A Statistical Properties of RIPW Estimators

A.1 Setup and preliminaries

We will consider the more general setting in Section 3 since it nests the setting in Section 2. For ease of reference, we state the framework here, along with the list of assumptions some of which are weaker than those stated in the main text.

Suppose the i-th unit is characterized by potential outcomes $\boldsymbol{Y}_i(1) = (Y_{i1}(1), \ldots, Y_{iT}(1)), \boldsymbol{Y}_i(0) = (Y_{i1}(0), \ldots, Y_{iT}(0))$, the treatment path $\boldsymbol{W}_i = (W_{i1}, \ldots, W_{iT})$, and a set of covariates $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{iT})$. The vector of time-varying treatment effects for unit i is denoted by $\boldsymbol{\tau}_i = (\tau_{i1}, \ldots, \tau_{iT}) = \boldsymbol{Y}_i(1) - \boldsymbol{Y}_i(0)$. We treat covariates as fixed and consider $\{(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{W}_i) : i \in [n]\}$ as a random vector (jointly) drawn from a distribution (conditional on $\{\boldsymbol{X}_i : i \in [n]\}$). We let $\mathbb{P}$ denote the joint distribution of the entire random vector $\{(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{W}_i) : i \in [n]\}$ (conditional on $\{\boldsymbol{X}_i : i \in [n]\}$) and $\mathbb{E}$ denote the expectation over this distribution.

The assignment model is characterized by the generalized propensity score defined as

$$\boldsymbol{\pi}_i(\boldsymbol{w}) = \mathbb{P}(\boldsymbol{W}_i = \boldsymbol{w}).$$

The outcome model is characterized by $\{(\boldsymbol{m}_i, \boldsymbol{\nu}_i) : i \in [n]\}$ where $\boldsymbol{m}_i = (m_{i1}, \ldots, m_{iT}), \boldsymbol{\nu}_i = (\nu_{i1}, \ldots, \nu_{iT})$,

$$\begin{aligned}
m_{it} &= \mathbb{E}[Y_{it}(0)] - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[Y_{it}(0)] - \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[Y_{it}(0)] + \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{E}[Y_{it}(0)],\\
\tau_{it} &= \mathbb{E}[Y_{it}(1)] - \mathbb{E}[Y_{it}(0)],\\
\nu_{it} &= \tau_{it} - \tau^*(\xi).
\end{aligned}$$

Let $\{(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{\mu}}_i(0), \hat{\boldsymbol{\mu}}_i(1)) : i \in [n]\}$ be an estimate of $\{(\boldsymbol{\pi}_i, \boldsymbol{\mu}_i(0), \boldsymbol{\mu}_i(1))\}$. Further let $\hat{\boldsymbol{m}}_i = (\hat{m}_{i1}, \ldots, \hat{m}_{iT})$ and $\hat{\boldsymbol{\nu}}_i = (\hat{\nu}_{i1}, \ldots, \hat{\nu}_{iT})$, where

$$\begin{aligned}
\hat{m}_{it} &\triangleq \hat{\mu}_{it}(0) - \frac{1}{n}\sum_{i=1}^{n}\hat{\mu}_{it}(0) - \frac{1}{T}\sum_{t=1}^{T}\hat{\mu}_{it}(0) + \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\mu}_{it}(0),\\
\hat{\tau}_{it} &\triangleq \hat{\mu}_{it}(1) - \hat{\mu}_{it}(0),
\end{aligned}$$

$$\hat{\nu}_{it} \triangleq \hat{\tau}_{it} - \sum_{t=1}^{T} \frac{\xi_t}{n} \sum_{i=1}^{n} \hat{\tau}_{it}.$$

The results in Section 2 are given by the special case where $\hat{\boldsymbol{\pi}}_i = \boldsymbol{\pi}_i, \hat{\boldsymbol{m}}_i = \hat{\boldsymbol{\nu}}_i = \mathbf{0}_T$.

Define the modified potential outcomes as

$$\tilde{\boldsymbol{Y}}_i(0) = \boldsymbol{Y}_i(0) - \hat{\boldsymbol{m}}_i, \quad \tilde{\boldsymbol{Y}}_i(1) = \boldsymbol{Y}_i(1) - \hat{\boldsymbol{m}}_i - \hat{\boldsymbol{\nu}}_i,$$

and the modified treatment effects as

$$\tilde{\boldsymbol{\tau}}_i = \mathbb{E}[\tilde{\boldsymbol{Y}}_i(1) - \tilde{\boldsymbol{Y}}_i(0)] = \boldsymbol{\tau}_i - \mathbb{E}[\hat{\boldsymbol{\nu}}_i]. \tag{A.1}$$

By definition,

$$\tau^*(\xi) = \sum_{t=1}^{T} \frac{\xi_t}{n} \sum_{i=1}^{n} \tau_{it} = \sum_{t=1}^{T} \frac{\xi_t}{n} \sum_{i=1}^{n} \tilde{\tau}_{it}. \tag{A.2}$$

Then the modified observed outcome is $\tilde{\boldsymbol{Y}}_i = (\tilde{Y}_{i1}, \ldots, \tilde{Y}_{iT})$ where

$$\tilde{Y}_{it} = \tilde{Y}_{it}(1)W_{it} + \tilde{Y}_{it}(0)(1 - W_{it}) = Y_{it} - \hat{m}_{it} - \hat{\nu}_{it}W_{it}.$$

With a reshaped distribution $\boldsymbol{\Pi}$ on $\{0,1\}^T$, the RIPW estimator is defined as

$$\hat{\tau}(\boldsymbol{\Pi}) \triangleq \underset{\tau,\mu,\sum_i \alpha_i = \sum_t \lambda_t = 0}{\arg\min} \sum_{i=1}^{n} \sum_{t=1}^{T} (\tilde{Y}_{it} - \mu - \alpha_i - \lambda_t - W_{it}\tau)^2 \frac{\boldsymbol{\Pi}(\boldsymbol{W}_i)}{\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i)}.$$

We will suppress ξ from $\tau^*(\xi)$ and $\boldsymbol{\Pi}$ from $\hat{\tau}(\boldsymbol{\Pi})$ throughout the section.

Since $\hat{\tau}(\boldsymbol{\Pi})$ remains invariant if we replace Y_{it} by $Y_{it} - \mu' - \alpha'_i - \lambda'_t$, we assume that

$$\boldsymbol{Y}_i(0) = \boldsymbol{m}_i \Longleftrightarrow \tilde{\boldsymbol{Y}}_i(0) = \boldsymbol{m}_i - \hat{\boldsymbol{m}}_i, \tag{A.3}$$

by setting $\mu' = (1/nT)\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{E}[Y_{it}(0)], \alpha'_i = (1/T)\sum_{t=1}^{T}(\mathbb{E}[Y_{it}(0)] - \mu')$, and $\lambda'_t = (1/n)\sum_{i=1}^{n}(\mathbb{E}[Y_{it}(0)] - \mu')$.

The accuracy of the assignment model and the outcome model for unit i are defined as

$$\delta_{\pi i} = \sqrt{\mathbb{E}[|\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) - \boldsymbol{\pi}_i(\boldsymbol{W}_i)|^2]}, \quad \delta_{yi} = \sqrt{\mathbb{E}[\|\hat{\boldsymbol{m}}_i - \boldsymbol{m}_i\|_2^2 + \|\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i\|_2^2]}.$$

In the proofs, we need the conditional version of these measures

$$\Delta_{\pi i} = \sqrt{\mathbb{E}\left[|\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) - \boldsymbol{\pi}_i(\boldsymbol{W}_i)|^2 \mid \hat{\boldsymbol{\pi}}_i\right]^2}, \quad \Delta_{yi} = \sqrt{\mathbb{E}[\|\hat{\boldsymbol{m}}_i - \boldsymbol{m}_i\|_2^2 \mid \hat{\boldsymbol{m}}_i] + \mathbb{E}[\|\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i\|_2^2 \mid \hat{\boldsymbol{\nu}}_i]}.$$

We then define the unconditional and conditional average accuracy measures

$$\bar{\delta}_{\pi} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \delta_{\pi i}^2}, \quad \bar{\delta}_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \delta_{yi}^2},$$

and

$$\bar{\Delta}_{\pi} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \Delta_{\pi i}^2}, \quad \bar{\Delta}_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \Delta_{yi}^2}.$$

By law of iterated expectations,

$$\mathbb{E}[\bar{\Delta}_{\pi}^2] = \bar{\delta}_{\pi}^2, \quad \mathbb{E}[\bar{\Delta}_y^2] = \bar{\delta}_y^2.$$

By Markov inequality,

$$\bar{\Delta}_{\pi} = O_{\mathbb{P}}(\bar{\delta}_{\pi}), \quad \bar{\Delta}_y = O_{\mathbb{P}}(\bar{\delta}_y). \tag{A.4}$$

Therefore, if we can prove the result only assuming $\bar{\Delta}_{\pi}\bar{\Delta}_y = o(1)$ conditional on $(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i)_{i=1}^n$, we can prove it assuming that $\bar{\delta}_{\pi}\bar{\delta}_y = o(1)$ as in Section 3.

