{00}(X_j) - \tau_l \right)^2 = V_1 + o_p(1)
$$

Second term:

$$
\begin{aligned}
&\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\epsilon_{il}^{11}-\sum_{j\in\mathcal{I}_{01}}w_{ij}\epsilon_{jl}^{01}-\sum_{j\in\mathcal{I}_{10}}w_{ij}\epsilon_{jl}^{10}+\sum_{j\in\mathcal{I}_{00}}w_{ij}\epsilon_{jl}^{00}\right)^2\\
&=\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\left(\epsilon_{il}^{11}\right)^2+\sum_{j\in\mathcal{I}_{01}}w_{ij}^2\left(\epsilon_{jl}^{01}\right)^2+\sum_{j\in\mathcal{I}_{10}}w_{ij}^2\left(\epsilon_{jl}^{10}\right)^2+\sum_{j\in\mathcal{I}_{00}}w_{ij}^2\left(\epsilon_{jl}^{00}\right)^2\right)\\
&\quad+\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\sum_{j\in\mathcal{I}_{01}}\sum_{\substack{j'\in\mathcal{I}_{01}\\ j'\neq j}}w_{ij}w_{ij'}\epsilon_{jl}^{01}\epsilon_{j'l}^{01}+\sum_{j\in\mathcal{I}_{10}}\sum_{\substack{j'\in\mathcal{I}_{10}\\ j'\neq j}}w_{ij}w_{ij'}\epsilon_{jl}^{10}\epsilon_{j'l}^{10}+\sum_{j\in\mathcal{I}_{00}}\sum_{\substack{j'\in\mathcal{I}_{00}\\ j'\neq j}}w_{ij}w_{ij'}\epsilon_{jl}^{00}\epsilon_{j'l}^{00}\right)\\
&\quad+\frac{2}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\sum_{j\in\mathcal{I}_{01}}\sum_{j'\in\mathcal{I}_{10}}w_{ij}w_{ij'}\epsilon_{jl}^{01}\epsilon_{j'l}^{10}-\sum_{j\in\mathcal{I}_{01}}\sum_{j'\in\mathcal{I}_{00}}w_{ij}w_{ij'}\epsilon_{jl}^{01}\epsilon_{j'l}^{00}-\sum_{j\in\mathcal{I}_{10}}\sum_{j'\in\mathcal{I}_{00}}w_{ij}w_{ij'}\epsilon_{jl}^{10}\epsilon_{j'l}^{00}\right)\\
&\quad-\frac{2}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\sum_{j\in\mathcal{I}_{01}}w_{ij}\epsilon_{il}^{11}\epsilon_{jl}^{01}+\sum_{j\in\mathcal{I}_{10}}w_{ij}\epsilon_{il}^{11}\epsilon_{jl}^{10}-\sum_{j\in\mathcal{I}_{00}}w_{ij}\epsilon_{il}^{11}\epsilon_{jl}^{00}\right)
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\epsilon_{il}^{11}\right)^2-\sigma_{11,l}^2(X_i)\,\middle|\,\mathbf{G},\mathbf{S},\mathbf{X}\right]\\
&=\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\mathbb{E}\left[\left(\epsilon_{il}^{11}\right)^2-\sigma_{11,l}^2(X_i)\mid G_i,S_i,X_i\right]\\
&=0
\end{aligned}
$$

$$
\begin{aligned}
Var\left(\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\epsilon_{il}^{11}\right)^2-\sigma_{11,l}^2(X_i)\,\middle|\,\mathbf{G},\mathbf{S},\mathbf{X}\right)&=\frac{1}{n_{11}}Var\left(\left(\epsilon_{il}^{11}\right)^2-\sigma_{11,l}^2(X_i)\,\middle|\,\mathbf{G},\mathbf{S},\mathbf{X},G_i=1,S_i=1\right)\\
&=\frac{1}{n_{11}}Var\left(\left(\epsilon_{il}^{11}\right)^2\,\middle|\,\mathbf{G},\mathbf{S},\mathbf{X},G_i=1,S_i=1\right)\\
&\leq\frac{1}{n_{11}}\mathbb{E}\left[\left(\epsilon_{il}^{11}\right)^4\,\middle|\,\mathbf{G},\mathbf{S},\mathbf{X},G_i=1,S_i=1\right]\\
&\leq\frac{C}{n_{11}}\\
&=o(1)
\end{aligned}
$$

The last inequality follows from lemma 12. Then by lemma 9,

$$
\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\epsilon_{il}^{11}\right)^2-\sigma_{11,l}^2(X_i)=o_p(1)
$$

For any $(g,s) \in \{(0,1),(1,0),(0,0)\}$,

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \sum_{j \in \mathcal{I}_{gs}} w_{ij}^2 \left[\left(\epsilon_{jl}^{gs}\right)^2 - \sigma_{gs,l}^2(X_j) \right] \middle| \mathbf{G}, \mathbf{S}, \mathbf{X} \right] \\
&= \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \sum_{j \in \mathcal{I}_{gs}} w_{ij}^2 \mathbb{E}\left[\left(\epsilon_{jl}^{gs}\right)^2 - \sigma_{gs,l}^2(X_j) \middle| G_j, S_j, X_j \right] \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
&Var\left(\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \sum_{j \in \mathcal{I}_{gs}} w_{ij}^2 \left[\left(\epsilon_{jl}^{gs}\right)^2 - \sigma_{gs,l}^2(X_j) \right] \middle| \mathbf{G}, \mathbf{S}, \mathbf{X} \right) \\
&= Var\left(\frac{1}{n_{11}} \sum_{j \in \mathcal{I}_{gs}} \sum_{i \in \mathcal{I}_{11}} w_{ij}^2 \left[\left(\epsilon_{jl}^{gs}\right)^2 - \sigma_{gs,l}^2(X_j) \right] \middle| \mathbf{G}, \mathbf{S}, \mathbf{X} \right) \\
&= \frac{1}{n_{11}^2} \sum_{j \in \mathcal{I}_{gs}} Var\left(\sum_{i \in \mathcal{I}_{11}} w_{ij}^2 \left(\epsilon_{jl}^{gs}\right)^2 \middle| \mathbf{G}, \mathbf{S}, \mathbf{X} \right) \\
&\leq \frac{1}{n_{11}^2} \sum_{j \in \mathcal{I}_{gs}} \mathbb{E}\left[\left(\sum_{i \in \mathcal{I}_{11}} w_{ij}^2 \left(\epsilon_{jl}^{gs}\right)^2 \right)^2 \middle| \mathbf{G}, \mathbf{S}, \mathbf{X} \right] \\
&= \frac{1}{n_{11}^2} \sum_{j \in \mathcal{I}_{gs}} \mathbb{E}\left[\sum_{i \in \mathcal{I}_{11}} \sum_{i' \in \mathcal{I}_{11}} w_{ij}^2 w_{i'j}^2 \left(\epsilon_{jl}^{gs}\right)^4 \middle| \mathbf{G}, \mathbf{S}, \mathbf{X} \right] \\
&= \frac{1}{n_{11}^2} \sum_{j \in \mathcal{I}_{gs}} \sum_{i \in \mathcal{I}_{11}} \sum_{i' \in \mathcal{I}_{11}} w_{ij}^2 w_{i'j}^2 \mathbb{E}\left[\left(\epsilon_{jl}^{gs}\right)^4 \middle| \mathbf{G}, \mathbf{S}, \mathbf{X} \right] \\
&\leq \frac{C}{n_{11}^2} \sum_{j \in \mathcal{I}_{gs}} \sum_{i \in \mathcal{I}_{11}} \sum_{i' \in \mathcal{I}_{11}} w_{ij}^2 w_{i'j}^2 \\
&\leq \frac{C}{n_{11}^2} \sum_{j \in \mathcal{I}_{gs}} \sum_{i \in \mathcal{I}_{11}} w_{ij} \sum_{i' \in \mathcal{I}_{11}} w_{i'j} \\
&= \frac{C}{n_{11}^2} \sum_{j \in \mathcal{I}_{gs}} w_j^2 \\
&= o_p(1)
\end{aligned}
$$

This last line follows from lemma 7(i). By lemma 9,

$$
\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \sum_{j \in \mathcal{I}_{gs}} w_{ij}^2 \left[\left(\epsilon_{jl}^{gs}\right)^2 - \sigma_{gs,l}^2(X_j) \right] = o_p(1)
$$

For $(g,s) \in \{(0,1),(1,0),(0,0)\}$,

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}}\sum_{\substack{j'\in\mathcal{I}_{gs}\\ j'\neq j}} w_{ij}w_{ij'}\epsilon_{jl}^{gs}\epsilon_{j'l}^{gs}\right|\left.\mathbf{G},\mathbf{S},\mathbf{X}\vphantom{\sum_{\substack{j'\\j'}}}\right] &= \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}}\sum_{\substack{j'\in\mathcal{I}_{gs}\\ j'\neq j}} w_{ij}w_{ij'}\mathbb{E}\left[\epsilon_{jl}^{gs}\epsilon_{j'l}^{gs}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right] \\
&= \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}}\sum_{\substack{j'\in\mathcal{I}_{gs}\\ j'\neq j}} w_{ij}w_{ij'}\mathbb{E}\left[\epsilon_{jl}^{gs}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right]\mathbb{E}\left[\epsilon_{j'l}^{gs}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right] \\
&= 0
\end{aligned}$$

On the other hand,

$$\begin{aligned}
& Var\left(\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}}\sum_{\substack{j'\in\mathcal{I}_{gs}\\ j'\neq j}} w_{ij}w_{ij'}\epsilon_{jl}^{gs}\epsilon_{j'l}^{gs}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right) \\
&= \frac{1}{n_{11}^2}Var\left(\sum_{j\in\mathcal{I}_{gs}}\sum_{\substack{j'\in\mathcal{I}_{gs}\\ j'\neq j}}\sum_{i\in\mathcal{I}_{11}} w_{ij}w_{ij'}\epsilon_{jl}^{gs}\epsilon_{j'l}^{gs}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right) \\
&= \frac{1}{n_{11}^2}\sum_{j\in\mathcal{I}_{gs}}\sum_{\substack{j'\in\mathcal{I}_{gs}\\ j'\neq j}} Var\left(\sum_{i\in\mathcal{I}_{11}} w_{ij}w_{ij'}\epsilon_{jl}^{gs}\epsilon_{j'l}^{gs}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right) \\
&\leq \frac{1}{n_{11}^2}\sum_{j\in\mathcal{I}_{gs}}\sum_{\substack{j'\in\mathcal{I}_{gs}\\ j'\neq j}}\left(\sum_{i\in\mathcal{I}_{11}} w_{ij}w_{ij'}\right)^2\mathbb{E}\left[\left(\epsilon_{jl}^{gs}\right)^2\left(\epsilon_{j'l}^{gs}\right)^2\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right] \\
&\leq \frac{C}{n_{11}^2}\sum_{j\in\mathcal{I}_{gs}}\sum_{\substack{j'\in\mathcal{I}_{gs}\\ j'\neq j}}\left(\sum_{i\in\mathcal{I}_{11}} w_{ij}w_{ij'}\right)^2 \\
&\leq \frac{C}{n_{11}^2}\max_{r\in\mathcal{I}_{gs}} w_r\sum_{j\in\mathcal{I}_{gs}}\sum_{\substack{j'\in\mathcal{I}_{gs}\\ j'\neq j}}\sum_{i\in\mathcal{I}_{11}} w_{ij}w_{ij'} \\
&\leq \frac{C}{n_{11}^2}\max_{r\in\mathcal{I}_{gs}} w_r\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\sum_{j'\in\mathcal{I}_{gs}} w_{ij'} \\
&= \frac{C}{n_{11}}O_p(n^{1/2}) \\
&= o_p(1)
\end{aligned}$$

The second equality follows because all the cross product term has zero conditional covariance. By lemma 9,

$$\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \sum_{j \in \mathcal{I}_{gs}} \sum_{\substack{j' \in \mathcal{I}_{gs} \\ j' \neq j}} w_{ij} w_{ij'} \epsilon_{jl}^{gs} \epsilon_{j'l}^{gs} = o_p(1)$$

Therefore,

$$\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\sum_{j \in \mathcal{I}_{01}} \sum_{\substack{j' \in \mathcal{I}_{01} \\ j' \neq j}} w_{ij} w_{ij'} \epsilon_{jl}^{01} \epsilon_{j'l}^{01} + \sum_{j \in \mathcal{I}_{10}} \sum_{\substack{j' \in \mathcal{I}_{10} \\ j' \neq j}} w_{ij} w_{ij'} \epsilon_{jl}^{10} \epsilon_{j'l}^{10} + \sum_{j \in \mathcal{I}_{00}} \sum_{\substack{j' \in \mathcal{I}_{00} \\ j' \neq j}} w_{ij} w_{ij'} \epsilon_{jl}^{00} \epsilon_{j'l}^{00} \right) = o_p(1)$$

For $(g, s) \in \{(0,1), (1,0), (0,0)\}$ and $(g', s') \in \{(0,1), (1,0), (0,0)\}$

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \sum_{j \in \mathcal{I}_{gs}} \sum_{j \in \mathcal{I}_{g's'}} w_{ij} w_{ij'} \epsilon_{jl}^{gs} \epsilon_{j'l}^{g's'} \middle| \mathbf{G}, \mathbf{S}, \mathbf{X} \right] &= \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \sum_{j \in \}\int} \sum_{j \in \}'\int'} w_{ij} w_{ij'} \mathbb{E}\left[\epsilon_{jl}^{gs} \epsilon_{j'l}^{g's'} \middle| \mathbf{G}, \mathbf{S}, \mathbf{X} \right] \\
&= \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \sum_{j \in \}\int} \sum_{j \in \}'\int'} w_{ij} w_{ij'} \mathbb{E}\left[\epsilon_{jl}^{gs} \middle| \mathbf{G}, \mathbf{S}, \mathbf{X} \right] \mathbb{E}\left[\epsilon_{j'l}^{g's'} \middle| \mathbf{G}, \mathbf{S}, \mathbf{X} \right] \\
&= 0
\end{aligned}$$

On the other hand,

$$
\begin{aligned}
&Var\left(\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}}\sum_{j'\in\mathcal{I}_{g's'}} w_{ij}w_{ij'}\epsilon_{jl}^{gs}\epsilon_{j'l}^{g's'}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\\
&=\frac{1}{n_{11}^2}Var\left(\sum_{j\in\mathcal{I}_{gs}}\sum_{j'\in\mathcal{I}_{g's'}}\sum_{i\in\mathcal{I}_{11}} w_{ij}w_{ij'}\epsilon_{jl}^{gs}\epsilon_{j'l}^{g's'}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\\
&=\frac{1}{n_{11}^2}\sum_{j\in\mathcal{I}_{gs}}\sum_{j'\in\mathcal{I}_{g's'}}Var\left(\sum_{i\in\mathcal{I}_{11}} w_{ij}w_{ij'}\epsilon_{jl}^{gs}\epsilon_{j'l}^{g's'}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\\
&=\frac{1}{n_{11}^2}\sum_{j\in\mathcal{I}_{gs}}\sum_{j'\in\mathcal{I}_{g's'}}\left(\sum_{i\in\mathcal{I}_{11}} w_{ij}w_{ij'}\right)^2 Var\left(\epsilon_{jl}^{gs}\epsilon_{j'l}^{g's'}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\\
&\leq\frac{1}{n_{11}^2}\sum_{j\in\mathcal{I}_{gs}}\sum_{j'\in\mathcal{I}_{g's'}}\left(\sum_{i\in\mathcal{I}_{11}} w_{ij}w_{ij'}\right)^2 \mathbb{E}\left[\left(\epsilon_{jl}^{gs}\right)^2\left(\epsilon_{j'l}^{g's'}\right)^2\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right]\\
&\leq\frac{C}{n_{11}^2}\sum_{j\in\mathcal{I}_{gs}}\sum_{j'\in\mathcal{I}_{g's'}}\left(\sum_{i\in\mathcal{I}_{11}} w_{ij}w_{ij'}\right)^2\\
&\leq\frac{C}{n_{11}^2}\sum_{j\in\mathcal{I}_{gs}}\sum_{j'\in\mathcal{I}_{g's'}}w_j\sum_{i\in\mathcal{I}_{11}} w_{ij}w_{ij'}\\
&\leq\frac{C}{n_{11}^2}\max_{r\in\mathcal{I}_{gs}}w_r\sum_{j\in\mathcal{I}_{gs}}\sum_{j'\in\mathcal{I}_{g's'}}\sum_{i\in\mathcal{I}_{11}} w_{ij}w_{ij'}\\
&=\frac{C}{n_{11}^2}\max_{r\in\mathcal{I}_{gs}}w_r\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}}w_{ij}\sum_{j'\in\mathcal{I}_{g's'}} w_{ij'}\\
&=\frac{C}{n_{11}}\max_{r\in\mathcal{I}_{gs}}w_r\\
&=\frac{C}{n_{11}}O_p(n^{1/2})\\
&=o_p(1)
\end{aligned}
$$