To be self-contained, we list all quantities involved in the DATE equationand the asymptotically linear expansion of the RIPW estimator. Let $J = I_T - \mathbf{1}_T\mathbf{1}_T^\top/T$,

$$\Theta_i = \boldsymbol{\Pi}(\boldsymbol{W}_i)/\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i),$$

$$\Gamma_\theta \triangleq \frac{1}{n}\sum_{i=1}^{n}\Theta_i, \quad \Gamma_{ww} \triangleq \frac{1}{n}\sum_{i=1}^{n}\Theta_i \boldsymbol{W}_i^\top J \boldsymbol{W}_i, \quad \Gamma_{wy} \triangleq \frac{1}{n}\sum_{i=1}^{n}\Theta_i \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i,$$

$$\boldsymbol{\Gamma}_w \triangleq \frac{1}{n}\sum_{i=1}^{n}\Theta_i J \boldsymbol{W}_i, \quad \boldsymbol{\Gamma}_y \triangleq \frac{1}{n}\sum_{i=1}^{n}\Theta_i J \tilde{\boldsymbol{Y}}_i,$$

and

$$\begin{aligned}\mathcal{V}_i = \Theta_i \Big\{ & (\mathbb{E}[\Gamma_{wy}] - \tau^*\mathbb{E}[\Gamma_{ww}]) - (\mathbb{E}[\boldsymbol{\Gamma}_y] - \tau^*\mathbb{E}[\boldsymbol{\Gamma}_w])^\top J \boldsymbol{W}_i \\ & + \mathbb{E}[\Gamma_\theta]\boldsymbol{W}_i^\top J\left(\tilde{\boldsymbol{Y}}_i - \tau^*\boldsymbol{W}_i\right) - \mathbb{E}[\boldsymbol{\Gamma}_w]^\top J\left(\tilde{\boldsymbol{Y}}_i - \tau^*\boldsymbol{W}_i\right)\Big\}.\end{aligned}$$

This coincides with the definition in Theorem 2.2 when $\hat{\boldsymbol{\pi}}_i = \boldsymbol{\pi}_i$ and $\hat{\boldsymbol{m}}_i = \hat{\boldsymbol{\nu}}_i = \mathbf{0}_T$.

Finally, we state the core assumptions, some of which are repeated and combined for ease of reference and the rest of which are weakened. We start by restating the unit-specific mean ignorability assumption.

Assumption A.1. *For each $i \in [n]$,*

$$\mathbb{E}[(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0)) \mid \boldsymbol{W}_i] = \mathbb{E}[(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0))]. \tag{A.5}$$

Next, we combine the overlap condition for the true propensity scores (Assumption 2.2) and that for the estimated propensity scores (Assumption 3.2) with the constant c replaced by c_π to be more informative in the proofs.

Assumption A.2. *There exists a universal constant $c > 0$ and a non-stochastic subset $\mathbb{S}^* \subset \{0,1\}^T$ with at least two elements and at least one element not in $\{\boldsymbol{0}_T, \boldsymbol{1}_T\}$, such that*

$$\hat{\boldsymbol{\pi}}_i(\boldsymbol{w}) > c_\pi, \boldsymbol{\pi}_i(\boldsymbol{w}) > c_\pi, \quad \forall \boldsymbol{w} \in \mathbb{S}^*, i \in [n], \quad \textit{almost surely.} \tag{A.6}$$

Lastly, we state the following assumption that unifies and weakens Assumptions 2.1, 2.3, and 3.3.

Assumption A.3. *There exists* $q \in (0,1]$,

$$\frac{1}{n^2}\sum_{i=1}^{n}\rho_i\left\{\mathbb{E}\|\tilde{\boldsymbol{Y}}_i(1)\|_2^2 + \mathbb{E}\|\tilde{\boldsymbol{Y}}_i(0)\|_2^2 + 1\right\} = O(n^{-q}),$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{E}\|\tilde{\boldsymbol{Y}}_i(1)\|_2^2 + \mathbb{E}\|\tilde{\boldsymbol{Y}}_i(0)\|_2^2\right\} = O(1).$$

We close this section by a basic property of the maximal correlation.

Lemma A.1. *Let* $\mathcal{Z}_i = (\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{X}_i)$ *and* f_i *be any deterministic function on the domain of* $\mathcal{Z}_i$. *Then*

$$\mathrm{Var}\left[\sum_{i=1}^{n} f_i(\mathcal{Z}_i)\right] \le \frac{1}{2}\sum_{i=1}^{n}\mathrm{Var}[f_i(\mathcal{Z}_i)]\rho_i.$$

Proof. By definition of ρ_{ij},

$$\mathrm{Cov}\left(f_i(\mathcal{Z}_i), f_j(\mathcal{Z}_j)\right) \le \rho_{ij}\sqrt{\mathrm{Var}[f_i(\mathcal{Z}_i)]\mathrm{Var}[f_j(\mathcal{Z}_j)]} \le \frac{\rho_{ij}}{2}\left\{\mathrm{Var}[f_i(\mathcal{Z}_i)] + \mathrm{Var}[f_j(\mathcal{Z}_j)]\right\}.$$

Thus,

$$\mathrm{Var}\left[\sum_{i=1}^{n} f_i(\mathcal{Z}_i)\right] = \sum_{i,j=1}^{n}\mathrm{Cov}(f_i(\mathcal{Z}_i), f_j(\mathcal{Z}_j))$$
$$\le \sum_{i,j=1}^{n}\frac{\rho_{ij}}{2}\left\{\mathrm{Var}[f_i(\mathcal{Z}_i)] + \mathrm{Var}[f_j(\mathcal{Z}_j)]\right\} = \sum_{i=1}^{n}\mathrm{Var}[f_i(\mathcal{Z}_i)]\rho_i.$$

□

A.2 A non-stochastic formula of RIPW estimators

Theorem A.1. *With the same notation as Theorem 2.2,* $\hat{\tau} = \mathcal{N}/\mathcal{D}$, *where*

$$\mathcal{N} = \Gamma_{wy}\Gamma_\theta - \boldsymbol{\Gamma}_w^\top\boldsymbol{\Gamma}_y, \quad \mathcal{D} = \Gamma_{ww}\Gamma_\theta - \boldsymbol{\Gamma}_w^\top\boldsymbol{\Gamma}_w. \tag{A.7}$$

Proof. Let $\boldsymbol{\gamma} = (\lambda_1, \ldots, \lambda_t)$ be any vector with $\boldsymbol{\gamma}^\top \mathbf{1}_T = 0$. First we derive the optimum $\hat{\mu}(\boldsymbol{\gamma}, \tau), \hat{\alpha}_i(\boldsymbol{\gamma}, \tau)$ given any values of $\boldsymbol{\gamma}$ and τ. Recall that

$$(\hat{\mu}(\boldsymbol{\gamma}, \tau), \hat{\alpha}_i(\boldsymbol{\gamma}, \tau)) = \operatorname*{arg\,min}_{\sum_i \alpha_i = 0} \sum_{i=1}^n \left(\sum_{t=1}^T (\tilde{Y}_{it} - \mu - \alpha_i - \lambda_t - W_{it}\tau)^2 \right) \Theta_i.$$

Since the weight Θ_i only depends on i, it is easy to see that

$$\hat{\mu}(\boldsymbol{\gamma}, \tau) + \hat{\alpha}_i(\boldsymbol{\gamma}, \tau) = \frac{1}{T} \sum_{t=1}^T (\tilde{Y}_{it} - \lambda_t - W_{it}\tau), \quad \hat{\mu}(\boldsymbol{\gamma}, \tau) = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (\tilde{Y}_{it} - \lambda_t - W_{it}\tau).$$

As a result,

$$\begin{aligned}
&\sum_{t=1}^T (\tilde{Y}_{it} - \hat{\mu}(\boldsymbol{\gamma}, \mu) - \hat{\alpha}_i(\boldsymbol{\gamma}, \mu) - \lambda_t - W_{it}\tau)^2 \\
=&\left\| \left(\tilde{\boldsymbol{Y}}_i - \boldsymbol{\gamma} - \boldsymbol{W}_i \tau \right) - \frac{\mathbf{1}_T \mathbf{1}_T^\top}{T} \left(\tilde{\boldsymbol{Y}}_i - \boldsymbol{\gamma} - \boldsymbol{W}_i \tau \right) \right\|_2^2 \\
=&\left\| J \left(\tilde{\boldsymbol{Y}}_i - \boldsymbol{\gamma} - \boldsymbol{W}_i \tau \right) \right\|_2^2.
\end{aligned}$$

This yields a profile loss function for $\boldsymbol{\gamma}$ and τ:

$$(\hat{\boldsymbol{\gamma}}, \hat{\tau}) = \operatorname*{arg\,min}_{\boldsymbol{\gamma}^\top \mathbf{1}_T = 0} \sum_{i=1}^n \left\| J \left(\tilde{\boldsymbol{Y}}_i - \boldsymbol{\gamma} - \boldsymbol{W}_i \tau \right) \right\|_2^2 \Theta_i = \operatorname*{arg\,min}_{\boldsymbol{\gamma}^\top \mathbf{1}_T = 0} \sum_{i=1}^n \left\| J \left(\tilde{\boldsymbol{Y}}_i - \boldsymbol{W}_i \tau \right) - \boldsymbol{\gamma} \right\|_2^2 \Theta_i,$$

where the last equality uses the fact that $J\boldsymbol{\gamma} = \boldsymbol{\gamma}$. Given τ, the optimizer $\hat{\boldsymbol{\gamma}}(\tau)$ is simply the weighted average of $\{J(\tilde{\boldsymbol{Y}}_i - \boldsymbol{W}_i\tau)\}_{i=1}^n$ in absence of the constraint $\boldsymbol{\gamma}^\top \mathbf{1}_T = 0$, i.e.

$$\hat{\boldsymbol{\gamma}}(\tau) = \frac{\sum_{i=1}^n \Theta_i J(\tilde{\boldsymbol{Y}}_i - \boldsymbol{W}_i\tau)}{\sum_{i=1}^n \Theta_i} = \frac{\boldsymbol{\Gamma}_y}{\Gamma_\theta} - \frac{\boldsymbol{\Gamma}_w}{\Gamma_\theta}\tau.$$