By lemma 9,

$$
\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}}\sum_{j'\in\mathcal{I}_{g's'}} w_{ij}w_{ij'}\epsilon_{jl}^{gs}\epsilon_{j'l}^{g's'}=o_p(1)
$$

This implies that,

$$\frac{2}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\sum_{j\in\mathcal{I}_{01}}\sum_{j'\in\mathcal{I}_{10}} w_{ij}w_{ij'}\epsilon_{jl}^{01}\epsilon_{j'l}^{10}-\sum_{j\in\mathcal{I}_{01}}\sum_{j'\in\mathcal{I}_{00}} w_{ij}w_{ij'}\epsilon_{jl}^{01}\epsilon_{j'l}^{00}-\sum_{j\in\mathcal{I}_{10}}\sum_{j'\in\mathcal{I}_{00}} w_{ij}w_{ij'}\epsilon_{jl}^{10}\epsilon_{j'l}^{00}\right)=o_p(1)$$

For $(g,s)\in\{(0,1),(1,0),(0,0)\}$,

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\epsilon_{il}^{11}\epsilon_{jl}^{gs}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right] &= \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\mathbb{E}\left[\epsilon_{il}^{11}\epsilon_{jl}^{gs}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right]\\
&= \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\mathbb{E}\left[\epsilon_{il}^{11}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right]\mathbb{E}\left[\epsilon_{jl}^{gs}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right]\\
&= 0
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&Var\left(\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\epsilon_{il}^{11}\epsilon_{jl}^{gs}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\\
&= \frac{1}{n_{11}^2}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}^2 Var\left(\epsilon_{il}^{11}\epsilon_{jl}^{gs}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\\
&\leq \frac{1}{n_{11}^2}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}^2 \mathbb{E}\left[\left(\epsilon_{il}^{11}\right)^2\left(\epsilon_{jl}^{gs}\right)^2\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right]\\
&\leq \frac{C}{n_{11}^2}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}^2\\
&\leq \frac{C}{n_{11}^2}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\\
&= \frac{C}{n_{11}}\\
&= o_p(1)
\end{aligned}$$

By lemma 9,

$$\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\epsilon_{il}^{11}\epsilon_{jl}^{gs}=o_p(1)$$

This implies that:

$$\frac{2}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\sum_{j \in \mathcal{I}_{01}} w_{ij} \epsilon_{il}^{11} \epsilon_{jl}^{01} + \sum_{j \in \mathcal{I}_{10}} w_{ij} \epsilon_{il}^{11} \epsilon_{jl}^{10} - \sum_{j \in \mathcal{I}_{00}} w_{ij} \epsilon_{il}^{11} \epsilon_{jl}^{00} \right) = o_p(1)$$

Therefore,

$$\begin{aligned}
&\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\epsilon_{il}^{11} - \sum_{j \in \mathcal{I}_{01}} w_{ij} \epsilon_{jl}^{01} - \sum_{j \in \mathcal{I}_{10}} w_{ij} \epsilon_{jl}^{10} + \sum_{j \in \mathcal{I}_{00}} w_{ij} \epsilon_{jl}^{00} \right)^2 \\
&= \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\sigma_{11,l}^2(X_i) + \sum_{j \in \mathcal{I}_{01}} w_{ij}^2 \sigma_{01,l}^2(X_j) + \sum_{j \in \mathcal{I}_{10}} w_{ij}^2 \sigma_{10,l}^2(X_j) + \sum_{j \in \mathcal{I}_{00}} w_{ij}^2 \sigma_{00,l}^2(X_j) \right) + o_p(1)
\end{aligned}$$

Third term:

$$\begin{aligned}
&\frac{2}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\mu_l^{11}(X_i) - \sum_{j \in \mathcal{I}_{01}} w_{ij} \mu_l^{01}(X_j) - \sum_{j \in \mathcal{I}_{10}} w_{ij} \mu_l^{10}(X_j) + \sum_{j \in \mathcal{I}_{00}} w_{ij} \mu_l^{00}(X_j) - \tau_l \right) \\
&\left(\epsilon_{il}^{11} - \sum_{j \in \mathcal{I}_{01}} w_{ij} \epsilon_{jl}^{01} - \sum_{j \in \mathcal{I}_{10}} w_{ij} \epsilon_{jl}^{10} + \sum_{j \in \mathcal{I}_{00}} w_{ij} \epsilon_{jl}^{00} \right)
\end{aligned}$$

Define $\Phi_{1,i} = \mu_l^{11}(X_i) - \sum_{j \in \mathcal{I}_{01}} w_{ij} \mu_l^{01}(X_j) - \sum_{j \in \mathcal{I}_{10}} w_{ij} \mu_l^{10}(X_j) + \sum_{j \in \mathcal{I}_{00}} w_{ij} \mu_l^{00}(X_j) - \tau_l$. First notice that

$$\begin{aligned}
|\Phi_{1,i}| &= \left| \mu_l^{11}(X_i) - \sum_{j \in \mathcal{I}_{01}} w_{ij} \mu_l^{01}(X_j) - \sum_{j \in \mathcal{I}_{10}} w_{ij} \mu_l^{10}(X_j) + \sum_{j \in \mathcal{I}_{00}} w_{ij} \mu_l^{00}(X_j) - \tau_l \right| \\
&\leq \left| \mu_l^{11}(X_i) \right| + \sum_{j \in \mathcal{I}_{01}} w_{ij} \left| \mu_l^{01}(X_j) \right| + \sum_{j \in \mathcal{I}_{01}} w_{ij} \left| \mu_l^{10}(X_j) \right| + \sum_{j \in \mathcal{I}_{01}} w_{ij} \left| \mu_l^{00}(X_j) \right| + |\tau_l| \\
&\leq C
\end{aligned}$$

$$\mathbb{E} \left[\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \Phi_{1,i} \epsilon_{il}^{11} \middle| \mathbf{G}, \mathbf{S}, \mathbf{X} \right] = \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \Phi_{1,i} \mathbb{E} \left[\epsilon_{il}^{11} \middle| \mathbf{G}, \mathbf{S}, \mathbf{X} \right] = 0$$

On the other hand,

$$\begin{aligned}
Var\left(\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\Phi_{1,i}\epsilon_{il}^{11}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right) &= \frac{1}{n_{11}^2}Var\left(\sum_{i\in\mathcal{I}_{11}}\Phi_{1,i}\epsilon_{il}^{11}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\\
&= \frac{1}{n_{11}^2}\sum_{i\in\mathcal{I}_{11}}\Phi_{1,i}^2 Var\left(\epsilon_{il}^{11}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\\
&\leq \frac{C}{n_{11}^2}\sum_{i\in\mathcal{I}_{11}}\sigma_{11,l}^2(X_i)\\
&\leq \frac{C\overline{\sigma}^2}{n_{11}}\\
&= o_p(1)
\end{aligned}$$

By lemma 9,

$$\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\Phi_{1,i}\epsilon_{il}^{11} = o_p(1)$$

For $(g,s)\in\{(0,1),(1,0),(0,0)\}$,

$$\mathbb{E}\left[\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\Phi_{1,i}\sum_{j\in\mathcal{I}_{gs}}w_{ij}\epsilon_{jl}^{gs}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right] = \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\Phi_{1,i}\sum_{j\in\mathcal{I}_{gs}}w_{ij}\mathbb{E}\left[\epsilon_{jl}^{gs}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right] = 0$$

On the other hand

$$
\begin{aligned}
&Var\left(\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\Phi_{1,i}\sum_{j\in\mathcal{I}_{gs}}w_{ij}\epsilon_{jl}^{gs}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\\
&=\frac{1}{n_{11}^2}Var\left(\sum_{j\in\mathcal{I}_{gs}}\sum_{i\in\mathcal{I}_{11}}\Phi_{1,i}w_{ij}\epsilon_{jl}^{gs}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\\
&=\frac{1}{n_{11}^2}\sum_{j\in\mathcal{I}_{gs}}Var\left(\sum_{i\in\mathcal{I}_{11}}\Phi_{1,i}w_{ij}\epsilon_{jl}^{gs}\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\\
&=\frac{1}{n_{11}^2}\sum_{j\in\mathcal{I}_{gs}}\left(\sum_{i\in\mathcal{I}_{11}}\Phi_{1,i}w_{ij}\right)^2Var\left(\epsilon_{jl}^{gs}\middle|G_i=g,S_i=s,X_i\right)\\
&=\frac{C\overline{\sigma}^2}{n_{11}^2}\sum_{j\in\mathcal{I}_{gs}}\left(\sum_{i\in\mathcal{I}_{11}}w_{ij}\right)^2\\
&=\frac{C\overline{\sigma}^2}{n_{11}^2}\sum_{j\in\mathcal{I}_{gs}}w_j^2\\
&=\frac{n_{gs}C\overline{\sigma}^2}{n_{11}^2}\frac{1}{n_{gs}}\sum_{j\in\mathcal{I}_{gs}}w_j^2\\
&=\frac{1}{n_{11}}\frac{n_{gs}C\overline{\sigma}^2}{n_{11}}O(1)\\
&=o_p(1)
\end{aligned}
$$

By lemma 9,

$$
\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\Phi_{1,i}\sum_{j\in\mathcal{I}_{gs}}w_{ij}\epsilon_{jl}^{gs}=o_p(1)
$$

Therefore,

$$
\begin{aligned}
&\frac{2}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\mu_l^{11}(X_i)-\sum_{j\in\mathcal{I}_{01}}w_{ij}\mu_l^{01}(X_j)-\sum_{j\in\mathcal{I}_{10}}w_{ij}\mu_l^{10}(X_j)+\sum_{j\in\mathcal{I}_{00}}w_{ij}\mu_l^{00}(X_j)-\tau_l\right)\\
&\left(\epsilon_{il}^{11}-\sum_{j\in\mathcal{I}_{01}}w_{ij}\epsilon_{jl}^{01}-\sum_{j\in\mathcal{I}_{10}}w_{ij}\epsilon_{jl}^{10}+\sum_{j\in\mathcal{I}_{00}}w_{ij}\epsilon_{jl}^{00}\right)\\
&=\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\Phi_{1,i}\left(\epsilon_{il}^{11}-\sum_{j\in\mathcal{I}_{01}}w_{ij}\epsilon_{jl}^{01}-\sum_{j\in\mathcal{I}_{10}}w_{ij}\epsilon_{jl}^{10}+\sum_{j\in\mathcal{I}_{00}}w_{ij}\epsilon_{jl}^{00}\right)\\
&=o_p(1)
\end{aligned}
$$

Combine the results from the three terms, we have:

$$\begin{aligned}
\widehat{V}_1 &= \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\Delta Y_{il} - \widehat{\Delta Y_{il}^{01}} - \widehat{\Delta Y_{il}^{10}} + \widehat{\Delta Y_{il}^{00}} - \tau_l \right)^2 + o_p(1) \\
&= V_1 + \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\sigma_{11,l}^2(X_i) + \sum_{j \in \mathcal{I}_{01}} w_{ij}^2 \sigma_{01,l}^2(X_j) + \sum_{j \in \mathcal{I}_{10}} w_{ij}^2 \sigma_{10,l}^2(X_j) + \sum_{j \in \mathcal{I}_{00}} w_{ij}^2 \sigma_{00,l}^2(X_j) \right) + o_p(1)
\end{aligned}$$

Next, for $(g, s) \in \{(0,1), (1,0), (0,0)\}$,

$$\begin{aligned}
\widehat{\sigma}_{gs,l}^2(X_j) &= \frac{K}{K+1} \left(\Delta Y_{jl} - \frac{1}{K} \sum_{m=1}^{K} \Delta Y_{k_{m(j)} l} \right)^2 \\
&= \frac{K}{K+1} \left(\Delta Y_{jl} - \mu_l^{gs}(X_j) - \frac{1}{K} \sum_{m=1}^{K} \left[\Delta Y_{k_{m(j)} l} - \mu_l^{gs}(X_{k_{m(j)}}) \right] + \frac{1}{K} \sum_{m=1}^{K} \left[\mu_l^{gs}(X_j) - \mu_l^{gs}(X_{k_{m(j)}}) \right] \right)^2 \\
&= \frac{K}{K+1} \left(\epsilon_{jl}^{gs} - \frac{1}{K} \sum_{m=1}^{K} \epsilon_{k_{m(j)} l}^{gs} + \psi_{K,j} \right)^2
\end{aligned}$$

where $\psi_{K,j} = \frac{1}{K} \sum_{m=1}^{K} \left[\mu_l^{gs}(X_j) - \mu_l^{gs}(X_{k_{m(j)}}) \right]$

$$\begin{aligned}
\mathbb{E}\left[\widehat{\sigma}_{gs,l}^2(X_j) \middle| \mathbf{G}, \mathbf{S}, \mathbf{X} \right] &= \frac{K}{K+1} \mathbb{E}\left[\left(\epsilon_{jl}^{gs} - \frac{1}{K} \sum_{m=1}^{K} \epsilon_{k_{m(j)} l}^{gs} + \psi_{K,j} \right)^2 \middle| \mathbf{G}, \mathbf{S}, \mathbf{X} \right] \\
&= \frac{K}{K+1} \left(\mathbb{E}\left[\left(\epsilon_{jl}^{gs} \right)^2 \middle| \mathbf{G}, \mathbf{S}, \mathbf{X} \right] + \frac{1}{K^2} \sum_{m=1}^{K} \mathbb{E}\left[\left(\epsilon_{k_{m(j)} l}^{gs} \right)^2 \middle| \mathbf{G}, \mathbf{S}, \mathbf{X} \right] + \psi_{K,j}^2 \right) \\
&= \frac{K}{K+1} \left(\sigma_{gs,l}^2(X_j) + \frac{1}{K^2} \sum_{m=1}^{K} \sigma_{gs,l}^2(X_{k_m(j)}) + \psi_{K,j}^2 \right) \\
&= \sigma_{gs,l}^2(X_j) + \frac{1}{K+1} \sum_{m=1}^{K} \left(\sigma_{gs,l}^2(X_{k_m(j)}) - \sigma_{gs,l}^2(X_j) \right) + \frac{K}{K+1} \psi_{K,j}^2
\end{aligned}$$

$$\begin{aligned}
|\psi_{K,j}| = \left| \frac{1}{K} \sum_{m=1}^{K} \left[\mu_l^{gs}(X_j) - \mu_l^{gs}(X_{k_{m(j)}}) \right] \right| &\leq \frac{1}{K} \sum_{m=1}^{K} \left| \left[\mu_l^{gs}(X_j) - \mu_l^{gs}(X_{k_{m(j)}}) \right] \right| \\
\leq \frac{C}{K} \sum_{m=1}^{K} \left\| X_j - X_{k_{m(j)}} \right\| &\leq C \cdot \max_{r \in \mathcal{I}_{gs}} \left\| X_r - X_{k_{m(r)}} \right\| = o_p(1)
\end{aligned}$$

The last equality follows from lemma 11.

$$
\begin{aligned}
&\left|\frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\left(\mathbb{E}\left[\widehat{\sigma}^2_{gs,l}(X_j)\middle|\,\mathbf{G},\mathbf{S},\mathbf{X}\right]-\sigma^2_{gs,l}(X_j)\right)\right|\\
&=\left|\frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\left(\frac{1}{K+1}\sum_{m=1}^{K}\left(\sigma^2_{gs,l}(X_{k_m(j)})-\sigma^2_{gs,l}(X_j)\right)+\frac{K}{K+1}\psi^2_{K,j}\right)\right|\\
&=\left|\frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\left(\frac{1}{K+1}\sum_{m=1}^{K}\left(\sigma^2_{gs,l}(X_{k_m(j)})-\sigma^2_{gs,l}(X_j)\right)\right)\right|+\left|\frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\left(\frac{K}{K+1}\psi^2_{K,j}\right)\right|\\
&\le\frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\left(\frac{1}{K+1}\sum_{m=1}^{K}\left|\left(\sigma^2_{gs,l}(X_{k_m(j)})-\sigma^2_{gs,l}(X_j)\right)\right|\right)+\frac{1}{n_{11}}\frac{K}{K+1}\sum_{j\in\mathcal{I}_{gs}} w_j^2\left|\psi_{K,j}\right|^2\\
&\le\frac{C}{n_{11}}\frac{1}{K+1}\sum_{j\in\mathcal{I}_{gs}} w_j^2\left(\sum_{m=1}^{K}\left\|X_j-X_{k_{m(j)}}\right\|\right)+\frac{CKn_{gs}}{(K+1)n_{11}}\left(\max_{r\in\mathcal{I}_{gs}}\left\|X_r-X_{k_{m(r)}}\right\|\right)^2\frac{1}{n_{gs}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\\
&=\frac{CKn_{gs}}{(K+1)n_{11}}\max_{r\in\mathcal{I}_{gs}}\left\|X_r-X_{k_{m(r)}}\right\|\frac{1}{n_{gs}}\sum_{j\in\mathcal{I}_{gs}} w_j^2+o_p(1)\\
&=o_p(1)
\end{aligned}
$$

This implies that,

$$
\frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\left(\mathbb{E}\left[\widehat{\sigma}^2_{gs,l}(X_j)\middle|\,\mathbf{G},\mathbf{S},\mathbf{X}\right]-\sigma^2_{gs,l}(X_j)\right)=o_p(1)
$$

Since the support of X is compact so $|\psi_{K,j}|$ is uniformly bounded.