Noting that $\hat{\boldsymbol{\gamma}}(\tau)^\top \mathbf{1}_T = 0$ since $J\mathbf{1}_T = 0$, $\hat{\boldsymbol{\gamma}}(\tau)$ is also the minimizer of the constrained problem, i.e.

$$\hat{\boldsymbol{\gamma}}(\tau) = \operatorname*{arg\,min}_{\boldsymbol{\gamma}^\top \mathbf{1}_T = 0} \sum_{i=1}^n \left\| J \left(\tilde{\boldsymbol{Y}}_i - \boldsymbol{W}_i \tau \right) - \boldsymbol{\gamma} \right\|_2^2 \Theta_i.$$

Plugging in $\hat{\boldsymbol{\gamma}}(\tau)$ yields a profile loss function for τ

$$\hat{\tau} = \arg\min \sum_{i=1}^{n} \left\| J\left(\tilde{\boldsymbol{Y}}_i - \boldsymbol{W}_i\tau\right) - \hat{\boldsymbol{\gamma}}(\tau) \right\|_2^2 \Theta_i \triangleq L(\tau).$$

A direct calculation shows that

$$\begin{aligned}
\frac{L'(\tau)}{2n} =& \frac{1}{n}\sum_{i=1}^{n}\Theta_i\left(-J\boldsymbol{W}_i + \frac{\boldsymbol{\Gamma}_w}{\Gamma_\theta}\right)^\top\left(J\left(\tilde{\boldsymbol{Y}}_i - \boldsymbol{W}_i\tau\right) - \frac{\boldsymbol{\Gamma}_y}{\Gamma_\theta} + \frac{\boldsymbol{\Gamma}_w}{\Gamma_\theta}\tau\right) \\
=& \frac{1}{n}\left\{\sum_{i=1}^{n}\Theta_i\left(J\boldsymbol{W}_i - \frac{\boldsymbol{\Gamma}_w}{\Gamma_\theta}\right)^\top\left(J\boldsymbol{W}_i - \frac{\boldsymbol{\Gamma}_w}{\Gamma_\theta}\right)\right\}\tau - \frac{1}{n}\left\{\sum_{i=1}^{n}\Theta_i\left(J\boldsymbol{W}_i - \frac{\boldsymbol{\Gamma}_w}{\Gamma_\theta}\right)^\top\left(J\tilde{\boldsymbol{Y}}_i - \frac{\boldsymbol{\Gamma}_y}{\Gamma_\theta}\right)\right\} \\
=& \left\{\Gamma_{ww} - \frac{\boldsymbol{\Gamma}_w^\top\boldsymbol{\Gamma}_w}{\Gamma_\theta}\right\}\tau - \left\{\Gamma_{wy} - \frac{\boldsymbol{\Gamma}_w^\top\boldsymbol{\Gamma}_y}{\Gamma_\theta}\right\}
\end{aligned}$$

Since $L(\tau)$ is a convex quadratic function of τ, the first-order condition is sufficient and necessary to determine the optimality. The proof is then completed by solving $L'(\hat{\tau}) = 0$. □

A.3 Statistical properties of RIPW estimators with deterministic $(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i)$

A.3.1 Asymptotic linear expansion of RIPW estimators

As a warm-up, we assume that $(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i)_{i=1}^n$ are deterministic. This, for example, includes the pure design-based inference where $\hat{\boldsymbol{\pi}}_i = \boldsymbol{\pi}_i$ and $\hat{\boldsymbol{m}}_i = \hat{\boldsymbol{\nu}}_i = 0$. In this case, the measures of accuracy can be simplified as

$$\Delta_{\pi i} = \sqrt{\mathbb{E}\left[\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) - \boldsymbol{\pi}_i(\boldsymbol{W}_i)\right]^2}, \quad \Delta_{yi} = \sqrt{\|\hat{\boldsymbol{m}}_i - \boldsymbol{m}_i\|_2^2 + \|\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i\|_2^2}. \tag{A.8}$$

As a result, $(\Delta_{\pi i}, \Delta_{yi})$ are deterministic.

We start by a lemma showing that $\Gamma_\theta, \Gamma_{wy}, \Gamma_{ww}, \boldsymbol{\Gamma}_w, \boldsymbol{\Gamma}_y$ concentrate around their means. For notational convenience, we let $\mathrm{Var}(Z)$ denote $\mathbb{E}\|Z - \mathbb{E}[Z]\|_2^2$ for a random vector Z.

Lemma A.2. *Under Assumptions A.2 and A.3,*

$$|\mathbb{E}[\Gamma_\theta]| + |\mathbb{E}[\Gamma_{wy}]| + |\mathbb{E}[\Gamma_{ww}]| + \|\mathbb{E}[\boldsymbol{\Gamma}_w]\|_2 + \|\mathbb{E}[\boldsymbol{\Gamma}_y]\|_2 = O(1),$$

and

$$\mathrm{Var}(\Gamma_\theta) + \mathrm{Var}(\Gamma_{wy}) + \mathrm{Var}(\Gamma_{ww}) + \mathrm{Var}(\boldsymbol{\Gamma}_w) + \mathrm{Var}(\boldsymbol{\Gamma}_y) = O(n^{-q}).$$

As a consequence,

$$\left|\Gamma_\theta - \mathbb{E}[\Gamma_\theta]\right| + \left|\Gamma_{wy} - \mathbb{E}[\Gamma_{wy}]\right| + \left|\Gamma_{ww} - \mathbb{E}[\Gamma_{ww}]\right| + \left\|\boldsymbol{\Gamma}_w - \mathbb{E}[\boldsymbol{\Gamma}_w]\right\|_2 + \left\|\boldsymbol{\Gamma}_y - \mathbb{E}[\boldsymbol{\Gamma}_y]\right\|_2 = O_{\mathbb{P}}\left(n^{-q/2}\right).$$

Proof. By Assumption A.2, $\Theta_i \leq 1/c_\pi$ almost surely. Moreover, $\|\boldsymbol{W}_i\|_2 \leq \sqrt{T}$ since $W_{it} \in \{0, 1\}$. Thus,

$$\|\boldsymbol{\Gamma}_w\|_2 \leq \frac{\sqrt{T}}{c_\pi}, \quad |\Gamma_{ww}| \leq \frac{T}{c_\pi}, \quad |\Gamma_\theta| \leq \frac{1}{c_\pi} \Longrightarrow \mathbb{E}\|\boldsymbol{\Gamma}_w\|_2 + \mathbb{E}|\Gamma_{ww}| + \mathbb{E}|\Gamma_\theta| = O(1).$$

Next, we derive bounds for $(\mathbb{E}[\Gamma_{wy}])^2$ and $\|\mathbb{E}[\boldsymbol{\Gamma}_y]\|_2^2$ separately. For $(\mathbb{E}[\Gamma_{wy}])^2$,

$$\begin{aligned}
(\mathbb{E}[\Gamma_{wy}])^2 &\leq \left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}[\Theta_i \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i]\right)^2 \leq \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\Theta_i \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i]^2 \\
&\leq \frac{1}{nc_\pi^2}\sum_{i=1}^n \mathbb{E}[\boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i]^2 \leq \frac{T}{nc_\pi^2}\sum_{i=1}^n \mathbb{E}\|\tilde{\boldsymbol{Y}}_i\|_2^2 \\
&\leq \frac{T}{nc_\pi^2}\sum_{i=1}^n \left\{\mathbb{E}\|\tilde{\boldsymbol{Y}}_i(0)\|_2^2 + \mathbb{E}\|\tilde{\boldsymbol{Y}}_i(1)\|_2^2\right\} \\
&= O(1),
\end{aligned}$$

where the last step follows from the Assumption A.3. For $\|\mathbb{E}[\boldsymbol{\Gamma}_y]\|_2^2$,

$$\begin{aligned}
\|\mathbb{E}[\boldsymbol{\Gamma}_y]\|_2^2 &\leq \frac{1}{n}\sum_{i=1}^n (\mathbb{E}[\Theta_i J \tilde{\boldsymbol{Y}}_i])^2 \leq \frac{1}{nc_\pi^2}\sum_{i=1}^n \|\tilde{\boldsymbol{Y}}_i\|_2^2 \\
&\leq \frac{1}{nc_\pi^2}\sum_{i=1}^n \left\{\mathbb{E}\|\tilde{\boldsymbol{Y}}_i(0)\|_2^2 + \mathbb{E}\|\tilde{\boldsymbol{Y}}_i(1)\|_2^2\right\} \\
&= O(1),
\end{aligned}$$

where the last step follows from the Assumption A.3. Putting the pieces together, the bound on the sum of expectations is proved.

Next, we turn to the bound on the variances. By Lemma A.1,

$$\mathrm{Var}(\Gamma_\theta) \leq \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(\Theta_i)\rho_i \leq \frac{1}{n^2 c_\pi^2}\sum_{i=1}^{n}\rho_i.$$

The Assumption A.2 implies that

$$\frac{1}{n^2}\sum_{i=1}^{n}\rho_i = O(n^{-q}).$$

Therefore, $\mathrm{Var}(\Gamma_\theta) = O(n^{-q})$. For Γ_{ww},

$$\begin{aligned}\mathrm{Var}(\Gamma_{ww}) &\leq \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(\Theta_i \boldsymbol{W}_i^\top J \boldsymbol{W}_i)\rho_i \leq \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}(\Theta_i \boldsymbol{W}_i^\top J \boldsymbol{W}_i)^2\rho_i \\ &\leq \frac{1}{n^2 c_\pi^2}\sum_{i=1}^{n}\mathbb{E}\|\boldsymbol{W}_i\|_2^2\rho_i \leq \frac{T}{n^2 c_\pi^2}\sum_{i=1}^{n}\rho_i = O(n^{-q}),\end{aligned}$$

where the last equality uses the fact that $\|\boldsymbol{W}_i\|_2 \leq \sqrt{T}$. For Γ_{wy},

$$\begin{aligned}\mathrm{Var}(\Gamma_{wy}) &\leq \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(\Theta_i \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i)\rho_i \leq \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}(\Theta_i \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i)^2\rho_i \\ &\overset{(i)}{\leq} \frac{1}{n^2 c_\pi^2}\sum_{i=1}^{n}\mathbb{E}\left[\|\boldsymbol{W}_i\|_2^2 \cdot \|\tilde{\boldsymbol{Y}}_i\|_2^2\right]\rho_i \\ &\overset{(ii)}{\leq} \frac{T}{n^2 c_\pi^2}\sum_{i=1}^{n}\left(\mathbb{E}\|\tilde{\boldsymbol{Y}}_i(1)\|_2^2 + \mathbb{E}\|\tilde{\boldsymbol{Y}}_i(0)\|_2^2\right)\rho_i \\ &\overset{(iii)}{=} O(n^{-q}),\end{aligned}$$

where (i) follows from the Cauchy-Schwarz inequality and that $\|J\|_{\mathrm{op}} = 1$, (ii) is obtained from the fact that $\|\boldsymbol{W}_i\|_2^2 \leq T$ and $\tilde{\boldsymbol{Y}}_i \in \{\tilde{\boldsymbol{Y}}_i(1), \tilde{\boldsymbol{Y}}_i(0)\}$, and (iii) follows from the Assumption A.3.

For $\boldsymbol{\Gamma}_w$, recall that $\mathrm{Var}(\boldsymbol{\Gamma}_w)$ is the sum of the variance of each coordinate of $\boldsymbol{\Gamma}_w$. By Lemma

A.1,

$$\begin{aligned}\mathrm{Var}(\boldsymbol{\Gamma}_w) &\leq \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(\Theta_i J\boldsymbol{W}_i)\rho_i \leq \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}\|\Theta_i J\boldsymbol{W}_i\|_2^2\rho_i \\ &\leq \frac{1}{n^2 c_\pi^2}\sum_{i=1}^{n}\mathbb{E}\|\boldsymbol{W}_i\|_2^2\rho_i \leq \frac{T}{n^2 c_\pi^2}\sum_{i=1}^{n}\rho_i = O(n^{-q}).\end{aligned}$$

For $\boldsymbol{\Gamma}_y$, analogues to inequalities (i) - (iii) for Γ_{wy}, we obtain that

$$\begin{aligned}\mathrm{Var}(\boldsymbol{\Gamma}_y) &\leq \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(\Theta_i J\tilde{\boldsymbol{Y}}_i)\rho_i \leq \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}\|\Theta_i J\tilde{\boldsymbol{Y}}_i\|_2^2\rho_i \\ &\leq \frac{1}{n^2 c_\pi^2}\sum_{i=1}^{n}\left(\mathbb{E}[\|\tilde{\boldsymbol{Y}}_i(1)\|_2^2] + \mathbb{E}[\|\tilde{\boldsymbol{Y}}_i(0)\|_2^2]\right)\rho_i = O(n^{-q}),\end{aligned}$$

where the last step follows from the Assumption A.3.

Finally, by Markov's inequality,

$$\begin{aligned}&\left|\Gamma_\theta - \mathbb{E}[\Gamma_\theta]\right| + \left|\Gamma_{wy} - \mathbb{E}[\Gamma_{wy}]\right| + \left|\Gamma_{ww} - \mathbb{E}[\Gamma_{ww}]\right| + \left\|\boldsymbol{\Gamma}_w - \mathbb{E}[\boldsymbol{\Gamma}_w]\right\|_2 + \left\|\boldsymbol{\Gamma}_y - \mathbb{E}[\boldsymbol{\Gamma}_y]\right\|_2 \\ =&O_{\mathbb{P}}\left(\sqrt{\mathrm{Var}(\Gamma_\theta) + \mathrm{Var}(\Gamma_{wy}) + \mathrm{Var}(\Gamma_{ww}) + \mathrm{Var}(\boldsymbol{\Gamma}_w) + \mathrm{Var}(\boldsymbol{\Gamma}_y)}\right) = O_{\mathbb{P}}(n^{-q/2}).\end{aligned}$$

□

The following lemma shows that the denominator of $\hat{\tau}$ is bounded away from 0.

Lemma A.3. *Under Assumptions A.3, regardless of the dependence between $(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i)$ and the data,*

$$\mathcal{D} \geq c_{\mathcal{D}}^2\left(\frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{W}_i = \boldsymbol{w}_1)\right)\left(\frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{W}_i = \boldsymbol{w}_2)\right),$$

for some constant $c_{\mathcal{D}}$ that only depends on $\boldsymbol{\Pi}$. As a result, $\mathcal{D} \geq 0$ almost surely. If Assumption

A.2 *also holds,* [6]

$$\mathbb{E}[\mathcal{D}] \geq c_{\mathcal{D}}^2 \left(c_\pi^2 - \frac{1}{n^2} \sum_{i=1}^n \rho_i \right), \quad \mathcal{D} \geq c_{\mathcal{D}}^2 c_\pi^2 - o_{\mathbb{P}}(1).$$

Proof. By definition,

$$\begin{aligned}
\mathcal{D} &= \left(\frac{1}{n} \sum_{i=1}^n \Theta_i \tilde{\boldsymbol{W}}_i^\top J \tilde{\boldsymbol{W}}_i \right) \left(\frac{1}{n} \sum_{i=1}^n \Theta_i \right) - \left\| \frac{1}{n} \sum_{i=1}^n \Theta_i J \tilde{\boldsymbol{W}}_i \right\|_2^2 \\
&= \frac{1}{n^2} \sum_{i,j=1}^n \Theta_i \Theta_j \left(\tilde{\boldsymbol{W}}_i^\top J \tilde{\boldsymbol{W}}_i + \tilde{\boldsymbol{W}}_j^\top J \tilde{\boldsymbol{W}}_j - 2 \tilde{\boldsymbol{W}}_i^\top J \tilde{\boldsymbol{W}}_j \right) \\
&= \frac{1}{n^2} \sum_{i,j=1}^n \Theta_i \Theta_j \| J(\boldsymbol{W}_i - \boldsymbol{W}_j) \|_2^2.
\end{aligned}$$

Let $\boldsymbol{w}_1, \boldsymbol{w}_2$ be two distinct elements from $\mathbb{S}^*$ with $\boldsymbol{w}_1 \notin \{\mathbf{0}_T, \mathbf{1}_T\}$ and

$$\frac{1}{n} \sum_{i=1}^n \boldsymbol{\pi}_i(\boldsymbol{w}_k) > c_\pi, \quad k \in \{1, 2\}. \tag{A.9}$$

This is enabled by Assumption A.2. Note that $J(\boldsymbol{w}_1 - \boldsymbol{w}_2) = 0$ iff $\boldsymbol{w}_1 - \boldsymbol{w}_2 = a\mathbf{1}_T$ for some $a \in \mathbb{R}$, which is impossible since $\boldsymbol{w}_1 \notin \{\mathbf{0}_T, \mathbf{1}_T\}$ and all entries of $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are binary. In addition, since $\boldsymbol{\Pi}$ has support $\mathbb{S}^*$, $\boldsymbol{\Pi}(\boldsymbol{w}_1), \boldsymbol{\Pi}(\boldsymbol{w}_2) > 0$. Let

$$c_{\mathcal{D}} = \min\{\boldsymbol{\Pi}(\boldsymbol{w}_1), \boldsymbol{\Pi}(\boldsymbol{w}_2)\} \| J(\boldsymbol{w}_1 - \boldsymbol{w}_2) \|_2 > 0.$$

Then

$$\begin{aligned}
\mathcal{D} &\geq \frac{c_{\mathcal{D}}^2}{n^2} \sum_{i,j=1}^n \frac{1}{\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) \hat{\boldsymbol{\pi}}_j(\boldsymbol{W}_j)} I(\boldsymbol{W}_i = \boldsymbol{w}_1, \boldsymbol{W}_j = \boldsymbol{w}_2) \\
&\geq \frac{c_{\mathcal{D}}^2}{n^2} \sum_{i,j=1}^n I(\boldsymbol{W}_i = \boldsymbol{w}_1, \boldsymbol{W}_j = \boldsymbol{w}_2) \\
&= c_{\mathcal{D}}^2 \left(\frac{1}{n} \sum_{i=1}^n I(\boldsymbol{W}_i = \boldsymbol{w}_1) \right) \left(\frac{1}{n} \sum_{i=1}^n I(\boldsymbol{W}_i = \boldsymbol{w}_2) \right),
\end{aligned}$$

[6] A more rigorous version of the second statement is $\max\{c_{\mathcal{D}}^2 c_\pi^2 - \mathcal{D}, 0\} = o_{\mathbb{P}}(1)$

where the second inequality follows from the fact that $\hat{\boldsymbol{\pi}}_i(\boldsymbol{w}) \leq 1$. By (A.9),

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{W}_i = \boldsymbol{w}_k)\right] = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\pi}_i(\boldsymbol{w}_k) > c_\pi, \quad k \in \{1, 2\}.$$