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\left(\widehat{\sigma}^2_{gs,l}(X_j)-\mathbb{E}\left[\widehat{\sigma}^2_{gs,l}(X_j)\middle|\,\mathbf{G},\mathbf{S},\mathbf{X}\right]\right)\middle|\,\mathbf{G},\mathbf{S},\mathbf{X}\right]\\
&=\frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\left(\mathbb{E}\left[\widehat{\sigma}^2_{gs,l}(X_j)\middle|\,\mathbf{G},\mathbf{S},\mathbf{X}\right]-\mathbb{E}\left[\widehat{\sigma}^2_{gs,l}(X_j)\middle|\,\mathbf{G},\mathbf{S},\mathbf{X}\right]\right)\\
&=0
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
&Var\left(\frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\left(\widehat{\sigma}^2_{gs,l}(X_j)-\mathbb{E}\left[\widehat{\sigma}^2_{gs,l}(X_j)\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right]\right)\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\\
&=Var\left(\frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\widehat{\sigma}^2_{gs,l}(X_j)\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\\
&=\frac{K^2}{n_{11}^2(K+1)^2}Var\left(\sum_{j\in\mathcal{I}_{gs}} w_j^2\left(\epsilon^{gs}_{jl}-\frac{1}{K}\sum_{m=1}^{K}\epsilon^{gs}_{k_{m(j)}l}+\psi_{K,j}\right)^2\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\\
&=\frac{K^2}{n_{11}^2(K+1)^2}\left(\sum_{j\in\mathcal{I}_{gs}} Var\left(w_j^2\left(\epsilon^{gs}_{jl}-\frac{1}{K}\sum_{m=1}^{K}\epsilon^{gs}_{k_{m(j)}l}+\psi_{K,j}\right)^2\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\right.\\
&\left.+\sum_{j\in\mathcal{I}_{gs}}\sum_{\substack{j'\in\mathcal{I}_{gs}\\ j'\neq j}} Cov\left(w_j^2\left(\epsilon^{gs}_{jl}-\frac{1}{K}\sum_{m=1}^{K}\epsilon^{gs}_{k_{m(j)}l}+\psi_{K,j}\right)^2, w_{j'}^2\left(\epsilon^{gs}_{j'l}-\frac{1}{K}\sum_{m=1}^{K}\epsilon^{gs}_{k_{m(j')}l}+\psi_{K,j'}\right)^2\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\right)
\end{aligned}
$$

Now for the variance term,

$$
\begin{aligned}
&\sum_{j\in\mathcal{I}_{gs}} Var\left(w_j^2\left(\epsilon^{gs}_{jl}-\frac{1}{K}\sum_{m=1}^{K}\epsilon^{gs}_{k_{m(j)}l}+\psi_{K,j}\right)^2\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\\
&\leq\sum_{j\in\mathcal{I}_{gs}} w_j^4\mathbb{E}\left[\left(\epsilon^{gs}_{jl}-\frac{1}{K}\sum_{m=1}^{K}\epsilon^{gs}_{k_{m(j)}l}+\psi_{K,j}\right)^4\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right]\\
&\leq 27\sum_{j\in\mathcal{I}_{gs}} w_j^4\left(\mathbb{E}\left[\left(\epsilon^{gs}_{jl}\right)^4\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right]+\mathbb{E}\left[\left(\frac{1}{K}\sum_{m=1}^{K}\epsilon^{gs}_{k_{m(j)}l}\right)^4\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right]+\mathbb{E}\left[\left(\psi_{K,j}\right)^4\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right]\right)\\
&\leq 27\sum_{j\in\mathcal{I}_{gs}} w_j^4\left(\mathbb{E}\left[\left(\epsilon^{gs}_{jl}\right)^4\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right]+\frac{1}{K}\sum_{m=1}^{K}\mathbb{E}\left[\left(\epsilon^{gs}_{k_{m(j)}l}\right)^4\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right]+\mathbb{E}\left[\left(\psi_{K,j}\right)^4\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right]\right)\\
&\leq 27C\sum_{j\in\mathcal{I}_{gs}} w_j^4
\end{aligned}
$$

The third inequality follows from Jensen's inequality.

$$
\begin{aligned}
&\frac{K^2}{n_{11}^2(K+1)^2}\sum_{j\in\mathcal{I}_{gs}} Var\left(w_j^2\left(\epsilon_{jl}^{gs}-\frac{1}{K}\sum_{m=1}^{K}\epsilon_{k_{m(j)}l}^{gs}+\psi_{K,j}\right)^2\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\\
&\leq \frac{27n_{gs}CK^2}{n_{11}^2(K+1)^2}\frac{1}{n_{gs}}\sum_{j\in\mathcal{I}_{gs}} w_j^4\\
&= o_p(1)
\end{aligned}
$$

For the covariance term, by Cauchy Schwarz inequality and the derivation above

$$
\begin{aligned}
&\left|Cov\left(w_j^2\left(\epsilon_{jl}^{gs}-\frac{1}{K}\sum_{m=1}^{K}\epsilon_{k_{m(j)}l}^{gs}+\psi_{K,j}\right)^2, w_{j'}^2\left(\epsilon_{j'l}^{gs}-\frac{1}{K}\sum_{m=1}^{K}\epsilon_{k_{m(j')}l}^{gs}+\psi_{K,j'}\right)^2\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\right|\\
&\leq \sqrt{Var\left(w_j^2\left(\epsilon_{jl}^{gs}-\frac{1}{K}\sum_{m=1}^{K}\epsilon_{k_{m(j)}l}^{gs}+\psi_{K,j}\right)^2\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)}\sqrt{Var\left(w_{j'}^2\left(\epsilon_{j'l}^{gs}-\frac{1}{K}\sum_{m=1}^{K}\epsilon_{k_{m(j')}l}^{gs}+\psi_{K,j'}\right)^2\middle|\mathbf{G},\mathbf{S},\right.}\\
&\leq Cw_j^2w_{j'}^2
\end{aligned}
$$

On the other hand, recall $\mathcal{K}_K(j)$ and $\mathcal{K}_K(j')$ are the within-subgroup matched sets of unit j, and j', define $\overline{\mathcal{K}}_K(j)=\{j\}\cup K_K(j)$, and $\overline{\mathcal{K}}_K(j')=\{j'\}\cup K_K(j')$ if $\overline{\mathcal{K}}_K(j)\cap\overline{\mathcal{K}}_K(j')=\phi$, then notice that the two terms in the covariance are independent after conditioning on $(\mathbf{G},\mathbf{S},\mathbf{X})$, which implies that the covariance is none zero if and only if $\overline{\mathcal{K}}_K(j)\cap\overline{\mathcal{K}}_K(j')\neq\phi$.

Therefore,

$$\left|\sum_{j\in\mathcal{I}_{gs}}\sum_{\substack{j'\in\mathcal{I}_{gs}\\ j'\neq j}} Cov\left(w_j^2\left(\epsilon_{jl}^{gs}-\frac{1}{K}\sum_{m=1}^{K}\epsilon_{k_{m(j)}l}^{gs}+\psi_{K,j}\right)^2, w_{j'}^2\left(\epsilon_{j'l}^{gs}-\frac{1}{K}\sum_{m=1}^{K}\epsilon_{k_{m(j')}l}^{gs}+\psi_{K,j'}\right)^2\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\right|$$

$$=\left|\sum_{j\in\mathcal{I}_{gs}}\sum_{\substack{j'\in\mathcal{I}_{gs}\\ j'\neq j}} Cov\left(w_j^2\left(\epsilon_{jl}^{gs}-\frac{1}{K}\sum_{m=1}^{K}\epsilon_{k_{m(j)}l}^{gs}+\psi_{K,j}\right)^2, w_{j'}^2\left(\epsilon_{j'l}^{gs}-\frac{1}{K}\sum_{m=1}^{K}\epsilon_{k_{m(j')}l}^{gs}+\psi_{K,j'}\right)^2\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\right.$$
$$\left.\mathbb{1}(\overline{\mathcal{K}}_K(j)\cap\overline{\mathcal{K}}_K(j')\neq\phi)\right|$$

$$\leq\sum_{j\in\mathcal{I}_{gs}}\sum_{\substack{j'\in\mathcal{I}_{gs}\\ j'\neq j}}\left|Cov\left(w_j^2\left(\epsilon_{jl}^{gs}-\frac{1}{K}\sum_{m=1}^{K}\epsilon_{k_{m(j)}l}^{gs}+\psi_{K,j}\right)^2, w_{j'}^2\left(\epsilon_{j'l}^{gs}-\frac{1}{K}\sum_{m=1}^{K}\epsilon_{k_{m(j')}l}^{gs}+\psi_{K,j'}\right)^2\middle|\mathbf{G},\mathbf{S},\mathbf{X}\right)\right|$$
$$\mathbb{1}(\overline{\mathcal{K}}_K(j)\cap\overline{\mathcal{K}}_K(j')\neq\phi)$$

$$\leq C\sum_{j\in\mathcal{I}_{gs}}\sum_{\substack{j'\in\mathcal{I}_{gs}\\ j'\neq j}} w_j^2 w_{j'}^2\mathbb{1}(\overline{\mathcal{K}}_K(j)\cap\overline{\mathcal{K}}_K(j')\neq\phi)$$

$$\leq C\sum_{j\in\mathcal{I}_{gs}}\sum_{\substack{j'\in\mathcal{I}_{gs}\\ j'\neq j}}\frac{w_j^4+w_{j'}^4}{2}\mathbb{1}(\overline{\mathcal{K}}_K(j)\cap\overline{\mathcal{K}}_K(j')\neq\phi)$$

$$\leq\frac{C}{2}\sum_{j\in\mathcal{I}_{gs}}\sum_{j'\in\mathcal{I}_{gs}} w_j^4\mathbb{1}(\overline{\mathcal{K}}_K(j)\cap\overline{\mathcal{K}}_K(j')\neq\phi)+\frac{C}{2}\sum_{j'\in\mathcal{I}_{gs}}\sum_{j\in\mathcal{I}_{gs}} w_{j'}^4\mathbb{1}(\overline{\mathcal{K}}_K(j')\cap\overline{\mathcal{K}}_K(j)\neq\phi)$$

$$=C\sum_{j\in\mathcal{I}_{gs}}\sum_{j'\in\mathcal{I}_{gs}} w_j^4\mathbb{1}(\overline{\mathcal{K}}_K(j)\cap\overline{\mathcal{K}}_K(j')\neq\phi)$$

$$=C\sum_{j\in\mathcal{I}_{gs}} w_j^4\sum_{j'\in\mathcal{I}_{gs}}\mathbb{1}(\overline{\mathcal{K}}_K(j)\cap\overline{\mathcal{K}}_K(j')\neq\phi)$$

Notice that $\sum_{j'\in\mathcal{I}_{gs}}\mathbb{1}(\overline{\mathcal{K}}_K(j)\cap\overline{\mathcal{K}}_K(j')\neq\phi)$ is the total number of the sets that has a nonempty intersection with $\overline{\mathcal{K}}_K(i)$. This is bounded by,

$$\sum_{j'\in\mathcal{I}_{gs}}\mathbb{1}(\overline{\mathcal{K}}_K(j)\cap\overline{\mathcal{K}}_K(j')\neq\phi)\leq\sum_{r\in\overline{\mathcal{K}}_K(j)} K_K(r)\leq\sum_{r\in\overline{\mathcal{K}}_K(j)}\max_{r\in\mathcal{I}_{gs}} K_K(r)=(J+1)\max_{r\in\mathcal{I}_{gs}} K_K(r)\leq(K+1)K\gamma_q$$

The last line follows from stone's lemma. (Add stone's lemma when I incorporate the result.)

Therefore,

$$
\begin{aligned}
&\frac{K^2}{n_{11}^2(K+1)^2}\left|\sum_{j\in\mathcal{I}_{gs}}\sum_{\substack{j'\in\mathcal{I}_{gs}\\ j'\neq j}} Cov\left(w_j^2\left(\epsilon_{jl}^{gs}-\frac{1}{K}\sum_{m=1}^{K}\epsilon_{k_{m(j)}l}^{gs}+\psi_{K,j}\right)^2, w_{j'}^2\left(\epsilon_{j'l}^{gs}-\frac{1}{K}\sum_{m=1}^{K}\epsilon_{k_{m(j')}l}^{gs}+\psi_{K,j'}\right)^2\right|\mathbf{G},\mathbf{S},\right.\\
&\leq \frac{CK^2}{n_{11}^2(K+1)^2}\sum_{j\in\mathcal{I}_{gs}} w_j^4 \sum_{j'\in\mathcal{I}_{gs}} \mathbb{1}(\overline{\mathcal{K}}_K(j)\cap\overline{\mathcal{K}}_K(j')\neq\phi)\\
&\leq \frac{C\gamma_d K^3 n_{gs}}{n_{11}^2(K+1)}\frac{1}{n_{gs}}\sum_{j\in\mathcal{I}_{gs}} w_j^4\\
&= o_p(1)
\end{aligned}
$$

So,

$$
Var\left(\frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\left(\widehat{\sigma}_{gs,l}^2(X_j)-\mathbb{E}\left[\widehat{\sigma}_{gs,l}^2(X_j)\middle|\,\mathbf{G},\mathbf{S},\mathbf{X}\right]\right)\middle|\,\mathbf{G},\mathbf{S},\mathbf{X}\right)=o(1)
$$

By lemma 9,

$$
\frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\left(\widehat{\sigma}_{gs,l}^2(X_j)-\mathbb{E}\left[\widehat{\sigma}_{gs,l}^2(X_j)\middle|\,\mathbf{G},\mathbf{S},\mathbf{X}\right]\right)=o_p(1)
$$

Therefore,

$$
\begin{aligned}
&\frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\widehat{\sigma}_{gs,l}^2(X_j)-\frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\sigma_{gs,l}^2(X_j)\\
&=\frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\left(\widehat{\sigma}_{gs,l}^2(X_j)-\sigma_{gs,l}^2(X_j)\right)\\
&=\frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\left(\widehat{\sigma}_{gs,l}^2(X_j)-\mathbb{E}\left[\widehat{\sigma}_{gs,l}^2(X_j)\middle|\,\mathbf{G},\mathbf{S},\mathbf{X}\right]+\mathbb{E}\left[\widehat{\sigma}_{gs,l}^2(X_j)\middle|\,\mathbf{G},\mathbf{S},\mathbf{X}\right]-\sigma_{gs,l}^2(X_j)\right)\\
&=\frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\left(\widehat{\sigma}_{gs,l}^2(X_j)-\mathbb{E}\left[\widehat{\sigma}_{gs,l}^2(X_j)\middle|\,\mathbf{G},\mathbf{S},\mathbf{X}\right]\right)+\frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} w_j^2\left(\mathbb{E}\left[\widehat{\sigma}_{gs,l}^2(X_j)\middle|\,\mathbf{G},\mathbf{S},\mathbf{X}\right]-\sigma_{gs,l}^2(X_j)\right)\\
&=o_p(1)
\end{aligned}
$$

Lastly, denote $\widetilde{w}_j^2 = \sum_{i\in\mathcal{I}_{11}} w_{ij}^2$, so

$$\widetilde{w}_j^2 = \sum_{i\in\mathcal{I}_{11}} w_{ij}^2 \leq \left(\sum_{i\in\mathcal{I}_{11}} w_{ij}\right)^2 = w_j^2$$

This implies that,

$$\frac{1}{n_{gs}}\sum_{j\in\mathcal{I}_{gs}} \widetilde{w}_j^2 = O(1), \quad \frac{1}{n_{gs}}\sum_{j\in\mathcal{I}_{gs}} \widetilde{w}_j^4 = O(1)$$

Also notice that

$$\begin{aligned}
\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}^2\widehat{\sigma}_{gs,l}^2(X_j) &= \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}}\sum_{i\in\mathcal{I}_{11}} w_{ij}^2\widehat{\sigma}_{gs,l}^2(X_j) = \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} \widetilde{w}_j^2\widehat{\sigma}_{gs,l}^2(X_j)\\
\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}^2\sigma_{gs,l}^2(X_j) &= \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}}\sum_{i\in\mathcal{I}_{11}} w_{ij}^2\sigma_{gs,l}^2(X_j) = \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{gs}} \widetilde{w}_j^2\sigma_{gs,l}^2(X_j)
\end{aligned}$$

Then following the same method as above, we could show that,

$$\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}^2\widehat{\sigma}_{gs,l}^2(X_j) - \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}^2\sigma_{gs,l}^2(X_j) = o_p(1)$$

This implies that,

$$\begin{aligned}
\widehat{V}_2 &= \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{01}}\left(w_j^2 - \widetilde{w}_j^2\right)\widehat{\sigma}_{01,l}^2(X_j) + \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{10}}\left(w_j^2 - \widetilde{w}_j^2\right)\widehat{\sigma}_{10,l}^2(X_j) + \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{00}}\left(w_j^2 - \widetilde{w}_j^2\right)\widehat{\sigma}_{00,l}^2(X_j)\\
&= \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{01}}\left(w_j^2 - \widetilde{w}_j^2\right)\sigma_{01,l}^2(X_j) + \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{10}}\left(w_j^2 - \widetilde{w}_j^2\right)\sigma_{10,l}^2(X_j) + \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{00}}\left(w_j^2 - \widetilde{w}_j^2\right)\sigma_{00,l}^2(X_j) + o_p(1)
\end{aligned}$$

Combining all the results above,

$$\begin{aligned}
&\widehat{V}_1 + \widehat{V}_2\\
&= \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\Delta Y_{il} - \widehat{\Delta Y_{il}^{01}} - \widehat{\Delta Y_{il}^{10}} + \widehat{\Delta Y_{il}^{00}} - \tau_l\right)^2 + o_p(1)\\
&= V_1 + \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sigma_{11,l}^2(X_i) + \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{01}}\widetilde{w}_j^2\sigma_{01,l}^2(X_j) + \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{10}}\widetilde{w}_j^2\sigma_{10,l}^2(X_j) + \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{00}}\widetilde{w}_j^2\sigma_{00,l}^2(X_j)\\
&+ \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{01}}\left(w_j^2 - \widetilde{w}_j^2\right)\sigma_{01,l}^2(X_j) + \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{10}}\left(w_j^2 - \widetilde{w}_j^2\right)\sigma_{10,l}^2(X_j) + \frac{1}{n_{11}}\sum_{j\in\mathcal{I}_{00}}\left(w_j^2 - \widetilde{w}_j^2\right)\sigma_{00,l}^2(X_j) + o_p(1)\\
&= V_1 + V_2 + o_p(1)
\end{aligned}$$

This completes the proof. ■

Proof. Proof of theorem 4.

For any $(g,s) \in \{(0,1),(1,0),(0,0)\}$,

$$
\begin{aligned}
&\left|\frac{1}{n_{11}}\sum_{i=1}^{n}\left[\mathbb{1}(i\in\mathcal{I}_{11})-\mathbb{1}(i\in\mathcal{I}_{gs})w_i\right]\left(\mu_l^{gs}(X_i)-\widehat{\mu}_l^{gs}(X_i)\right)\right| \\
&=\left|\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left[\mu_l^{gs}(X_i)-\mu_l^{gs}(X_j)-\left(\widehat{\mu}_l^{gs}(X_i)-\widehat{\mu}_l^{gs}(X_j)\right)\right]\right| \\
&\leq \frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left|\mu_l^{gs}(X_i)-\mu_l^{gs}(X_j)-\left(\widehat{\mu}_l^{gs}(X_i)-\widehat{\mu}_l^{gs}(X_j)\right)\right|
\end{aligned}
$$

The first equality follows from the same logic as the proof of lemma 13.

By Taylor expansion to $k-1$-th order,

$$
\begin{aligned}
\left|\mu_l^{gs}(X_j)-\mu_l^{gs}(X_i)-\sum_{r=1}^{p-1}\frac{1}{r!}\sum_{m\in\Lambda_r}\partial^m\mu_l^{gs}(X_i)(X_j-X_i)^m\right| &\leq \max_{m\in\Lambda_p}\left\|\partial^m\mu_l^{gs}(x)\right\|_\infty\frac{1}{p!}\sum_{m\in\Lambda_p}\left\|X_j-X_i\right\|^p \\
&= O(1)\left\|X_j-X_i\right\|^p
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\left|\widehat{\mu}_l^{gs}(X_j)-\widehat{\mu}_l^{gs}(X_i)-\sum_{r=1}^{p-1}\frac{1}{r!}\sum_{m\in\Lambda_r}\partial^m\widehat{\mu}_l^{gs}(X_i)(X_j-X_i)^m\right| &\leq \max_{m\in\Lambda_p}\left\|\partial^m\widehat{\mu}_l^{gs}(x)\right\|_\infty\frac{1}{p!}\sum_{m\in\Lambda_p}\left\|X_j-X_i\right\|^p \\
&= O_p(1)\left\|X_j-X_i\right\|^p
\end{aligned}
$$

Lastly,

$$
\begin{aligned}
&\left|\sum_{r=1}^{p-1}\frac{1}{r!}\sum_{m\in\Lambda_r}\left(\partial^m\widehat{\mu}_l^{gs}(X_i)-\partial^m\mu_l^{gs}(X_i)\right)(X_j-X_i)^m\right| \\
&\leq \sum_{r=1}^{p-1}\max_{t\in\Lambda_r}\left\|\partial^m\widehat{\mu}^{gs}(x)-\partial^m\mu^{gs}(x)\right\|_\infty\frac{1}{r!}\sum_{m\in\Lambda_r}\left\|X_j-X_i\right\|^r \\
&\leq \sum_{r=1}^{p-1}O_p(-\gamma_r)\left\|X_j-X_i\right\|^r
\end{aligned}
$$

Now,

$$
\begin{aligned}
&\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \sum_{j \in \mathcal{I}_{gs}} w_{ij} \left| \mu_l^{gs}(X_i) - \mu_l^{gs}(X_j) - (\widehat{\mu}_l^{gs}(X_i) - \widehat{\mu}_l^{gs}(X_j)) \right| \\
&\leq \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \sum_{j \in \mathcal{I}_{gs}} w_{ij} \left| \mu_l^{gs}(X_j) - \mu_l^{gs}(X_i) - \sum_{r=1}^{p-1} \frac{1}{r!} \sum_{m \in \Lambda_r} \partial^m \mu_l^{gs}(X_i)(X_j - X_i)^m \right| \\
&+ \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \sum_{j \in \mathcal{I}_{gs}} w_{ij} \left| \widehat{\mu}_l^{gs}(X_j) - \widehat{\mu}_l^{gs}(X_i) - \sum_{r=1}^{p-1} \frac{1}{r!} \sum_{m \in \Lambda_r} \partial^m \widehat{\mu}_l^{gs}(X_i)(X_j - X_i)^m \right| \\
&+ \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \sum_{j \in \mathcal{I}_{gs}} w_{ij} \left| \sum_{r=1}^{p-1} \frac{1}{r!} \sum_{m \in \Lambda_r} \left(\partial^m \widehat{\mu}_l^{gs}(X_i) - \partial^m \mu_l^{gs}(X_i) \right) (X_j - X_i)^m \right| \\
&\leq O_p(1) \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \sum_{j \in \mathcal{I}_{gs}} w_{ij} \left\| X_j - X_i \right\|^p + O_p(1) \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \sum_{j \in \mathcal{I}_{gs}} w_{ij} \left\| X_j - X_i \right\|^p \\
&+ \sum_{r=1}^{p-1} O_p(-\gamma_r) \left(\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \sum_{j \in \mathcal{I}_{gs}} w_{ij} \left\| X_j - X_i \right\|^r \right) \\
&= o_p(n^{-1/2}) + o_p(n^{-1/2}) + \sum_{r=1}^{p-1} O_p(-\gamma_r) o_p(n^{-1/2} + \gamma_r) \\
&= o_p(n^{-1/2})
\end{aligned}
$$

The first equality follows from assumption 9 and the similar argument as the proof of lemma 7(iv).
Therefore,

$$
\begin{aligned}
\left| B - \widehat{B} \right| \leq & \left| \frac{1}{n_{11}} \sum_{i=1}^{n} \left[\mathbb{1}(i \in \mathcal{I}_{11}) - \mathbb{1}(i \in \mathcal{I}_{gs}) w_i \right] \left(\mu_l^{01}(X_i) - \widehat{\mu}_l^{01}(X_i) \right) \right| \\
& + \left| \frac{1}{n_{11}} \sum_{i=1}^{n} \left[\mathbb{1}(i \in \mathcal{I}_{11}) - \mathbb{1}(i \in \mathcal{I}_{gs}) w_i \right] \left(\mu_l^{10}(X_i) - \widehat{\mu}_l^{10}(X_i) \right) \right| \\
& + \left| \frac{1}{n_{11}} \sum_{i=1}^{n} \left[\mathbb{1}(i \in \mathcal{I}_{11}) - \mathbb{1}(i \in \mathcal{I}_{gs}) w_i \right] \left(\mu_l^{00}(X_i) - \widehat{\mu}_l^{00}(X_i) \right) \right| \\
& = o_p(n^{-1/2})
\end{aligned}
$$

By Slutsky theorem, this shows that $\sqrt{n_{11}}(B - \widehat{B}) \xrightarrow{p} 0$.

Lastly,

$$
\begin{aligned}
&(V_1+V_2)^{-1/2}\sqrt{n_{11}}\left(\widehat{\tau}_l-\tau_l-\widehat{B}\right)\\
&=(V_1+V_2)^{-1/2}\sqrt{n_{11}}\left(\widehat{\tau}_l-\tau_l-B\right)+(V_1+V_2)^{-1/2}\sqrt{n_{11}}(B-\widehat{B})\\
&\xrightarrow{d}\mathcal{N}(0,1)
\end{aligned}
$$

The convergence result follows by the argument above and theorem 2. This completes the proof. ∎

F Proof of Lemmas

Proof. Proof of lemma 1. For each $(g,s)\in\{(0,1),(1,0),(0,0)\}$, we have:

$$
\begin{aligned}
\mathbb{E}[\mathbb{E}[\Delta Y_l\mid G=g,S=s,X]\mid G=1,S=1]&=\int_{\mathcal{X}}\mathbb{E}[\Delta Y_l\mid G=g,S=s,X=x]f_{11}(x)dx\\
&=\int_{\mathcal{X}}\mathbb{E}[\Delta Y_l\mid G=g,S=s,X=x]\frac{f_{11}(x)}{f_{gs}(x)}f_{gs}(x)dx\\
&=\mathbb{E}\left[\mathbb{E}[\Delta Y_l\mid G=g,S=s,X]\frac{f_{11}(X)}{f_{gs}(X)}\middle|G=g,S=s\right]\\
&=\mathbb{E}\left[\mathbb{E}\left[\frac{f_{11}(X)}{f_{gs}(X)}\Delta Y_l\middle|G=g,S=s,X\right]\middle|G=g,S=s\right]\\
&=\mathbb{E}\left[\frac{f_{11}(X)}{f_{gs}(X)}\Delta Y_l\middle|G=g,S=s\right]
\end{aligned}
$$

Plug this into the right hand side of theorem 1 yields the desired result. ∎

Proof. Proof of lemma 2.
To prove the first equation, notice that the first specification can be written as:

$$
\begin{aligned}
&Y_{it}=\sum_{g,s}\alpha_{gs}\mathbb{1}(G_i=g,S_i=s)+\sum_{g}\alpha_{gt}\mathbb{1}(G_i=g)+\sum_{s}\gamma_{st}\mathbb{1}(S_i=s)+\sum_{l\neq-1}\tau_lD_i\mathbb{1}(t-t^*=l)+X_i'\theta+u_{it}\\
&\implies\Delta Y_{il}=(\alpha_{0,t^*+l-1}-\alpha_{0,t^*}+\gamma_{0,t^*+l-1}-\gamma_{0,t^*})+(\alpha_{0,t^*+l-1}-\alpha_{0,t^*})G_i+(\gamma_{0,t^*+l-1}-\gamma_{0,t^*})S_i+\tau_lD_i\\
&+(u_{i,t^*+l-1}-u_{i,t^*})
\end{aligned}
$$

So $\widehat{\tau}_l^1$ can be obtained by regressing ΔY_{il} on the constant and (G_i,S_i,D_i). Therefore, its corresponding estimand τ_l^1 is the coefficient of D of the population linear projection of ΔY_l on $1,G,S,D$, which is a fully saturated regression. Therefore, it is straight forward to solve that,

$$\tau_l^1 = \mathbb{E}[\Delta Y_l|G=1,S=1] - \mathbb{E}[\Delta Y_l|G=0,S=1] - \mathbb{E}[\Delta Y_l|G=1,S=0] + \mathbb{E}[\Delta Y_l|G=0,S=0]$$

To prove the second equation, notice by the similar technique as above, we can show that $\widehat{\tau}_l^2$ can be obtained by regressing ΔY_{il} on the constant and (G_i, S_i, D_i, X_i). Therefore, its corresponding estimand τ_l^2 is the coefficient of D of the population linear projection of ΔY_l on $1, G, S, D, X$. Apply the FWL theorem, we have:

$$\begin{aligned}
\tau_l^2 &= \frac{\mathbb{E}[(D - L(\Delta Y_l|G,S,X))\Delta Y]}{\mathbb{E}[(D - L(\Delta Y_l|G,S,X))^2]} \\
&= \mathbb{E}\left[\frac{p_{11}\left(1 - L(D|G,S,X)\right)}{\mathbb{E}[(D-L(D|G,S,X))^2]}\Delta Y_l \middle| G=1,S=1\right] - \mathbb{E}\left[\frac{p_{01}L(D|G,S,X)}{\mathbb{E}[(D-L(D|G,S,X))^2]}\Delta Y_l \middle| G=0,S=1\right] \\
&\quad - \mathbb{E}\left[\frac{p_{10}L(D|G,S,X)}{\mathbb{E}[(D-L(D|G,S,X))^2]}\Delta Y_l \middle| G=1,S=0\right] + \mathbb{E}\left[\frac{-p_{00}L(D|G,S,X)}{\mathbb{E}[(D-L(D|G,S,X))^2]}\Delta Y_l \middle| G=0,S=0\right]
\end{aligned}$$

The second equality follows from law of iterated expectation. ■

Proof. Proof of lemma 3.