Furthermore, by Lemma A.1,

$$\begin{aligned}
&\left| \operatorname{Cov}\left[\frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{W}_i = \boldsymbol{w}_1), \frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{W}_i = \boldsymbol{w}_2)\right] \right| \\
&= \frac{1}{n^2}\left| \sum_{i,j=1}^{n} \operatorname{Cov}(I(\boldsymbol{W}_i = \boldsymbol{w}_1), I(\boldsymbol{W}_j = \boldsymbol{w}_2)) \right| \\
&\leq \frac{1}{n^2}\sum_{i,j=1}^{n} \left| \operatorname{Cov}(I(\boldsymbol{W}_i = \boldsymbol{w}_1), I(\boldsymbol{W}_j = \boldsymbol{w}_2)) \right| \\
&\leq \frac{1}{n^2}\sum_{i,j=1}^{n} \rho_{ij}\sqrt{\operatorname{Var}(I(\boldsymbol{W}_i = \boldsymbol{w}_1))\operatorname{Var}(I(\boldsymbol{W}_j = \boldsymbol{w}_2))} \\
&\leq \frac{1}{n^2}\sum_{i,j=1}^{n} \rho_{ij} = \frac{1}{n^2}\sum_{i=1}^{n} \rho_i.
\end{aligned}$$

Putting pieces together, we obtain that

$$\begin{aligned}
\mathbb{E}[\mathcal{D}] &\geq c_{\mathcal{D}}^2 \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{W}_i = \boldsymbol{w}_1)\right)\left(\frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{W}_i = \boldsymbol{w}_2)\right)\right] \\
&= c_{\mathcal{D}}^2 \Bigg\{ \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{W}_i = \boldsymbol{w}_1)\right]\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{W}_i = \boldsymbol{w}_2)\right] \\
&\qquad + \operatorname{Cov}\left[\frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{W}_i = \boldsymbol{w}_1), \frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{W}_i = \boldsymbol{w}_2)\right] \Bigg\} \\
&\geq c_{\mathcal{D}}^2 \left(c_\pi^2 - \frac{1}{n^2}\sum_{i=1}^{n} \rho_i\right).
\end{aligned}$$

On the other hand, by Lemma A.1, for $k \in \{1, 2\}$,

$$\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{W}_i = \boldsymbol{w}_k)\right) \leq \frac{1}{n^2}\sum_{i=1}^{n} \rho_i = O(n^{-q}) = o(1).$$

By Markov's inequality, for $k \in \{1, 2\}$,

$$\frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{W}_i = \boldsymbol{w}_k) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}(\boldsymbol{W}_i = \boldsymbol{w}_1) - o_{\mathbb{P}}(1) \geq c_\pi - o_{\mathbb{P}}(1).$$

Therefore,

$$\mathcal{D} \geq c_{\mathcal{D}}^2(c_\pi - o_{\mathbb{P}}(1))(c_\pi - o_{\mathbb{P}}(1)) \geq c_{\mathcal{D}}^2 c_\pi^2 - o_{\mathbb{P}}(1).$$

□

Based on Lemma A.2 and A.3, we can derive an asymptotic linear expansion for the RIPW estimator.

Theorem A.2. *Under Assumptions A.2 and A.3,*

$$\mathcal{D}(\hat{\tau} - \tau^*) = N_* + \frac{1}{n}\sum_{i=1}^{n}(\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i]) + O_{\mathbb{P}}\left(n^{-q}\right),$$

where

$$N_* = \frac{1}{2n}\sum_{i=1}^{n}\mathbb{E}[\mathcal{V}_i] = \mathbb{E}[\Gamma_{wy}]\mathbb{E}[\Gamma_\theta] - \mathbb{E}[\boldsymbol{\Gamma}_w]^\top \mathbb{E}[\boldsymbol{\Gamma}_y] - \tau^*\left(\mathbb{E}[\Gamma_{ww}]\mathbb{E}[\Gamma_\theta] - \mathbb{E}[\boldsymbol{\Gamma}_w]^\top \mathbb{E}[\boldsymbol{\Gamma}_w]\right).$$

Furthermore,

$$\hat{\tau} - \tau^* = O_{\mathbb{P}}(|N_*|) + O_{\mathbb{P}}(n^{-q/2}).$$

Proof. Note that

$$\mathcal{D}(\hat{\tau} - \tau^*) = \mathcal{N} - \tau^*\mathcal{D}.$$

By Lemma A.2,

$$\begin{aligned}
&\left|(\Gamma_{wy} - \mathbb{E}[\Gamma_{wy}])(\Gamma_\theta - \mathbb{E}[\Gamma_\theta])\right| + \left|(\boldsymbol{\Gamma}_w - \mathbb{E}[\boldsymbol{\Gamma}_w])^\top(\boldsymbol{\Gamma}_y - \mathbb{E}[\boldsymbol{\Gamma}_y])\right| \\
\leq &\frac{1}{2}\left\{(\Gamma_{wy} - \mathbb{E}[\Gamma_{wy}])^2 + (\Gamma_\theta - \mathbb{E}[\Gamma_\theta])^2 + \|\boldsymbol{\Gamma}_w - \mathbb{E}[\boldsymbol{\Gamma}_w]\|_2^2 + \|\boldsymbol{\Gamma}_y - \mathbb{E}[\boldsymbol{\Gamma}_y]\|_2^2\right\}
\end{aligned}$$

$$=O_{\mathbb{P}}\left(\mathrm{Var}(\Gamma_{wy}) + \mathrm{Var}(\Gamma_\theta) + \mathrm{Var}(\boldsymbol{\Gamma}_w) + \mathrm{Var}(\boldsymbol{\Gamma}_y)\right) = O_{\mathbb{P}}(n^{-q}).$$

Let

$$\mathcal{V}_{i1} = \Theta_i \left\{ \mathbb{E}[\Gamma_{wy}] - \mathbb{E}[\boldsymbol{\Gamma}_y]^\top J \boldsymbol{W}_i + \mathbb{E}[\Gamma_\theta] \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i - \mathbb{E}[\boldsymbol{\Gamma}_w]^\top J \tilde{\boldsymbol{Y}}_i \right\}.$$

Then,

$$\mathcal{N} = \mathbb{E}[\Gamma_{wy}]\mathbb{E}[\Gamma_\theta] - \mathbb{E}[\boldsymbol{\Gamma}_w]^\top \mathbb{E}[\boldsymbol{\Gamma}_y] + \frac{1}{n}\sum_{i=1}^n (\mathcal{V}_{i1} - \mathbb{E}[\mathcal{V}_{i1}]) + O_{\mathbb{P}}(n^{-q}),$$

Similarly,

$$\mathcal{D} = \mathbb{E}[\Gamma_{ww}]\mathbb{E}[\Gamma_\theta] - \mathbb{E}[\boldsymbol{\Gamma}_w]^\top \mathbb{E}[\boldsymbol{\Gamma}_w] + \frac{1}{n}\sum_{i=1}^n (\mathcal{V}_{i2} - \mathbb{E}[\mathcal{V}_{i2}]) + O_{\mathbb{P}}(n^{-q}),$$

where

$$\mathcal{V}_{i2} = \Theta_i \left\{ \mathbb{E}[\Gamma_{ww}] - \mathbb{E}[\boldsymbol{\Gamma}_w]^\top J \boldsymbol{W}_i + \mathbb{E}[\Gamma_\theta] \boldsymbol{W}_i^\top J \boldsymbol{W}_i - \mathbb{E}[\boldsymbol{\Gamma}_w]^\top J \boldsymbol{W}_i \right\}.$$

Since $\mathcal{V}_i = \mathcal{V}_{i1} - \tau^* \mathcal{V}_{i2}$,

$$\mathcal{D}(\hat{\tau} - \tau^*) = \mathcal{N} - \tau^* \mathcal{D} = N_* + \frac{1}{n}\sum_{i=1}^n (\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i]) + O_{\mathbb{P}}(n^{-q}).$$

This proves the first statement.

Next, we prove the second statement on $\hat{\tau} - \tau^*$. By Lemma A.3, $1/\mathcal{D} = O_{\mathbb{P}}(1)$. It is left to show that

$$\frac{1}{n}\sum_{i=1}^n (\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i]) = O_{\mathbb{P}}(n^{-q/2}).$$

Applying the inequality that $\mathrm{Var}(Z_1 + Z_2) = 2\mathrm{Var}(Z_1) + 2\mathrm{Var}(Z_2) - \mathrm{Var}(Z_1 - Z_2) \le 2(\mathrm{Var}(Z_1) +$

$\mathrm{Var}(Z_2)$), we obtain that

$$
\begin{aligned}
&\frac{1}{4}\mathrm{Var}(\mathcal{V}_{i1})\\
\leq&\mathrm{Var}\left(\Theta_i\mathbb{E}[\Gamma_{wy}]\right)+\mathrm{Var}\left(\Theta_i\mathbb{E}[\Gamma_\theta]\boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i\right)+\mathrm{Var}\left(\Theta_i\mathbb{E}[\boldsymbol{\Gamma}_w]^\top J\tilde{\boldsymbol{Y}}_i\right)+\mathrm{Var}\left(\Theta_i\mathbb{E}[\boldsymbol{\Gamma}_y]^\top J\boldsymbol{W}_i\right)\\
\leq&\mathbb{E}\left(\Theta_i\mathbb{E}[\Gamma_{wy}]\right)^2+\mathbb{E}\left(\Theta_i\mathbb{E}[\Gamma_\theta]\boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i\right)^2+\mathbb{E}\left(\Theta_i\mathbb{E}[\boldsymbol{\Gamma}_w]^\top J\tilde{\boldsymbol{Y}}_i\right)^2+\mathbb{E}\left(\Theta_i\mathbb{E}[\boldsymbol{\Gamma}_y]^\top J\boldsymbol{W}_i\right)^2\\
\overset{(i)}{\leq}&\frac{1}{c_\pi^2}\left\{(\mathbb{E}[\Gamma_{wy}])^2+(\mathbb{E}[\Gamma_\theta])^2\mathbb{E}(\boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i)^2+\mathbb{E}\left(\mathbb{E}[\boldsymbol{\Gamma}_w]^\top J\tilde{\boldsymbol{Y}}_i\right)^2+\mathbb{E}\left(\mathbb{E}[\boldsymbol{\Gamma}_y]^\top J\boldsymbol{W}_i\right)^2\right\}\\
\overset{(ii)}{\leq}&\frac{1}{c_\pi^2}\left\{(\mathbb{E}[\Gamma_{wy}])^2+(\mathbb{E}[\Gamma_\theta])^2\mathbb{E}\|\boldsymbol{W}_i\|_2^2\mathbb{E}\|\tilde{\boldsymbol{Y}}_i\|_2^2+\|\mathbb{E}[\boldsymbol{\Gamma}_w]\|_2^2\mathbb{E}\|\tilde{\boldsymbol{Y}}_i\|_2^2+\|\mathbb{E}[\boldsymbol{\Gamma}_y]\|_2^2\mathbb{E}\|\boldsymbol{W}_i\|_2^2\right\}\\
\overset{(iii)}{\leq}&\frac{1}{c_\pi^2}\left\{(\mathbb{E}[\Gamma_{wy}])^2+T(\mathbb{E}[\Gamma_\theta])^2\mathbb{E}\|\tilde{\boldsymbol{Y}}_i\|_2^2+\|\mathbb{E}[\boldsymbol{\Gamma}_w]\|_2^2\mathbb{E}\|\tilde{\boldsymbol{Y}}_i\|_2^2+T\|\mathbb{E}[\boldsymbol{\Gamma}_y]\|_2^2\right\},
\end{aligned}
$$

where (i) follows from the Assumption A.2 that $\Theta_i \leq 1/c_\pi$ almost surely, (ii) follows from the Cauchy-Schwarz inequality and the fact that $\|J\|_{\mathrm{op}} = 1$, and (iii) follows from the fact that $\|\boldsymbol{W}_i\|_2^2 \leq T$. By Lemma A.2, we obtain that for all $i \in [n]$,

$$
\mathrm{Var}(\mathcal{V}_{i1}) \leq C_1\left(1+\mathbb{E}\|\tilde{\boldsymbol{Y}}_i\|_2^2\right) \leq C_1\left(1+\mathbb{E}\|\tilde{\boldsymbol{Y}}_i(0)\|_2^2+\mathbb{E}\|\tilde{\boldsymbol{Y}}_i(1)\|_2^2\right), \tag{A.10}
$$

for some constant C_1 that only depends on c_π and T. Similarly, we have that $\mathrm{Var}(\mathcal{V}_{i2}) \leq C_2$ for some constant C_2 that only depends on c_π and T. By Assumption A.3,

$$
\tau^* = \sum_{t=1}^{T}\xi_t\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{E}[\tilde{Y}_{it}(1)]-\mathbb{E}[\tilde{Y}_{it}(0)]\right)\right\} = O(1).
$$

Therefore,

$$
\mathrm{Var}(\mathcal{V}_i) \leq 2\mathrm{Var}(\mathcal{V}_{i1})+2(\tau^*)^2\mathrm{Var}(\mathcal{V}_{i2}) \leq C\left(1+\mathbb{E}\|\tilde{\boldsymbol{Y}}_i(0)\|_2^2+\mathbb{E}\|\tilde{\boldsymbol{Y}}_i(1)\|_2^2\right).
$$

for some constant C that only depends on c_π and T. Since $\mathcal{V}_i$ is a function of $(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{X}_i)$, by Lemma A.1 and Assumption A.3,

$$
\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{V}_i\right) \leq \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(\mathcal{V}_i)\rho_i = O(n^{-q}).
$$

By Chebyshev's inequality,

$$\frac{1}{n}\sum_{i=1}^{n}(\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i]) = O_{\mathbb{P}}(n^{-q/2}).$$

The proof is then completed. $\square$

A.3.2 DATE equation and consistency

Theorem A.2 shows that the asymptotic limit of $\mathcal{D}(\hat{\tau} - \tau^*)$ is N_*. For consistency, it remains to prove that $N_* = o(1)$. We start by proving that the asymptotic bias is zero when either the treatment or the outcome model is perfectly estimated.

Lemma A.4. *Under Assumptions A.1, A.2, and A.3, $N_* = 0$, if either (1) $\Delta_{yi} = 0$ for all $i \in [n]$, or (2) $\Delta_{\pi i} = 0$ for all $i \in [n]$, and $\mathbf{\Pi}$ satisfies the DATE equation (2.14).*

Proof. Without loss of generality, we assume that $\tau^* = 0$; otherwise, we replace $Y_{it}(1)$ by $Y_{it}(1) - \tau^*$ and the resulting $\hat{\tau}$ becomes $\hat{\tau} - \tau^*$. Then

$$N_* = \mathbb{E}[\Gamma_{wy}]\mathbb{E}[\Gamma_\theta] - \mathbb{E}[\mathbf{\Gamma}_w]^\top \mathbb{E}[\mathbf{\Gamma}_y].$$

It remains to prove that $N_* = 0$. Since $(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i)$ are deterministic, by Assumption A.1 and (A.1),

$$\begin{aligned}\mathbb{E}[\Gamma_{wy}] &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\Theta_i \boldsymbol{W}_i^\top J \tilde{\mathbf{Y}}_i] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\Theta_i \boldsymbol{W}_i^\top J\{\tilde{\mathbf{Y}}_i(0) + \mathrm{diag}(\boldsymbol{W}_i)(\tilde{\mathbf{Y}}_i(1) - \tilde{\mathbf{Y}}_i(0))\}] \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\Theta_i J \boldsymbol{W}_i]^\top \mathbb{E}[\tilde{\mathbf{Y}}_i(0)] + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\Theta_i \boldsymbol{W}_i^\top J\,\mathrm{diag}(\boldsymbol{W}_i)]\tilde{\boldsymbol{\tau}}_i.\end{aligned}$$

Similarly,

$$\mathbb{E}[\mathbf{\Gamma}_y] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\Theta_i] J \mathbb{E}[\tilde{\mathbf{Y}}_i(0)] + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\Theta_i\,\mathrm{diag}(\boldsymbol{W}_i)]\tilde{\boldsymbol{\tau}}_i.$$

As a result,

$$N_* = \frac{1}{n}\sum_{i=1}^{n}\{\mathbb{E}[\Theta_i J\boldsymbol{W}_i]\mathbb{E}[\Gamma_\theta] - \mathbb{E}[\Theta_i]\mathbb{E}[\boldsymbol{\Gamma}_w]\}^\top \mathbb{E}[\tilde{\boldsymbol{Y}}_i(0)]$$
$$+ \frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{E}[\Theta_i \boldsymbol{W}_i^\top J \operatorname{diag}(\boldsymbol{W}_i)]\mathbb{E}[\Gamma_\theta] - \mathbb{E}[\boldsymbol{\Gamma}_w]^\top \mathbb{E}[\Theta_i \operatorname{diag}(\boldsymbol{W}_i)]\right\} \tilde{\boldsymbol{\tau}}_i. \tag{A.11}$$

If $\Delta_{yi} = 0$, $\hat{\boldsymbol{m}}_i = \boldsymbol{m}_i$ and $\hat{\boldsymbol{\nu}}_i = \boldsymbol{\nu}_i$. Since we have assumed $\tau^* = 0$, $\tilde{\boldsymbol{\tau}}_i = \boldsymbol{0}_T$. By (A.3), $\mathbb{E}[\tilde{\boldsymbol{Y}}_i(0)] = \boldsymbol{0}_T$. It is then obvious from (A.11) that $N_* = 0$.

If $\Delta_{\pi i} = 0$, $\hat{\boldsymbol{\pi}}_i = \boldsymbol{\pi}_i$ and thus for any function $f(\cdot)$,

$$\mathbb{E}[\Theta_i f(\boldsymbol{W}_i)] = \sum_{\boldsymbol{w}\in\{0,1\}^T} \frac{\boldsymbol{\Pi}(\boldsymbol{w})}{\boldsymbol{\pi}_i(\boldsymbol{w})} f(\boldsymbol{w})\boldsymbol{\pi}_i(\boldsymbol{w}) = \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[f(\boldsymbol{W})]. \tag{A.12}$$

As a result,

$$\mathbb{E}[\Theta_i J\boldsymbol{W}_i] = \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[J\boldsymbol{W}] = \mathbb{E}[\boldsymbol{\Gamma}_w], \quad \mathbb{E}[\Theta_i] = 1 = \mathbb{E}[\Gamma_\theta],$$

and

$$\mathbb{E}[\Theta_i \boldsymbol{W}_i^\top J \operatorname{diag}(\boldsymbol{W}_i)] = \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W} J \operatorname{diag}(\boldsymbol{W})], \quad \mathbb{E}[\Theta_i \operatorname{diag}(\boldsymbol{W}_i)] = \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\operatorname{diag}(\boldsymbol{W})].$$

Then

$$\mathbb{E}[\Theta_i J\boldsymbol{W}_i]\mathbb{E}[\Gamma_\theta] - \mathbb{E}[\Theta_i]\mathbb{E}[\boldsymbol{\Gamma}_w] = \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[J\boldsymbol{W}] - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[J\boldsymbol{W}] = 0,$$

and by DATE equation,

$$\begin{aligned}
&\mathbb{E}[\Theta_i \boldsymbol{W}_i^\top J \operatorname{diag}(\boldsymbol{W}_i)]\mathbb{E}[\Gamma_\theta] - \mathbb{E}[\boldsymbol{\Gamma}_w]^\top \mathbb{E}[\Theta_i \operatorname{diag}(\boldsymbol{W}_i)] \\
&= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])^\top J \operatorname{diag}(\boldsymbol{W})] \\
&= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])^\top J\boldsymbol{W}]\xi^\top.
\end{aligned}$$

By (A.11) and (A.2),

$$
\begin{aligned}
N_* &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[(\boldsymbol{W}-\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])^\top J\boldsymbol{W}]\xi^\top\tilde{\boldsymbol{\tau}}_i \\
&= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[(\boldsymbol{W}-\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])^\top J\boldsymbol{W}]\left(\frac{1}{n}\sum_{i=1}^{n}\xi^\top\tilde{\boldsymbol{\tau}}_i\right) \\
&= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[(\boldsymbol{W}-\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])^\top J\boldsymbol{W}]\tau^* = 0.
\end{aligned}
$$

□

Next, we prove a general bound for the asymptotic bias N_* as a function of $(\Delta_{yi}, \Delta_{\pi i})_{i=1}^n$.