Proof of part (a): By construction of the nearest neighbor matching algorithm, w_{ij}^{fix} and w_j^{fix} satisfy assumption 5 (i). To show (ii), conditional on the sample size $n_{g's'}$, notice that under nearest neighbor matching,

$$\begin{aligned}
&\mathbb{E}\left[\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left\|X_j - X_i\right\| \middle| i \in \mathcal{I}_{11}\right] \\
&= \mathbb{E}\left[\sum_{m=1}^{M}\frac{1}{M}\left\|X_i - X_{j_m(i,g,s)}\right\| \middle| i \in \mathcal{I}_{11}\right] \\
&\leq \mathbb{E}\left[\left\|X_i - X_{j_M(i,g,s)}\right\| \mid i \in \mathcal{I}_{11}\right] \\
&= n_{gs}^{-1/q}\mathbb{E}\left[n_{gs}^{1/q}\left\|X_i - X_{j_M(i,g,s)}\right\| \mid i \in \mathcal{I}_{11}\right] \\
&\leq n_{gs}^{-1/q}C \\
&= o(1)
\end{aligned}$$

The last inequality follows lemma 2 in Abadie and Imbens (2006).

To show (iii), we can also conduct the other pairwise matching by reversing the role of reference group and the comparison group in the original pairwise matching. Then each unit $j \in \mathcal{I}_{11} \cup \mathcal{I}_{gs}$ gets a matching weight w_j, then under our assumption, we can invoke lemma

3(i) in Abadie and Imbens (2006) to get that $\mathbb{E}[w_j^{4+\delta}|j \in \mathcal{I}_{11} \cup \mathcal{I}_{gs}] = O(1)$. Therefore,

$$\begin{aligned}
\mathbb{E}[w_j^{4+\delta}|j \in \mathcal{I}_{gs}] &= \mathbb{E}[w_j^{4+\delta}\mathbb{1}(j \in \mathcal{I}_{gs})|i \in \mathcal{I}_{11} \cup \mathcal{I}_{gs}]\frac{p_{11}+p_{gs}}{p_{gs}} \\
&\leq \mathbb{E}[w_j^{4+\delta}|i \in \mathcal{I}_{11} \cup \mathcal{I}_{gs}]\frac{p_{11}+p_{gs}}{p_{gs}} \\
&= O(1)
\end{aligned}$$

(iv) is shown in lemma 10.

Proof of part (b): By construction of the nearest neighbor matching algorithm, $w_{ij}^{diverge}$ and $w_j^{diverge}$ satisfy assumption 5 (i).

To show (ii), conditional on the sample size n_{gs}, we have:

$$\begin{aligned}
&\mathbb{E}\left[\sum_{j \in \mathcal{I}_{gs}} w_{ij} \left\|X_j - X_i\right\| \middle| i \in \mathcal{I}_{11}\right] \\
&= \mathbb{E}\left[\sum_{m=1}^{M} \frac{1}{M} \left\|X_i - X_{j_m(i,g,s)}\right\| \middle| i \in \mathcal{I}_{11}\right] \\
&\leq \mathbb{E}\left[\left\|X_i - X_{j_M(i,g,s)}\right\| \,\middle|\, i \in \mathcal{I}_{11}\right] \\
&\leq C\left(\frac{M}{n_{gs}}\right)^{1/d} \\
&= o(1)
\end{aligned}$$

The second inequality follows from lemma $C.2$ in Lin et al. (2023). The last line follows from assumption 7.

To show (iii), we still conduct the other pairwise matching by reversing the role of reference group and the comparison group in the original pairwise matching. Then each unit $j \in \mathcal{I}_{11} \cup \mathcal{I}_{gs}$ gets a matching weight w_j. Conditional on the sample size n_{11}, and n_{gs}, by theorem $B.2$ in Lin et al. (2023), we have:

$$\mathbb{E}\left[\left|\frac{n_{gs}}{n_{11}} w_j - \frac{f_{11}(X)}{f_{gs}(X)}\right|^{4+\delta}\right] \to 0$$

Use the C-r inequality, we have:

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{n_{gs}}{n_{11}}w_j\right)^{4+\delta}\right] &\leq 2^{3+\delta}\mathbb{E}\left[\left|\frac{n_{gs}}{n_{11}}w_j - \frac{f_{11}(X)}{f_{gs}(X)}\right|^{4+\delta}\right] + \mathbb{E}\left[\left(\frac{f_{11}(X)}{f_{gs}(X)}\right)^{4+\delta}\right] \\
&\leq o(1) + C \\
&= O(1)
\end{aligned}
$$

The inequality follows from assumption 6 (i). Therefore,

$$
\mathbb{E}[w_j^{4+\delta}] = \left(\frac{n_{11}}{n_{gs}}\right)^{4+\delta}\mathbb{E}\left[\left(\frac{n_{gs}}{n_{11}}w_j\right)^{4+\delta}\right] = O(1)
$$

Then by the similar argument as the proof in part (a), we can show that

$$
\mathbb{E}[w_j^{4+\delta}|j \in \mathcal{I}_{gs}] = O(1)
$$

(iv) is shown in lemma 10.

Proof of part (c): By construction of the nearest neighbor matching algorithm, w_{ij}^{kernel} and w_j^{kernel} satisfy assumption 5 (i).

To prove (ii), notice that under assumption 8 (i), if $\left\|\frac{X_j - X_i}{h}\right\| > R$

$$
w_{ij}^{kernel}\left\|X_j - X_i\right\| = \frac{K\left(\frac{X_j - X_i}{h}\right)}{\sum_{l\in\mathcal{I}_{g's'}} K\left(\frac{X_l - X_i}{h}\right)}\left\|X_j - X_i\right\| = 0
$$

if $\left\|\frac{X_j - X_i}{h}\right\| < R$

$$
w_{ij}^{kernel}\left\|X_j - X_i\right\| \leq w_{ij}^{kernel}Rh
$$

Therefore,

$$
\mathbb{E}\left[\sum_{j\in\mathcal{I}_{g's'}} w_{ij}\left\|X_j - X_i\right\|\middle| i \in \mathcal{I}_{gs}\right] \leq Rh\cdot\mathbb{E}\left[\sum_{j\in\mathcal{I}_{g's'}} w_{ij}\middle| i \in \mathcal{I}_{gs}\right] = Rh \to 0
$$

The equality follows by the fact the pairwise weight w_{ij}^{kernel} satisfies assumption 5(i) and the convergence result follows by assumption 8.

To prove (iii), it suffices to show that

$$\mathbb{E}\big[w_j^5 \mid j \in \mathcal{I}_{g's'}\big] = O(1)$$

Conditional on the subgroup sample size n_{gs}, $n_{g's'}$, and $j \in \mathcal{I}_{g's'}$, by definition,

$$\begin{aligned} w_j^5 &= \left(\sum_{i \in \mathcal{I}_{gs}} w_{ij}\right)^5 \\ &= \sum_{i_1 \in \mathcal{I}_{gs}} \cdots \sum_{i_5 \in \mathcal{I}_{gs}} w_{i_1 j} \cdots w_{i_5 j} \end{aligned}$$

Consider a tuple $(i_1, \ldots, i_5)$ containing exactly k distinct indices, with multiplicities $a_1, \ldots, a_k$ satisfying

$$a_1 + \cdots + a_k = 5.$$

Conditional on all comparison group covariates, and $j \in \mathcal{I}_{g's'}$, $\{w_{ij}\}_{i \in \mathcal{I}_{gs}}$, are i.i.d random variables, so

$$\begin{aligned} &\mathbb{E}[w_{i_1 j} \cdots w_{i_5 j} \mid j \in \mathcal{I}_{g's'}, \{X_\ell : \ell \in \mathcal{I}_{g's'}\}] \\ &= \prod_{h=1}^{k} \mathbb{E}\big[w_{ij}^{a_h} \mid j \in \mathcal{I}_{g's'}, \{X_\ell : \ell \in \mathcal{I}_{g's'}\}\big] \\ &\leq \left(\mathbb{E}[w_{ij} \mid j \in \mathcal{I}_{g's'}, \{X_\ell : \ell \in \mathcal{I}_{g's'}\}]\right)^k \end{aligned}$$

The inequality follows from $0 \leq w_{ij} \leq 1$. On the other hand, the number of order five tuples containing exact k distinct indices is $n_{gs}(n_{gs}-1)\cdots(n_{gs}-k+1)C(5,k)$ which is bounded by $C(5,k)n_{gs}^k$, where $C(5,k)$ is the Stirling number of the second kind (a constant). It follows that

$$\mathbb{E}\big[w_j^5 \mid j \in \mathcal{I}_{g's'}, \{X_\ell : \ell \in \mathcal{I}_{g's'}\}\big] \leq \sum_{k=1}^{5} C(5,k) n_{gs}^k \left(\mathbb{E}[w_{ij} \mid j \in \mathcal{I}_{g's'}, \{X_\ell : \ell \in \mathcal{I}_{g's'}\}]\right)^k$$

By law of iterated expectation

$$\mathbb{E}\big[w_j^5 \mid j \in \mathcal{I}_{g's'}\big] \leq \sum_{k=1}^{5} C(5,k) n_{gs}^k \mathbb{E}\left[\left(\mathbb{E}[w_{ij} \mid \{X_j : j \in \mathcal{I}_{g's'}\}]\right)^k \mid j \in \mathcal{I}_{g's'}\right]$$

Now the key is to put an appropriate bound on $\mathbb{E}\left[\left(\mathbb{E}[w_{ij} \mid \{X_j : j \in \mathcal{I}_{g's'}\}]\right)^k \mid j \in \mathcal{I}_{g's'}\right]$

Fix $x \in \mathcal{X}$, and define

$$N_x = \sum_{\substack{\ell \in \mathcal{I}_{g's'} \\ \ell \neq j}} \mathbb{1}\{\|X_\ell - x\| \leq rh\} \, .$$

which is the total number of the comparison units other than unit j that falls into $B(x, rh)$. Since the sample is i.i.d

$$N_x \sim \text{Binomial}(n_{g's'} - 1, p_h(x)),$$

where

$$p_h(x) = \Pr(\|X_\ell - x\| \leq rh \mid \ell \in \mathcal{I}_{g's'}) \, .$$

By the lower bound of the density function and convex support condition imposed assumption 6 (i),

$$\begin{aligned}
p_h(x) &= \int_{\mathcal{X} \cap B(x,rh)} f_{g's'}(z) \, dz \\
&\geq \underline{f} \ \text{Leb}(\mathcal{X} \cap B(x, rh)) \\
&\geq \underline{f} \, c_{\mathcal{X}} Leb(B(x, rh)) \\
&\geq \underline{f} \, c_{\mathcal{X}} \frac{\pi^{q/2}}{\Gamma\left(\frac{q}{2} + 1\right)} r^q h^q
\end{aligned}$$

The second inequality follows from theorem 5 (pp. 12) of Viel et al. (2025). $\Gamma(\cdot)$ is the Gamma function. Thus, for some constant $C > 0$,

$$p_h(x) \geq C h^q$$

uniformly over $x \in \mathcal{X}$.

On the other hand,

$$\mathbb{E}[N_x \mid j \in \mathcal{I}_{g's'}] = (n_{g's'} - 1) p_h(x) \geq \frac{C n_{g's'} h^q}{2}$$

uniformly in x. Since $n_{g's'} h^q \to \infty$, $\mathbb{E}[N_x \mid j \in \mathcal{I}_{g's'}] \to \infty$ uniformly in x.

For the same positive integer k as above,

$$
\begin{aligned}
&\mathbb{E}\big[(N_x+1)^{-k} \mid j \in \mathcal{I}_{g's'}\big] \\
&= \mathbb{E}\left[(N_x+1)^{-k}\mathbb{1}\left(N_x > \frac{\mathbb{E}[N_x \mid j \in \mathcal{I}_{g's'}]}{2}\right)\middle| j \in \mathcal{I}_{g's'}\right] \\
&+ \mathbb{E}\left[(N_x+1)^{-k}\mathbb{1}\left(N_x \leq \frac{\mathbb{E}[N_x \mid j \in \mathcal{I}_{g's'}]}{2}\right)\middle| j \in \mathcal{I}_{g's'}\right] \\
&\leq \left(\frac{2}{\mathbb{E}[N_x \mid j \in \mathcal{I}_{g's'}]}\right)^k + \Pr\left(N_x < \frac{\mathbb{E}[N_x \mid j \in \mathcal{I}_{g's'}]}{2}\middle| j \in \mathcal{I}_{g's'}\right).
\end{aligned}
$$

By the Chernoff bound for binomial random variable,

$$
\Pr\left(N_x < \frac{\mathbb{E}[N_x \mid j \in \mathcal{I}_{g's'}]}{2}\middle| j \in \mathcal{I}_{g's'}\right) \leq \exp\left(-\frac{\mathbb{E}[N_x|j \in \mathcal{I}_{g's'}]}{8}\right).
$$

Since $\mathbb{E}[N_x|j \in \mathcal{I}_{g's'}] \to \infty$ uniformly in x and exponential decay dominates every fixed polynomial, for sufficiently large $n_{g's'}$,

$$
\exp\left(-\frac{\mathbb{E}[N_x|j \in \mathcal{I}_{g's'}]}{8}\right) \leq \left(\frac{2}{\mathbb{E}[N_x|j \in \mathcal{I}_{g's'}]}\right)^k.
$$

Therefore, for sufficiently large $n_{g's'}$

$$
\mathbb{E}\big[(N_x+1)^{-k}|j \in \mathcal{I}_{g's'}\big] \leq \frac{2^{k+1}}{(\mathbb{E}[N_x|j \in \mathcal{I}_{g's'}])^k} \leq \frac{C_k}{(n_{g's'}h^q)^k}.
$$

Now use the definition of the pairwise weight w_{ij} under the kernel matching

$$
w_{ij} = \frac{K\left(\frac{X_j - x}{h}\right)}{\sum_{\ell \in \mathcal{I}_{g's'}} K\left(\frac{X_\ell - x}{h}\right)} = \frac{K\left(\frac{X_j - x}{h}\right)}{K\left(\frac{X_j - x}{h}\right) + \sum_{\substack{\ell \in \mathcal{I}_{g's'} \\ \ell \neq j}} K\left(\frac{X_\ell - x}{h}\right)}
$$

By assumption 8(i), every comparison observation satisfying $\|X_\ell - x\| \leq rh$ contributes at least $\underline{K}$ to the denominator. Hence,

$$
\sum_{\ell \in \mathcal{I}_{g's'}} K\left(\frac{X_\ell - x}{h}\right) \geq K\left(\frac{X_j - x}{h}\right) + \underline{K}N_x.
$$

When $N_x \geq 1$, using the fact that the support of the kernel K is bounded above by $\overline{K}$, we have

$$w_{ij} \leq \frac{\overline{K}}{\underline{K} N_x}.$$

In addition, combining the fact that $w_{ij} \leq 1$, so

$$w_{ij} \leq \frac{2\overline{K}}{\underline{K}(N_x + 1)}.$$

For fixed x, N_x depends only on comparison observations other than j, and is therefore independent of X_j. As a result, for $k = 1, \ldots, 5$,

$$\begin{aligned}
&\mathbb{E}\big[\mathbb{1}\{\|X_j - x\| \leq Rh\}\, w_{ij}^k | j \in \mathcal{I}_{g's'}\big] \\
&\leq \mathbb{E}\left[\mathbb{1}\{\|X_j - x\| \leq Rh\} \left(\frac{2\overline{K}}{\underline{K}(N_x + 1)}\right)^k \middle| j \in \mathcal{I}_{g's'}\right] \\
&\leq \left(\frac{2\overline{K}}{\underline{K}}\right)^k \Pr(\|X_j - x\| \leq Rh | j \in \mathcal{I}_{g's'})\, \mathbb{E}\big[(N_x + 1)^{-k} | j \in \mathcal{I}_{g's'}\big].
\end{aligned}$$