Theorem A.3. *Let $\boldsymbol{\Pi}$ be an solution of the DATE equation (2.14). Under Assumptions A.1, A.2, and A.3,*

$$
|N_*| = O\left(\bar{\Delta}_\pi\bar{\Delta}_y\right).
$$

Proof. As in the proof of Lemma A.4, we assume that $\tau^* = 0$. Let

$$
\Theta_i^* = \frac{\boldsymbol{\Pi}(\boldsymbol{W}_i)}{\boldsymbol{\pi}_i(\boldsymbol{W}_i)}, \quad \tilde{\boldsymbol{Y}}_i^* = \boldsymbol{Y}_i - \boldsymbol{m}_i - \mathrm{diag}(\boldsymbol{W}_i)\boldsymbol{\nu}_i.
$$

Further, let Γ_θ^* and $\boldsymbol{\Gamma}_w^*$ be the counterpart of Γ_θ and $\boldsymbol{\Gamma}_w$ with $(\Theta_i, \tilde{\boldsymbol{Y}}_i)$ replaced by $(\Theta_i^*, \tilde{\boldsymbol{Y}}_i^*)$. For any function $f : \{0,1\}^T \mapsto \mathbb{R}$ such that $\mathbb{E}[f^2(\boldsymbol{W}_i)] \le C_1$ for some constant $C_1 > 0$, by Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\mathbb{E}[\Theta_i f(\boldsymbol{W}_i) - \Theta_i^* f(\boldsymbol{W}_i)] = \mathbb{E}[(\Theta_i - \Theta_i^*)f(\boldsymbol{W}_i)] \le \sqrt{C_1}\sqrt{\mathbb{E}\left(\Theta_i - \Theta_i^*\right)^2} \\
&= \sqrt{C_1}\sqrt{\mathbb{E}\left[\frac{\boldsymbol{\Pi}(\boldsymbol{W}_i)^2}{\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i)^2\boldsymbol{\pi}_i(\boldsymbol{W}_i)^2}\left(\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) - \boldsymbol{\pi}_i(\boldsymbol{W}_i)\right)^2\right]} \le \frac{\sqrt{C_1}}{c_\pi^2}\Delta_{\pi i}. \qquad \text{(A.13)}
\end{aligned}
$$

Thus, there exists a constant C_2 that only depends on c_π and T such that

$$
\begin{aligned}
&|\mathbb{E}[\Theta_i] - \mathbb{E}[\Theta_i^*]| + \|\mathbb{E}[\Theta_i J\boldsymbol{W}_i] - \mathbb{E}[\Theta_i^* J\boldsymbol{W}_i]\|_2 + \|\mathbb{E}[\Theta_i \boldsymbol{W}_i^\top J\,\mathrm{diag}(\boldsymbol{W}_i)] - \mathbb{E}[\Theta_i^* \boldsymbol{W}_i^\top J\,\mathrm{diag}(\boldsymbol{W}_i)]\|_2 \\
&\quad + \|\mathbb{E}[\Theta_i\,\mathrm{diag}(\boldsymbol{W}_i)] - \mathbb{E}[\Theta_i^*\,\mathrm{diag}(\boldsymbol{W}_i)]\|_{\mathrm{op}} \le C_2\Delta_{\pi i}.
\end{aligned}
$$

By triangle inequality and Cauchy-Schwarz inequality, we also have

$$|\mathbb{E}[\Gamma_\theta] - \mathbb{E}[\Gamma^*_\theta]| + \|\mathbb{E}[\boldsymbol{\Gamma}_w] - \mathbb{E}[\boldsymbol{\Gamma}^*_w]\|_2 \le \frac{C_2}{n}\sum_{i=1}^n \Delta_{\pi i} \le C_2 \bar{\Delta}_\pi.$$

On the other hand, by Lemma A.2, there exists a constant C_3 that only depends on c_π and T,

$$|\mathbb{E}[\Gamma_\theta]| + \|\mathbb{E}[\boldsymbol{\Gamma}_w]\|_2 \le C_3.$$

Without loss of generality, we assume that

$$C_3 \ge 1 + \sqrt{T} \ge 1 + \|\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[J\boldsymbol{W}]\|_2 = \mathbb{E}[\Theta^*_i] + \|\mathbb{E}[\Theta^*_i J\boldsymbol{W}_i]\|_2.$$

Putting pieces together,

$$\begin{aligned}
&\Big|\mathbb{E}[\Theta_i J\boldsymbol{W}_i]\mathbb{E}[\Gamma_\theta] - \mathbb{E}[\Theta_i]\mathbb{E}[\boldsymbol{\Gamma}_w] - (\mathbb{E}[\Theta^*_i J\boldsymbol{W}_i]\mathbb{E}[\Gamma^*_\theta] - \mathbb{E}[\Theta^*_i]\mathbb{E}[\boldsymbol{\Gamma}^*_w])\Big| \\
&\le \Big|\mathbb{E}[\Theta_i J\boldsymbol{W}_i] - \mathbb{E}[\Theta^*_i J\boldsymbol{W}_i]\Big| \cdot \mathbb{E}[\Gamma_\theta] + \Big|\mathbb{E}[\Theta_i] - \mathbb{E}[\Theta^*_i]\Big| \cdot \|\mathbb{E}[\boldsymbol{\Gamma}_w]\|_2 \\
&\quad + \Big|\mathbb{E}[\Gamma_\theta] - \mathbb{E}[\Gamma^*_\theta]\Big| \cdot \|\mathbb{E}[\Theta^*_i J\boldsymbol{W}_i]\|_2 + \Big\|\mathbb{E}[\boldsymbol{\Gamma}_w] - \mathbb{E}[\boldsymbol{\Gamma}^*_w]\Big\| \cdot \mathbb{E}[\Theta^*_i] \\
&\le 2C_3C_2(\Delta_{\pi i} + \bar{\Delta}_\pi).
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\Big|\mathbb{E}[\Theta_i \boldsymbol{W}_i^\top J \operatorname{diag}(\boldsymbol{W}_i)]\mathbb{E}[\Gamma_\theta] - \mathbb{E}[\boldsymbol{\Gamma}_w]^\top \mathbb{E}[\Theta_i \operatorname{diag}(\boldsymbol{W}_i)] \\
&- \big(\mathbb{E}[\Theta^*_i \boldsymbol{W}_i^\top J \operatorname{diag}(\boldsymbol{W}_i)]\mathbb{E}[\Gamma^*_\theta] - \mathbb{E}[\boldsymbol{\Gamma}^*_w]^\top \mathbb{E}[\Theta^*_i \operatorname{diag}(\boldsymbol{W}_i)]\big)\Big| \\
&\le 2C_3C_2(\Delta_{\pi i} + \bar{\Delta}_\pi).
\end{aligned}$$

Let

$$\begin{aligned}
N'_* = \frac{1}{n}\sum_{i=1}^n &\{\mathbb{E}[\Theta^*_i J\boldsymbol{W}_i]\mathbb{E}[\Gamma^*_\theta] - \mathbb{E}[\Theta^*_i]\mathbb{E}[\boldsymbol{\Gamma}^*_w]\}^\top \mathbb{E}[\tilde{\boldsymbol{Y}}_i(0)] \\
&+ \frac{1}{n}\sum_{i=1}^n \left\{\mathbb{E}[\Theta^*_i \boldsymbol{W}_i^\top J \operatorname{diag}(\boldsymbol{W}_i)]\mathbb{E}[\Gamma^*_\theta] - \mathbb{E}[\boldsymbol{\Gamma}^*_w]^\top \mathbb{E}[\Theta^*_i \operatorname{diag}(\boldsymbol{W}_i)]\right\} \tilde{\boldsymbol{\tau}}_i.
\end{aligned}$$

Using the same arguments as in the proof of Lemma A.4,

$$\mathbb{E}[\Theta_i^* J \boldsymbol{W}_i]\mathbb{E}[\Gamma_\theta^*] - \mathbb{E}[\Theta_i^*]\mathbb{E}[\boldsymbol{\Gamma}_w^*] = 0,$$

and

$$\mathbb{E}[\Theta_i^* \boldsymbol{W}_i^\top J \operatorname{diag}(\boldsymbol{W}_i)]\mathbb{E}[\Gamma_\theta^*] - \mathbb{E}[\boldsymbol{\Gamma}_w^*]^\top \mathbb{E}[\Theta_i^* \operatorname{diag}(\boldsymbol{W}_i)] = \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])^\top J \boldsymbol{W}]\xi^\top.$$

Then

$$N_*' = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])^\top J \boldsymbol{W}]\xi^\top \tilde{\boldsymbol{\tau}}_i = \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])^\top J \boldsymbol{W}]\tau^* = 0.$$

This entails that

$$|N_*| = |N_* - N_*'| \le \frac{2C_3C_2}{n}\sum_{i=1}^n (\Delta_{\pi i} + \bar{\Delta}_\pi)(\|\mathbb{E}[\tilde{\boldsymbol{Y}}_i(0)]\|_2 + \|\tilde{\boldsymbol{\tau}}_i\|_2).$$

By (A.1), (A.2), and (A.3),

$$\|\mathbb{E}[\tilde{\boldsymbol{Y}}_i(0)]\|_2 + \|\tilde{\boldsymbol{\tau}}_i\|_2 = \|\hat{\boldsymbol{m}}_i - \boldsymbol{m}_i\|_2 + \|\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i\|_2 \le 2\Delta_{yi}.$$

Since $(1/n)\sum_{i=1}^n \Delta_{yi} \le \sqrt{(1/n)\sum_{i=1}^n \Delta_{yi}^2}$,

$$|N_*| \le \frac{4C_3C_2}{n}\sum_{i=1}^n (\Delta_{\pi i} + \bar{\Delta}_\pi)\Delta_{yi} = 4C_3C_2\bar{\Delta}_\pi\bar{\Delta}_y.$$

The proof is then completed. □

A.3.3 Asymptotic inference under independence

Theorem A.2 and Theorem A.3 imply the following properties of RIPW estimators.