The upper density bound and the convex support condition imposed by assumption 6 yields

$$\begin{aligned}
\Pr(\|X_j - x\| \leq Rh | j \in \mathcal{I}_{g's'}) &= \int_{\mathcal{X} \cap B(x, Rh)} f_{g's'}(z)\, dz \\
&\leq \overline{f}\ \mathrm{Leb}(B(x, Rh)) \\
&\leq \overline{f} \frac{\pi^{q/2}}{\Gamma\left(\frac{q}{2} + 1\right)} R^q h^q.
\end{aligned}$$

Consequently, for sufficiently large $n_{g's'}$

$$\begin{aligned}
\mathbb{E}\big[\mathbb{1}\{\|X_j - x\| \leq Rh\}\, w_{ij}^k | j \in \mathcal{I}_{g's'}\big] &\leq \left(\frac{2\overline{K}}{\underline{K}}\right)^k \overline{f} \frac{\pi^{q/2}}{\Gamma\left(\frac{q}{2} + 1\right)} R^q h^q \frac{C_k}{(n_{g's'} h^q)^k} \\
&= \left(\frac{2\overline{K}}{\underline{K}}\right)^k \overline{f} \frac{\pi^{q/2}}{\Gamma\left(\frac{q}{2} + 1\right)} R^q C_k h^q (n_{g's'} h^q)^{-k}
\end{aligned}$$

Lastly,

$$
\begin{aligned}
\mathbb{E}[w_{ij}|j\in\mathcal{I}_{g's'},\{X_\ell:\ell\in\mathcal{I}_{g's'}\}] &= \int_{\mathcal{X}} f_{gs}(x)w_{ij}\,dx \\
&= \int_{\mathcal{X}} \mathbb{1}\{\|x-X_j\|\le Rh\}\,f_{gs}(x)w_{ij}\,dx \\
&\le \left(\int_{\mathcal{X}} \mathbb{1}\{\|x-X_j\|\le Rh\}\;dx\right)^{\frac{k-1}{k}} \left(\int_{\mathcal{X}} f_{gs}^k(x)w_{ij}^k\,dx\right)^{\frac{1}{k}} \\
&= \left[\mathrm{Leb}(B(X_j,Rh))\right]^{\frac{k-1}{k}} \left(\int_{\mathcal{X}} f_{gs}^k(x)w_{ij}^k\,dx\right)^{\frac{1}{k}}
\end{aligned}
$$

The inequality follows from the Hölder's inequality. Thus,

$$
\begin{aligned}
(\mathbb{E}[w_{ij}|j\in\mathcal{I}_{g's'},\{X_\ell:\ell\in\mathcal{I}_{g's'}\}])^k &\le \left[\mathrm{Leb}(B(X_j,Rh))\right]^{k-1}\times\int_{\mathcal{X}} f_{gs}(x)^k w_{ij}^k\,dx \\
&= \left[\mathrm{Leb}(B(X_j,Rh))\right]^{k-1}\times\int_{\mathcal{X}} \mathbb{1}\{\|x-X_j\|\le Rh\}\,f_{gs}(x)^k w_{ij}^k\,dx \\
&\le \overline{f}\left(\frac{\pi^{q/2}}{\Gamma\left(\frac{q}{2}+1\right)}R^q h^q\right)^{k-1}\int_{\mathcal{X}} \mathbb{1}\{\|x-X_j\|\le Rh\}\,w_{ij}^k\,dx
\end{aligned}
$$

Take expectation on both sides with respect to all the comparison group covariates, and the event $j\in\mathcal{I}_{g's'}$,

$$
\begin{aligned}
&\mathbb{E}\left[(\mathbb{E}[w_{ij}|j\in\mathcal{I}_{g's'},\{X_\ell:\ell\in\mathcal{I}_{g's'}\}])^k\right] \\
&\le \overline{f}\left(\frac{\pi^{q/2}}{\Gamma\left(\frac{q}{2}+1\right)}R^q h^q\right)^{k-1}\mathbb{E}\left[\int_{\mathcal{X}} \mathbb{1}\{\|x-X_j\|\le Rh\}\,w_{ij}^k\,dx\,\middle|\,j\in\mathcal{I}_{g's'}\right] \\
&= \overline{f}\left(\frac{\pi^{q/2}}{\Gamma\left(\frac{q}{2}+1\right)}R^q\right)^{k-1} h^{q(k-1)}\int_{\mathcal{X}} \mathbb{E}\left[\mathbb{1}\{\|x-X_j\|\le Rh\}\,w_{ij}^k|j\in\mathcal{I}_{g's'}\right]\,dx \\
&\le C_k h^{q(k-1)}\int_{\mathcal{X}} h^q(n_{g's'}h^q)^{-k}\,dx \\
&\le \frac{C_k}{n_{g's'}^k}
\end{aligned}
$$

The first equality follows by the fact that the integrand is non-negative so we can apply the Tonelli's theorem to exchange the expectation and integration.

Therefore,

$$\mathbb{E}\big[w_j^5 | j \in \mathcal{I}_{g's'}\big] \leq \sum_{k=1}^{5} C(5,k) n_{gs}^k \mathbb{E}\left[\left(\mathbb{E}[w_{ij} | j \in \mathcal{I}_{g's'}, \{X_j : j \in \mathcal{I}_{g's'}\}]\right)^k\right]$$
$$\leq \sum_{k=1}^{5} C(5,k) C_k \left(\frac{n_{gs}}{n_{g's'}}\right)^k$$

Notice that $\frac{n_{gs}}{n_{g's'}} \to \frac{p_{gs}}{p_{g's'}} > 0$, so $\left(\frac{n_{gs}}{n_{g's'}}\right)^k = O(1)$. This finally shows that,

$$\mathbb{E}\big[w_j^5 \mid j \in \mathcal{I}_{g's'}\big] = O(1)$$

(iv) is shown in lemma 10. ■

Proof. Proof of lemma 4. To show the first equality, apply the law of total probability, we have:

$$\mathbb{E}\left[\frac{(1-p_{11})L(D|G,S,X)}{\mathbb{E}[(D-L(D|G,S,X))^2]}X \middle| D=0\right] = \mathbb{E}\left[\frac{(1-p_{11})L(D|G,S,X)}{\mathbb{E}[(D-L(D|G,S,X))^2]}X \middle| G=0, S=1\right]\frac{p_{01}}{1-p_{11}}$$
$$+\mathbb{E}\left[\frac{(1-p_{11})L(D|G,S,X)}{\mathbb{E}[(D-L(D|G,S,X))^2]}X \middle| G=1, S=0\right]\frac{p_{10}}{1-p_{11}}$$
$$+\mathbb{E}\left[\frac{(1-p_{11})L(D|G,S,X)}{\mathbb{E}[(D-L(D|G,S,X))^2]}X \middle| G=0, S=0\right]\frac{p_{00}}{1-p_{11}}$$
$$= \mathbb{E}\left[\frac{p_{01}L(D|G,S,X)}{\mathbb{E}[(D-L(D|G,S,X))^2]}X \middle| G=0, S=1\right] + \mathbb{E}\left[\frac{p_{10}L(D|G,S,X)}{\mathbb{E}[(D-L(D|G,S,X))^2]}X \middle| G=1, S=0\right]$$
$$+\mathbb{E}\left[\frac{p_{00}L(D|G,S,X)}{\mathbb{E}[(D-L(D|G,S,X))^2]}X \middle| G=0, S=0\right]$$

To prove the second equality, it suffices to show that

$$\mathbb{E}[p_{11}\left(1-L(D|G,S,X)\right)X|G=1,S=1] = \mathbb{E}[(1-p_{11})L(D|G,S,X)X|D=0]$$

Note that

$$\mathbb{E}[p_{11}\left(1-L(D|G,S,X)\right)|G=1,S=1] - \mathbb{E}[(1-p_{11})L(D|G,S,X)|D=0]$$
$$= \mathbb{E}[(D(1-L(D|G,S,X)) - (1-D)L(D|G,S,X))X]$$
$$= \mathbb{E}[(D-L(D|G,S,X))X]$$
$$= 0$$

The last equality follows from the fact that X is orthogonal to the regression residual $D - L(D|G,S,X)$ ■

Proof. Proof of lemma 5.
Let $\mathbf{G} = (G_1, \ldots, G_n)$, $\mathbf{S} = (S_1, \ldots, S_n)$, and $\mathbf{X} = (X_1', \ldots, X_n')'$, Now,

$$\begin{aligned}
T_n = \frac{n}{n_{11}} \Bigg\{ & \frac{1}{n} \sum_{i=1}^{n} \left[\mathbb{1}(i \in \mathcal{I}_{11}) - \mathbb{1}(i \in \mathcal{I}_{gs}) w_i \right] (Z_i - \mu_Z(G_i, S_i, X_i)) \\
& + \frac{1}{n} \sum_{i \in \mathcal{I}_{11}} (\mu_Z(1,1,X_i) - \mu_Z(g,s,X_i)) \\
& + \frac{1}{n} \sum_{i=1}^{n} \left[\mathbb{1}(i \in \mathcal{I}_{11}) - \mathbb{1}(i \in \mathcal{I}_{gs}) w_i \right] \mu_Z(g,s,X_i) \Bigg\} \\
:= & \frac{n}{n_{11}} \left(T_{1,n} + T_{2,n} + T_{3,n} \right).
\end{aligned}$$

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

$$\mathbb{V}[T_{1,n}|\mathbf{G}, \mathbf{S}, \mathbf{X}] = \frac{1}{n^2} \sum_{i=1}^{n} \left[\mathbb{1}(i \in \mathcal{I}_{11}) - \mathbb{1}(i \in \mathcal{I}_{gs}) w_i \right]^2 \sigma_Z^2(G_i, S_i, X_i)$$

from which

$$\begin{aligned}
\mathbb{V}[T_{1,n}] &= \frac{1}{n} \mathbb{E}\left[\left[\mathbb{1}(i \in \mathcal{I}_{11}) - \mathbb{1}(i \in \mathcal{I}_{gs}) w_i \right]^2 \sigma_Z^2(G_i, S_i, X_i) \right] \\
&\leq \frac{C}{n} \left(1 + p_{gs} \mathbb{E}\left[w_i^2 | i \in \mathcal{I}_{gs} \right] \right) \\
&\to 0
\end{aligned}$$

the last line follows by assumption 5. Thus, $T_{1,n} = o_p(1)$. Next,

$$\begin{aligned}
& Var[\mathbb{1}(i \in \mathcal{I}_{gs})(\mu_Z(1,1,X_i) - \mu_Z(g,s,X_i))] \\
& \leq \mathbb{E}[(\mu_Z(1,1,X_i) - \mu_Z(g,s,X_i))^2] \\
& \leq \mathbb{E}\left[(|\mu_Z(1,1,X_i)| + |\mu_Z(g,s,X_i)|)^2 \right] \\
& < \infty
\end{aligned}$$

The second to last inequality follows by the fact that the conditional mean function is

Lipschitz continuous and the support of X_i is bounded. Apply the WLLN,

$$
\begin{aligned}
T_{2,n} &= \frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\left(\mu_Z(1,1,X_i) - \mu_Z(g,s,X_i)\right) \\
&\xrightarrow{p} \mathbb{E}\left[\mu_Z(1,1,X) - \mu_Z(g,s,X)|G=1,S=1\right]p_{11} \\
&= \left(\mu_Z(1,1) - \mathbb{E}[\mu_Z(g,s,X)|G=1,S=1]\right)p_{11}.
\end{aligned}
$$

Finally, for the third term,

$$
\begin{aligned}
T_{3,n} &= \frac{1}{n}\sum_{i=1}^{n}\left[\mathbb{1}(i\in\mathcal{I}_{11}) - \mathbb{1}(i\in\mathcal{I}_{gs})w_i\right]\mu_Z(g,s,X_i) \\
&= \frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\mu_Z(g,s,X_i) - \frac{1}{n}\sum_{j\in\mathcal{I}_{gs}} w_i\mu_Z(g,s,X_j) \\
&= \frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\mu_Z(g,s,X_i) - \frac{1}{n}\sum_{j\in\mathcal{I}_{gs}}\sum_{i\in\mathcal{I}_{11}} w_{ij}\mu_Z(g,s,X_j) \\
&= \frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\mu_Z(g,s,X_i) - \frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\mu_Z(g,s,X_j) \\
&= \frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left[\mu_Z(g,s,X_i) - \mu_Z(g,s,X_j)\right] \\
&\leq \frac{C_2}{n}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left\|X_i - X_j\right\|
\end{aligned}
$$

The third equality follows from $\sum_{j\in\mathcal{I}_{gs}} w_{ij} = 1$. Notice the last term is non-negative because $w_{ij} \geq 0$ By Markov inequality,

$$
\begin{aligned}
P\left(\frac{C}{n}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left\|X_i - X_j\right\| > \epsilon\right) &\leq \frac{\mathbb{E}\left[\frac{C}{n}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left\|X_i - X_j\right\|\right]}{\epsilon} \\
&= \frac{C\mathbb{E}\left[\mathbb{1}(i\in\mathcal{I}_{11})\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left\|X_i - X_j\right\|\right]}{\epsilon} \\
&= \frac{Cp_{11}\mathbb{E}\left[\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left\|X_i - X_j\right\|\,\middle|\, G_i = 1, S_i = 1\right]}{\epsilon} \\
&\to 0
\end{aligned}
$$

The first inequality follows from the fact that the sample is i.i.d. and the pairwise weight is permutation equivariant. The last line follows from assumption 5(ii). Thus $T_{3,n} = o_p(1)$.

Therefore,

$$T_n = \frac{n}{n_{11}}\left(T_{1,n} + T_{2,n} + T_{3,n}\right) \xrightarrow{p} \mu_Z(1,1) - \mathbb{E}[\mu_Z(g,s,X)|G=1,S=1]$$

as required. ∎

Proof. Proof of lemma 6.

For the first set of claim, by Cheyshev's inequality,

$$\begin{aligned}
&P\left(\left|\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma_{11,l}^2(X_i) - \mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma_{11,l}^2(X_i)\right]\right| > \epsilon\right)\\
&\leq \frac{Var\left(\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma_{11,l}^2(X_i)\right)}{\epsilon^2}\\
&= \frac{n_{11}Var\left(\sigma_{11,l}^2(X_i)\right)}{n^2\epsilon^2}\\
&\leq \frac{n_{11}\mathbb{E}\left[\left(\sigma_{11,l}^2(X_i)\right)^2\right]}{n^2\epsilon^2}\\
&\leq \frac{n_{11}\overline{\sigma}^2}{n^2\epsilon^2}\\
&\to o_p(1)
\end{aligned}$$

The last inequality follows from assumption 6(iii). The last line follows from $n_{11} \asymp n$. This shows that

$$\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma_{11,l}^2(X_i) - \mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma_{11,l}^2(X_i)\right] = o_p(1)$$

On the other hand,

$$\mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma_{11,l}^2(X_i)\right] \leq p_{11}\overline{\sigma}^2 \implies \mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma_{11,l}^2(X_i)\right] = O(1)$$

The argument above shows that

$$Var\left(\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma_{11,l}^2(X_i)\right) = o(1)$$

Therefore,

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right)^2\right]=O(1)$$

Now,

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)-\frac{1}{p_{11}}\mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right]\right|\right]\\
&=\mathbb{E}\left[\left|\left(\frac{n}{n_{11}}-\frac{1}{p_{11}}\right)\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)+\frac{1}{p_{11}}\left(\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)-\mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right]\right)\right|\right]\\
&\leq\mathbb{E}\left[\left|\left(\frac{n}{n_{11}}-\frac{1}{p_{11}}\right)\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right|\right]+\mathbb{E}\left[\left|\frac{1}{p_{11}}\left(\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)-\mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right]\right)\right|\right]\\
&\leq\mathbb{E}\left[\left|\frac{n}{n_{11}}-\frac{1}{p_{11}}\right|^2\right]^{1/2}\mathbb{E}\left[\left|\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right|^2\right]^{1/2}+\frac{1}{p_{11}}\left(Var\left(\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right)\right)^{1/2}\\
&=o(1)O(1)+o(1)\\
&=o(1)
\end{aligned}$$

Then by Jensen's inequality,

$$\begin{aligned}
&\left|\mathbb{E}\left[\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right]-\frac{1}{p_{11}}\mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right]\right|\\
&=\left|\mathbb{E}\left[\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)-\frac{1}{p_{11}}\mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right]\right]\right|\\
&\leq\mathbb{E}\left[\left|\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)-\frac{1}{p_{11}}\mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right]\right|\right]\\
&=o_p(1)
\end{aligned}$$

Then by Markov inequality, for any $\epsilon>0$,

$$\begin{aligned}
&P\left(\left|\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)-\frac{1}{p_{11}}\mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right]\right|\right)\\
&\leq\frac{1}{\epsilon}\mathbb{E}\left[\left|\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)-\frac{1}{p_{11}}\mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right]\right|\right]\to 0
\end{aligned}$$

So,

$$\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)-\frac{1}{p_{11}}\mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right]=o_p(1)$$

Combining the two results above,

$$\begin{aligned}
&\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)-\mathbb{E}\left[\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right]\\
&=\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)-\frac{1}{p_{11}}\mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right]+\frac{1}{p_{11}}\mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right]-\mathbb{E}\left[\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sigma^2_{11,l}(X_i)\right]\\
&=o_p(1)
\end{aligned}$$

This finishes the proof of the first set of claims.

For the second claim, let $\Omega_n=\sum_{i\in\mathcal{I}_{gs}}w_i^2\sigma^2_{gs,l}(X_i)$, and $\Omega_n^{(j)}=\mathbb{1}(j\in\mathcal{I}_{gs})\left(w_j^{(j)}\right)^2\sigma^2_{gs,l}(X'_j)+\sum_{i\in\mathcal{I}_{gs}\setminus\{j\}}\left(w_i^{(j)}\right)^2\sigma^2_{gs,l}(X_i)$ be the counterpart of Ω_n with (G_j,S_j,X_j) being replaced by (G'_j,S'_j,X'_j).

$$\begin{aligned}
\left|\Omega_n-\Omega_n^{(j)}\right|&=\left|\mathbb{1}(j\in\mathcal{I}_{gs})w_j^2\sigma^2_{gs,l}(X_j)-\mathbb{1}(j\in\mathcal{I}_{gs})\left(w_j^{(j)}\right)^2\sigma^2_{gs,l}(X'_j)+\sum_{i\in\mathcal{I}_{gs}\setminus\{1\}}\left(w_i^2-\left(w_i^{(j)}\right)^2\right)\sigma^2_{gs,l}(X_i)\right|\\
&\le\overline{\sigma}^2\left(\mathbb{1}(j\in\mathcal{I}_{gs})(w_j^2)+\mathbb{1}'(j\in\mathcal{I}_{gs})\left(w_j^{(j)}\right)^2+\sum_{i\in\mathcal{I}_{gs}\setminus\{j\}}\left|w_i^2-\left(w_i^{(j)}\right)^2\right|\right)
\end{aligned}$$

By Efron Stein inequality,

$$
\begin{aligned}
Var(\Omega_n) &\leq \frac{1}{2}\sum_{j=1}^{n} \mathbb{E}\left[\left(\Omega_n - \Omega_n^{(j)}\right)^2\right] \\
&\leq \frac{1}{2}\sum_{j=1}^{n} \mathbb{E}\left[\overline{\sigma}^2 \left(\mathbb{1}(j \in \mathcal{I}_{gs})(w_j^2) + \mathbb{1}'(j \in \mathcal{I}_{gs})\left(w_j^{(j)}\right)^2 + \sum_{i \in \mathcal{I}_{gs}\backslash\{j\}} \left| w_i^2 - \left(w_i^{(j)}\right)^2\right|\right)^2\right] \\
&\leq \frac{1}{2}\sum_{j=1}^{n} \mathbb{E}\left[3\overline{\sigma}^2 \left(\mathbb{1}(j \in \mathcal{I}_{gs})(w_j^4) + \mathbb{1}'(j \in \mathcal{I}_{gs})\left(w_j^{(j)}\right)^4 + \left(\sum_{i \in \mathcal{I}_{gs}\backslash\{j\}} \left| w_i^2 - \left(w_i^{(j)}\right)^2\right|\right)^2\right)\right] \\
&= \frac{3n\overline{\sigma}^2}{2}\left(\mathbb{E}[\mathbb{1}(j \in \mathcal{I}_{gs})w_j^4] + \mathbb{E}\left[\mathbb{1}(j \in \mathcal{I}_{gs})\left(w_j^{(j)}\right)^4\right] + \mathbb{E}\left[\left(\sum_{i \in \mathcal{I}_{gs}\backslash\{j\}} \left| w_i^2 - \left(w_i^{(j)}\right)^2\right|\right)^2\right]\right) \\
&= \frac{3n\overline{\sigma}^2}{2}(O(1) + O(1) + O(n^{1/2})) \\
&= O(n^{3/2})
\end{aligned}
$$

The second inequality follows from $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$. The second equality follows from assumption 5(iii) and lemma 7(ii).

By Cheyshev's inequality,

$$
\begin{aligned}
&P\left(\left|\frac{1}{n}\sum_{i \in \mathcal{I}_{gs}} w_i^2 \sigma_{gs,l}^2(X_i) - \mathbb{E}\left[\frac{1}{n}\sum_{i \in \mathcal{I}_{gs}} w_i^2 \sigma_{gs,l}^2(X_i)\right]\right| > \epsilon\right) \\
&\leq \frac{\Omega_n}{n^2\epsilon^2} = \frac{n^{3/2}}{n^2\epsilon^2} \to 0
\end{aligned}
$$

This shows that

$$
\frac{1}{n}\sum_{i \in \mathcal{I}_{gs}} w_i^2 \sigma_{gs,l}^2(X_i) - \mathbb{E}\left[\frac{1}{n}\sum_{i \in \mathcal{I}_{gs}} w_i^2 \sigma_{gs,l}^2(X_i)\right] = o_p(1)
$$

On the other hand,

$$
\begin{aligned}
\mathbb{E}\left[\frac{1}{n}\sum_{i \in \mathcal{I}_{gs}} w_i^2 \sigma_{gs,l}^2(X_i)\right] &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[\mathbb{1}(j \in \mathcal{I}_{gs}) w_i^2 \sigma_{gs,l}^2(X_i)\right] \\
&\leq \overline{\sigma}^2 \mathbb{E}\left[w_i^2 | i \in \mathcal{I}_{gs}\right] \\
&= O(1)
\end{aligned}
$$

The above argument implies that

$$Var\left(\frac{1}{n}\sum_{i\in\mathcal{I}_{gs}} w_i^2\sigma_{gs,l}^2(X_i)\right) = \frac{1}{n^2}Var(\Omega_n) = \frac{1}{n^2}O(n^{3/2}) = o(1)$$

Therefore,

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i\in\mathcal{I}_{gs}} w_i^2\sigma_{gs,l}^2(X_i)\right)^2\right] = O(1)$$

Then following the same proof strategy as the proof of the first set of claim, we have:

$$\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{gs}} w_i^2\sigma_{gs,l}^2(X_i) - \mathbb{E}\left[\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{gs}} w_i^2\sigma_{gs,l}^2(X_i)\right] = o_p(1)$$

This proves the second set of claims. ■

Proof. Proof of lemma 7.
(i)

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n^2}\sum_{j\in\mathcal{I}_{gs}} w_j^2\right] &= \frac{1}{n}\mathbb{E}[w_j^2\mathbb{1}(i\in\mathcal{I}_{gs})] \\
&= \frac{p_{gs}}{n}\mathbb{E}[w_j^2 \mid i\in\mathcal{I}_{gs}] \\
&\le \frac{1}{n}\left(\mathbb{E}[w_j^{4+\delta} \mid i\in\mathcal{I}_{gs}]\right)^{\frac{2}{4+\delta}} \\
&= O(n^{-1})
\end{aligned}$$

The third line follows from Lyapunov inequality and the last line follows from assumption 5(iii).
So there exists a constant C and N, such that, for all $n \ge N$,

$$\mathbb{E}\left[\frac{1}{n^2}\sum_{j\in\mathcal{I}_{gs}} w_j^2\right] \le \frac{C}{N}$$

. For any $\epsilon > 0$, pick $M = \frac{C}{\epsilon}$, then for all $n \geq N$,

$$\begin{aligned} P\left(\frac{1}{n^2}\sum_{j\in\mathcal{I}_{gs}} w_j^2 \geq \frac{M}{N}\right) &\leq \frac{N}{M}\mathbb{E}\left[\frac{1}{n^2}\sum_{j\in\mathcal{I}_{gs}} w_j^2\right] \\ &\leq \frac{C}{M} = \epsilon \end{aligned}$$

$$\sum_{i\in\mathcal{I}_{11}} w_{ij}w_{ij'} \leq \sum_{i\in\mathcal{I}_{11}} w_{ij} \leq \max_{j\in\mathcal{I}_{gs}} w_j$$

$$\begin{aligned} &\max_{j\in\mathcal{I}_{gs}} w_j^2 \leq \sum_{j\in\mathcal{I}_{gs}} w_{gs}^2 \\ &\implies \left(\max_{j\in\mathcal{I}_{gs}} w_j\right)^2 \leq \sum_{j\in\mathcal{I}_{gs}} w_{gs}^2 \\ &\implies \max_{j\in\mathcal{I}_{gs}} w_j \leq \left(\sum_{j\in\mathcal{I}_{gs}} w_{gs}^2\right)^{1/2} = O_p(n^{1/2}) \end{aligned}$$

(ii) Notice that:

$$\begin{aligned} \sum_{j\in\mathcal{I}_{g's'}\setminus\{1\}} \left|w_j^2 - \left(w_j^{(1)}\right)^2\right| &= \sum_{j\in\mathcal{I}_{g's'}\setminus\{1\}} \left(w_j + w_j^{(1)}\right)\left|w_j - w_j^{(1)}\right| \\ &\leq \left(\max_{j\in\mathcal{I}_{g's'}\setminus\{1\}} w_j + \max_{j\in\mathcal{I}_{g's'}\setminus\{1\}} w_j^{(1)}\right) \sum_{j\in\mathcal{I}_{g's'}\setminus\{1\}} \left|w_j - w_j^{(1)}\right| \end{aligned}$$

Therefore,

$$
\begin{aligned}
&\mathbb{E}\left[\left(\sum_{j\in\mathcal{I}_{g's'}\setminus\{1\}}\left|w_j^2-\left(w_j^{(1)}\right)^2\right|\right)^2\right]\\
&\leq \mathbb{E}\left[\left(\max_{j\in\mathcal{I}_{g's'}\setminus\{1\}} w_j+\max_{j\in\mathcal{I}_{g's'}\setminus\{1\}} w_j^{(1)}\right)^2\left(\sum_{j\in\mathcal{I}_{g's'}\setminus\{1\}}\left|w_j-w_j^{(1)}\right|\right)^2\right]\\
&\leq \mathbb{E}\left[\left(\max_{j\in\mathcal{I}_{g's'}\setminus\{1\}} w_j+\max_{j\in\mathcal{I}_{g's'}\setminus\{1\}} w_j^{(1)}\right)^4\right]^{1/2}\mathbb{E}\left[\left(\sum_{j\in\mathcal{I}_{g's'}\setminus\{1\}}\left|w_j-w_j^{(1)}\right|\right)^4\right]^{1/2}\\
&\leq \left(8\mathbb{E}\left[\max_{j\in\mathcal{I}_{g's'}\setminus\{1\}} w_j^4+\max_{j\in\mathcal{I}_{g's'}\setminus\{1\}}\left(w_j^{(1)}\right)^4\right]\right)^{1/2}\mathbb{E}\left[\left(\sum_{j\in\mathcal{I}_{g's'}\setminus\{1\}}\left|w_j-w_j^{(1)}\right|\right)^4\right]^{1/2}\\
&\leq \left(8\mathbb{E}\left[\sum_{j\in\mathcal{I}_{g's'}\setminus\{1\}} w_j^4+\sum_{j\in\mathcal{I}_{g's'}\setminus\{1\}}\left(w_j^{(1)}\right)^4\right]\right)^{1/2}\mathbb{E}\left[\left(\sum_{j\in\mathcal{I}_{g's'}\setminus\{1\}}\left|w_j-w_j^{(1)}\right|\right)^4\right]^{1/2}\\
&= \left(8\sum_{j\in\mathcal{I}_{g's'}\setminus\{1\}}\mathbb{E}\left[w_j^4+\left(w_j^{(1)}\right)^4\right]\right)^{1/2}\mathbb{E}\left[\left(\sum_{j\in\mathcal{I}_{g's'}\setminus\{1\}}\left|w_j-w_j^{(1)}\right|\right)^4\right]^{1/2}\\
&= O(n^{1/2})O(1)\\
&= O(n^{1/2})
\end{aligned}
$$

The second equality follows from Cauchy Schwarz inequality. The third inequality follows from C_r inequality. The second last equality follows from assumption 5(ii) and (iii).

(iii) By Lyapunov inequality,

$$\mathbb{E}[w_j^r|j\in\mathcal{I}_{gs}]\leq\mathbb{E}[w_j^{4+\delta}|j\in\mathcal{I}_{gs}]^{\frac{r}{4+\delta}}=O(1)$$

This proves the first claim. So for any $\epsilon>0$ and $M>0$ there exists N such that, for any $n\geq N$, $\mathbb{E}[w_j^r|j\in\mathcal{I}_{gs}]\leq\frac{M}{\epsilon p_{gs}}$ By Markov's inequality, for any $n\geq N$

$$
\begin{aligned}
P\left(\frac{1}{n}\sum_{i\in\mathcal{I}_{gs}}w_j^r>M\right)&\leq\frac{\mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{gs}}w_j^r\right]}{M}\\
&=\frac{p_{gs}\mathbb{E}\left[w_j^r|j\in\mathcal{I}_{gs}\right]}{M}\\
&\leq\epsilon
\end{aligned}
$$

This proves the second claim.

(iv) By Markov inequality,

$$\begin{aligned}P\left(\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left\|X_i - X_j\right\| \geq \epsilon\right) &\leq \frac{\mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left\|X_i - X_j\right\|\right]}{\epsilon} \\ &= \frac{\mathbb{E}\left[\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left\|X_i - X_j\right\| |i \in \mathcal{I}_{11}\right] p_{11}}{\epsilon} \\ &\to 0\end{aligned}$$

The last line follows from assumption 5(ii).

By Slutsky theorem,

$$\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left\|X_i - X_j\right\| = \frac{n}{n_{11}}\frac{1}{n}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left\|X_i - X_j\right\| = o_p(1)$$

On the other hand, since the support is compact,

$$\begin{aligned}\left(\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left\|X_i - X_j\right\|\right)^2 &\leq C\cdot\sum_{j\in\mathcal{I}_{gs}} w_{ij}\cdot\left(\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left\|X_i - X_j\right\|\right) \\ &= C\left(\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left\|X_i - X_j\right\|\right)\end{aligned}$$

Therefore,

$$\frac{1}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\left(\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left\|X_i - X_j\right\|\right)^2 \leq \frac{C}{n_{11}}\sum_{i\in\mathcal{I}_{11}}\sum_{j\in\mathcal{I}_{gs}} w_{ij}\left\|X_i - X_j\right\| = o_p(1)$$

■

Proof. Proof of lemma 8.

Define the event $E_n = \{n_{gs} \geq np_{gs}/2\}$. By Hoeffding's Inequality,

$$P(E_n^c) = P\left(n_{11} - np_{11} < -\frac{np_{11}}{2}\right) \leq exp\left(-\frac{np_{11}^2}{2}\right)$$

On the event E_n,

$$\left|\frac{n}{n_{11}} - \frac{1}{p_{11}}\right| = \frac{|n_{11} - np_{11}|}{n_{11}p_{11}} \leq \frac{2\,|n_{11} - np_{11}|}{np_{11}^2}$$

So that

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}(E_n)\left(\frac{n}{n_{gs}} - \frac{1}{p_{gs}}\right)^2\right] &\leq \mathbb{E}\left[\mathbb{1}(E_n)\frac{4(n_{gs} - \mathbb{E}[n_{gs}])^2}{n^2p_{gs}^4}\right] \\
&\leq \frac{4Var(n_{gs})}{n^2p_{gs}^4} = \frac{4(1-p_{gs})}{np_{gs}^3} = O(n^{-1})
\end{aligned}$$

On the event E_n^c, $\frac{n}{n_{11}} \leq n$, so

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}(E_n^c)\left(\frac{n}{n_{gs}} - \frac{1}{p_{gs}}\right)^2\right] &\leq \mathbb{E}\left[\mathbb{1}(E_n^c)\left(n + \frac{1}{p_{gs}}\right)^2\right] \\
&= \left(n + \frac{1}{p_{gs}}\right)^2 P(E_n^c) \\
&\leq \left(n + \frac{1}{p_{gs}}\right)^2 exp\left(-\frac{np_{11}^2}{2}\right) \\
&= o(n^{-1})
\end{aligned}$$

Combine the two results we prove lemma 8. ■

Proof. Proof of lemma 9.

By conditional Chebyshev's inequality, for any $\epsilon > 0$,

$$P(|T_n| \geq \epsilon|\mathbf{G}, \mathbf{S}, \mathbf{X}) \leq \frac{Var(T_n|\mathbf{G}, \mathbf{S}, \mathbf{X})}{\epsilon^2}$$

By law of iterated expectation and the property of probability,

$$P(|T_n| \geq \epsilon) = \mathbb{E}[P(|T_n| \geq \epsilon|\mathbf{G}, \mathbf{S}, \mathbf{X})] \leq \min\left\{1, \frac{\mathbb{E}[Var(T_n|\mathbf{G}, \mathbf{S}, \mathbf{X})]}{\epsilon^2}\right\}$$

Let $Z_n = \min\left\{1, \frac{Var(T_n|\mathbf{G},\mathbf{S},\mathbf{X})}{\epsilon^2}\right\}$, now for any $0 < \delta < 1$,

$$\begin{aligned}\mathbb{E}\left[Z_n\right] &= \mathbb{E}\left[Z_n \mathbb{1}(Z_n \leq \delta)\right] + \mathbb{E}\left[Z_n \mathbb{1}(Z_n > \delta)\right] \\ &\leq \delta + P(Z_n > \delta) \\ &= \delta + P\left(Var(T_n|\mathbf{G},\mathbf{S},\mathbf{X}) > \delta\epsilon^2\right)\end{aligned}$$

This implies that for any $\delta > 0$,

$$\limsup_{n\to\infty} \mathbb{E}[Z_n] \leq \delta$$

Therefore,

$$\lim_{n\to\infty} \mathbb{E}[Z_n] = \lim_{n\to\infty} \min\left\{1, \frac{\mathbb{E}[Var(T_n|\mathbf{G},\mathbf{S},\mathbf{X})]}{\epsilon^2}\right\} = 0$$

Combine the results above, we have: $T_n \xrightarrow{p} 0$ ■

Proof. Proof of lemma 10.

Consider an intermediate status where only unit 1 is deleted and let $\widetilde{w}_{ij}^{(1)}$ and $\widetilde{w}_j^{(1)}$ denote the corresponding pairwise weight and matching weight under this status. Then by triangular inequality,

$$\sum_{j\in\mathcal{I}_{gs}\backslash\{1\}} \left|w_j - w_j^{(1)}\right| = \sum_{j\in\mathcal{I}_{gs}\backslash\{1\}} \left|w_j - \widetilde{w}_j^{(1)} + \widetilde{w}_j^{(1)} - w_j^{(1)}\right| \leq \sum_{j\in\mathcal{I}_{gs}\backslash\{1\}} \left|w_j - \widetilde{w}_j^{(1)}\right| + \sum_{j\in\mathcal{I}_{gs}\backslash\{1\}} \left|\widetilde{w}_j^{(1)} - w_j^{(1)}\right|$$

For the first term, there are three cases.

Case 1: $1 \notin \mathcal{I}_{11} \cup \mathcal{I}_{gs}$, then in this case nothing changes, so $\sum_{j\in\mathcal{I}_{gs}\backslash\{1\}} \left|w_j - \widetilde{w}_j^{(1)}\right| = 0$.

Case 2: $1 \in \mathcal{I}_{11}$, then for all three matching methods, since the pairwise matching weight w_{ij} only uses the information from the reference unit i and the comparison group, so in this case it removes the pairwise matching weight of unit 1 and leaves all other pairwise matching weight unchanged. Therefore,

$$\sum_{j\in\mathcal{I}_{gs}\backslash\{1\}} \left|w_j - \widetilde{w}_j^{(1)}\right| = \sum_{j\in\mathcal{I}_{gs}\backslash\{1\}} w_{ij} = 1$$

Case 3: $1 \in \mathcal{I}_{gs}$, by triangular inequality

$$\begin{aligned}
\sum_{j\in\mathcal{I}_{gs}\setminus\{1\}} \left|w_j - \widetilde{w}_j^{(1)}\right| &= \sum_{j\in\mathcal{I}_{gs}\setminus\{1\}} \left|\sum_{i\in\mathcal{I}_{11}} w_{ij} - \widetilde{w}_{ij}^{(1)}\right| \\
&\leq \sum_{j\in\mathcal{I}_{gs}\setminus\{1\}} \sum_{i\in\mathcal{I}_{11}} \left|w_{ij} - \widetilde{w}_{ij}^{(1)}\right| \\
&= \sum_{i\in\mathcal{I}_{11}} \sum_{j\in\mathcal{I}_{gs}\setminus\{1\}} \left|w_{ij} - \widetilde{w}_{ij}^{(1)}\right|
\end{aligned}$$

Subcase 1: Nearest neighbor matching with fixed and diverging neighbor. If $1 \notin \mathcal{M}(i)$, then $\sum_{j\in\mathcal{I}_{gs}\setminus\{1\}} \left|w_{ij} - \widetilde{w}_{ij}^{(1)}\right| = 0$. If $1 \in \mathcal{M}(i)$, then $\sum_{j\in\mathcal{I}_{gs}\setminus\{1\}} \left|w_{ij} - \widetilde{w}_{ij}^{(1)}\right| = \frac{1}{M} = w_{i1}$, so

$$\sum_{i\in\mathcal{I}_{11}} \sum_{j\in\mathcal{I}_{gs}\setminus\{1\}} \left|w_{ij} - \widetilde{w}_{ij}^{(1)}\right| \leq \sum_{i\in\mathcal{I}_{11}} w_{i1} = w_1$$

Subcase 2, Kernel matching,

$$\begin{aligned}
&\sum_{i\in\mathcal{I}_{11}} \sum_{j\in\mathcal{I}_{gs}\setminus\{1\}} \left|w_{ij} - \widetilde{w}_{ij}^{(1)}\right| \\
&= \sum_{i\in\mathcal{I}_{11}} \sum_{j\in\mathcal{I}_{gs}\setminus\{1\}} \left|\frac{K((X_i - X_j)/h)}{\sum_{r\in\mathcal{I}_{gs}} K((X_i - X_r)/h)} - \frac{K((X_i - X_j)/h)}{\sum_{r\in\mathcal{I}_{gs}} K((X_i - X_r)/h) - K((X_i - X_1)/h)}\right| \\
&= \sum_{i\in\mathcal{I}_{11}} \sum_{j\in\mathcal{I}_{gs}\setminus\{1\}} \frac{K((X_i - X_j)/h)K((X_i - X_1)/h)}{\sum_{r\in\mathcal{I}_{gs}} K((X_i - X_r)/h) \sum_{r\in\mathcal{I}_{gs}} K((X_i - X_r)/h) - K((X_i - X_1)/h)} \\
&= \sum_{i\in\mathcal{I}_{11}} \frac{K((X_i - X_1)/h)}{\sum_{r\in\mathcal{I}_{gs}} K((X_i - X_r)/h)} \\
&= \sum_{i\in\mathcal{I}_{11}} w_{i1} \\
&= w_1
\end{aligned}$$

Combine the results from the three cases,

$$\sum_{j\in\mathcal{I}_{gs}\setminus\{1\}} \left|w_j - \widetilde{w}_j^{(1)}\right| \leq w_1 + \mathbb{1}(1 \in \mathcal{I}_{11})$$

For the second term, notice that this case can be viewed as the symmetry case as the first term where we delete the observation $1'$ from the replaced sample, so by the same argument above, we have,

$$\sum_{j\in\mathcal{I}_{gs}\backslash\{1\}}\left|\widetilde{w}_j^{(1)}-w_j^{(1)}\right|=\sum_{j\in\mathcal{I}_{gs}\backslash\{1\}}\left|w_j^{(1)}-\widetilde{w}_j^{(1)}\right|\le w_{1'}+\mathbb{1}(1'\in\mathcal{I}_{11})$$

Therefore,

$$\sum_{j\in\mathcal{I}_{gs}\backslash\{1\}}\left|w_j-w_j^{(1)}\right|\le w_1+w_{1'}+\mathbb{1}(1\in\mathcal{I}_{11})+\mathbb{1}(1'\in\mathcal{I}_{11})$$

This proves the first claim.

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{j\in\mathcal{I}_{gs}\backslash\{1\}}\left|w_j-w_j^{(1)}\right|\right)^4\right]&\le\mathbb{E}\left[(w_1+w_{1'}+\mathbb{1}(1\in\mathcal{I}_{11})+\mathbb{1}(1'\in\mathcal{I}_{11}))^4\right]\\&\le 64\mathbb{E}\left[\left(w_1^4+w_{1'}^4+\mathbb{1}(1\in\mathcal{I}_{11})+\mathbb{1}(1'\in\mathcal{I}_{11})\right)\right]\\&=128(\mathbb{E}[w_1^4]+\mathbb{E}[\mathbb{1}(1\in\mathcal{I}_{11})])\\&=128(\mathbb{E}[w_1^4|1\in\mathcal{I}_{gs}]p_{gs}+p_{11})\\&=O(1)\end{aligned}$$

■

Proof. Proof of lemma 11.
Let $\xi>0$ be any fixed constant, we then pick a positive integer r (so r depends on ξ), such that $r\xi>1$, then

$$\begin{aligned}&P(n_{gs}^{-\xi}\max_{i\in\mathcal{I}_{gs}}n_{gs}^{1/q}\left\|X_i-X_{k_m(i)}\right\|>\epsilon|\mathbf{G},\mathbf{S})\\&\le\sum_{i\in\mathcal{I}_{gs}}P(n_{gs}^{-\xi}n_{gs}^{1/q}\left\|X_i-X_{k_m(i')}\right\|>\epsilon|\mathbf{G},\mathbf{S})\\&=\sum_{i=1}^{n}\mathbb{1}(i\in\mathcal{I}_{gs})P(n_{gs}^{-\xi}n_{gs}^{1/q}\left\|X_i-X_{k_m(i)}\right\|>\epsilon|\mathbf{G},\mathbf{S})\\&=n_{gs}P(n_{gs}^{-r\xi}n_{gs}^{r/q}\left\|X_i-X_{k_m(i)}\right\|^r>\epsilon^r|\mathbf{G},\mathbf{S},G_i=g,S_i=s)\\&\le\frac{\mathbb{E}[n_{gs}^{r/q}\left\|X_i-X_{k_m(i)}\right\|^r|\mathbf{G},\mathbf{S},G_i=g,S_i=s]}{\epsilon^r}\frac{n_{gs}}{n_{gs}^{r\xi}}\\&\le\frac{1}{n_{gs}^{r\xi-1}}\frac{C}{\epsilon^r}\end{aligned}$$

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 $n \to \infty$, $n_{gs}^{r\xi-1} \xrightarrow{a.s.} \infty$, , so $\frac{1}{n_{gs}^{r\xi-1}} \frac{C}{\epsilon^r} \xrightarrow{a.s.} 0$, as $n \to \infty$. This shows that:

$$P(n_{gs}^{-\xi} \max_{i \in \mathcal{I}_{gs}} n_{gs}^{1/q} \left\| X_i - X_{k_m(i)} \right\| > \epsilon | \mathbf{G}, \mathbf{S}) \xrightarrow{a.s.} 0$$

On the other hand, by law of iterated expectation

$$\begin{aligned} &P(n_{gs}^{-\xi} \max_{i \in \mathcal{I}_{gs}} n_{gs}^{1/q} \left\| X_i - X_{k_m(i)} \right\| > \epsilon) \\ &= \mathbb{E}\left[P(n_{gs}^{-\xi} \max_{i \in \mathcal{I}_{gs}} n_{gs}^{1/q} \left\| X_i - X_{k_m(i)} \right\| > \epsilon | \mathbf{G}, \mathbf{S}) \right] \end{aligned}$$

and combine with the fact that the conditional expectation is between $[0, 1]$, so by dominated convergence theorem,

$$n_{gs}^{-\xi} \max_{i \in \mathcal{I}_{gs}} n_{gs}^{1/q} \left\| X_i - X_{k_m(i)} \right\| = o_p(1)$$

The above equality holds for any fixed $\xi > 0$, specifically, pick $\xi = 1/q$, then:

$$\max_{i \in \mathcal{I}_{gs}} \left\| X_i - X_{k_m(i)} \right\| = o_p(1)$$

■

Proof. Proof of lemma 12.

$$\begin{aligned} \mathbb{E}\left[(\epsilon_{il}^{gs})^4 \middle| \mathbf{G}, \mathbf{S}, \mathbf{X}, G_i = g, S_i = s \right] &= \mathbb{E}\left[(\Delta Y_{il} - \mathbb{E}[\Delta Y_{il} | G_i = g, S_i = s, X_i])^4 \middle| G_i = g, S_i = s, X_i \right] \\ &\leq 8 \left(\mathbb{E}[\Delta Y_{il}^4 | G_i = g, S_i = s, X_i] + (\mathbb{E}[\Delta Y_{il} | G_i = g, S_i = s, X_i])^4 \right) \\ &< C \end{aligned}$$

The last inequality follows from assumption 6. ■

Proof. Proof of lemma 13.

By definition

$$
\begin{aligned}
\widehat{\tau}_l &= \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \Delta Y_{il} - \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{01}} w_i \Delta Y_{il} - \left(\frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{10}} w_i \Delta Y_{il} - \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{00}} w_i \Delta Y_{il} \right) \\
&= \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \Delta Y_{il} - \frac{1}{n_{11}} \sum_{j \in \mathcal{I}_{01}} w_j \Delta Y_{jl} - \left(\frac{1}{n_{11}} \sum_{j \in \mathcal{I}_{10}} w_j \Delta Y_{jl} - \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{00}} w_j \Delta Y_{jl} \right) \\
&= \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \Delta Y_{il} - \frac{1}{n_{11}} \sum_{j \in \mathcal{I}_{01}} \sum_{i \in \mathcal{I}_{11}} w_{ij} \Delta Y_{jl} - \frac{1}{n_{11}} \sum_{j \in \mathcal{I}_{10}} \sum_{i \in \mathcal{I}_{11}} w_{ij} \Delta Y_{jl} + \frac{1}{n_{11}} \sum_{j \in \mathcal{I}_{00}} \sum_{i \in \mathcal{I}_{11}} w_{ij} \Delta Y_{jl} \\
&= \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \Delta Y_{il} - \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \sum_{j \in \mathcal{I}_{01}} w_{ij} \Delta Y_{jl} - \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \sum_{j \in \mathcal{I}_{10}} w_{ij} \Delta Y_{jl} + \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \sum_{j \in \mathcal{I}_{00}} w_{ij} \Delta Y_{jl} \\
&= \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\Delta Y_{il} - \sum_{j \in \mathcal{I}_{01}} w_{ij} \Delta Y_{jl} - \sum_{j \in \mathcal{I}_{10}} w_{ij} \Delta Y_{jl} + \sum_{j \in \mathcal{I}_{00}} w_{ij} \Delta Y_{jl} \right) \\
&= \frac{1}{n_{11}} \sum_{i \in \mathcal{I}_{11}} \left(\Delta Y_{il} - \widehat{\Delta Y_{il}^{01}} - \widehat{\Delta Y_{il}^{10}} + \widehat{\Delta Y_{il}^{00}} \right)
\end{aligned}
$$

■