Theorem A.4. *Let* $\boldsymbol{\Pi}$ *be an solution of the DATE equation* (2.14). *Under Assumptions A.1,* *A.2, and A.3,*

$$\hat{\tau} - \tau^* = o_{\mathbb{P}}(1), \quad \textit{if} \;\; \bar{\Delta}_\pi \bar{\Delta}_y = o(1).$$

If, further, $q > 1/2$ in Assumption A.3 and $\bar{\Delta}_\pi \bar{\Delta}_y = o(1/\sqrt{n})$,

$$\mathcal{D} \cdot \sqrt{n}(\hat{\tau} - \tau^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i]) + o_{\mathbb{P}}(1).$$

Recalling (A.8) that $(\Delta_{\pi i}, \Delta_{yi})$ are deterministic, $\bar{\Delta}_\pi \bar{\Delta}_y = \mathbb{E}[\bar{\Delta}_\pi \bar{\Delta}_y]$. Since Assumptions A.2 and A.3 generalize Assumptions 2.1-2.3, Theorem A.4 implies Theorem 2.1 and 2.2. Similarly, Theorem 3.1 is implied by Theorem A.4.

Throughout the rest of the subsection, we focus on the special case where $\{(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{X}_i) : i \in [n]\}$ are independent. In this case, Assumption A.3 holds with $q = 1 > 1/2$ and thus the asymptotically linear expansion in Theorem A.4 holds. To obtain the asymptotic normality and a consistent variance estimator, we modify Assumption A.3 as follows.

Assumption A.4. *$\{(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{X}_i) : i = 1, \ldots, n\}$ are independent (but not necessarily identically distributed), and there exists $\omega > 0$ such that*

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ \mathbb{E}\|\tilde{\boldsymbol{Y}}_i(1)\|_2^{2+\omega} + \mathbb{E}\|\tilde{\boldsymbol{Y}}_i(0)\|_2^{2+\omega} \right\} = O(1).$$

To derive the asymptotic normality of the RIPW estimator, we need the following assumption that prevents the variance from being too small.

Assumption A.5. *There exists $v_0 > 0$ such that*

$$\sigma^2 \triangleq \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}(\mathcal{V}_i) \geq v_0.$$

The following lemma shows the asymptotic normality of the term $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i])$.

Lemma A.5. *Then under Assumptions A.2, A.4, and A.5,*

$$d_K \left(\mathcal{L} \left(\frac{1}{\sqrt{n}\sigma} \sum_{i=1}^{n} (\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i]) \right), N(0,1) \right) \to 0,$$

where $\mathcal{L}(\cdot)$ denotes the probability law, d_K denotes the Kolmogorov-Smirnov distance (i.e., the ℓ_∞-norm of the difference of CDFs)

Proof. Since $(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i)$ are deterministic, by Assumption A.4, $\{\mathcal{V}_i : i \in [n]\}$ are independent. Recalling the definition of $\mathcal{V}_i$, it is easy to see that Assumption A.4 implies

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}|\mathcal{V}_i|^{2+\omega} = O(1). \tag{A.14}$$

By Assumption A.4,

$$\sum_{i=1}^{n}\mathbb{E}\left|\frac{\mathcal{V}_i}{\sqrt{n}\sigma}\right|^{2+\omega} = O\left(n^{-\omega/2}\right) = o(1).$$

The proof is completed by the Berry-Esseen inequality (Proposition A.1) with $g(x) = x^\omega$. $\square$

Let $\hat{\mathcal{V}}_i$ denote the plug-in estimate of $\mathcal{V}_i$, i.e.,

$$\hat{\mathcal{V}}_i = \Theta_i\Bigg\{(\Gamma_{wy} - \hat{\tau}\Gamma_{ww}) - (\boldsymbol{\Gamma}_y - \hat{\tau}\boldsymbol{\Gamma}_w)^\top J\boldsymbol{W}_i + \Gamma_\theta \boldsymbol{W}_i^\top J\left(\tilde{\boldsymbol{Y}}_i - \hat{\tau}\boldsymbol{W}_i\right) - \boldsymbol{\Gamma}_w^\top J\left(\tilde{\boldsymbol{Y}}_i - \hat{\tau}\boldsymbol{W}_i\right)\Bigg\}. \tag{A.15}$$

We first prove that $\hat{\mathcal{V}}_i$ is an accurate approximation of $\mathcal{V}_i$ on average, even without the independence assumption.

Lemma A.6. *Let $\boldsymbol{\Pi}$ be a solution of the DATE equation. Under Assumptions A.1-A.3,*

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{\mathcal{V}}_i - \mathcal{V}_i)^2 = o_{\mathbb{P}}(1), \quad \text{if } \bar{\Delta}_\pi\bar{\Delta}_y = o(1).$$

Proof. Let

$$\hat{\mathcal{V}}_i' = \Theta_i\Bigg\{(\Gamma_{wy} - \tau^*\Gamma_{ww}) - (\boldsymbol{\Gamma}_y - \tau^*\boldsymbol{\Gamma}_w)^\top J\boldsymbol{W}_i + \Gamma_\theta \boldsymbol{W}_i^\top J\left(\tilde{\boldsymbol{Y}}_i - \tau^*\boldsymbol{W}_i\right) - \boldsymbol{\Gamma}_w^\top J\left(\tilde{\boldsymbol{Y}}_i - \tau^*\boldsymbol{W}_i\right)\Bigg\}.$$

Then

$$\hat{\mathcal{V}}_i - \hat{\mathcal{V}}_i' = (\hat{\tau} - \tau^*)\Theta_i\big\{-\Gamma_{ww} + \boldsymbol{\Gamma}_w^\top J\boldsymbol{W}_i - \Gamma_\theta \boldsymbol{W}_i^\top J\boldsymbol{W}_i + \boldsymbol{\Gamma}_w^\top J\boldsymbol{W}_i\big\}.$$

Under Assumption A.2, there exists a constant C that only depends on c_π and T such that

$$|\hat{\mathcal{V}}_i - \hat{\mathcal{V}}_i'| \le C|\hat{\tau} - \tau^*|.$$

By Theorem A.4,

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{\mathcal{V}}_i - \hat{\mathcal{V}}_i')^2 = O((\hat{\tau} - \tau^*)^2) = o_{\mathbb{P}}(1) \tag{A.16}$$

Next,

$$\begin{aligned}\hat{\mathcal{V}}_i' - \mathcal{V}_i = \Theta_i\Big\{ &((\Gamma_{wy} - \mathbb{E}[\Gamma_{wy}]) - \tau^*(\Gamma_{ww} - \mathbb{E}[\Gamma_{ww}])) - ((\boldsymbol{\Gamma}_y - \mathbb{E}[\boldsymbol{\Gamma}_y]) - \tau^*(\boldsymbol{\Gamma}_w - \mathbb{E}[\boldsymbol{\Gamma}_w]))^\top J\boldsymbol{W}_i \\ &+ (\Gamma_\theta - \mathbb{E}[\Gamma_\theta])\boldsymbol{W}_i^\top J\left(\tilde{\boldsymbol{Y}}_i - \tau^*\boldsymbol{W}_i\right) - (\boldsymbol{\Gamma}_w - \mathbb{E}[\boldsymbol{\Gamma}_w])^\top J\left(\tilde{\boldsymbol{Y}}_i - \tau^*\boldsymbol{W}_i\right)\Big\}.\end{aligned}$$

By Jensen's inequality and Assumption A.2,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}(\hat{\mathcal{V}}_i' - \mathcal{V}_i)^2 \\ &\le \frac{5}{nc_\pi^2}\sum_{i=1}^{n}\Big\{(\Gamma_{wy} - \mathbb{E}[\Gamma_{wy}])^2 + (\Gamma_{ww} - \mathbb{E}[\Gamma_{ww}])^2\cdot\tau^{*2} + \|(\boldsymbol{\Gamma}_y - \mathbb{E}[\boldsymbol{\Gamma}_y])\|_2^2\cdot\|J\boldsymbol{W}_i\|_2^2 \\ &\qquad + \|(\boldsymbol{\Gamma}_w - \mathbb{E}[\boldsymbol{\Gamma}_w])\|_2^2\cdot\|J(\tilde{\boldsymbol{Y}}_i - 2\tau^*\boldsymbol{W}_i)\|_2^2 + (\Gamma_\theta - \mathbb{E}[\Gamma_\theta])^2\left(\boldsymbol{W}_i^\top J\left(\tilde{\boldsymbol{Y}}_i - \tau^*\boldsymbol{W}_i\right)\right)^2\Big\} \\ &= \frac{5}{c_\pi^2}\Big\{(\Gamma_{wy} - \mathbb{E}[\Gamma_{wy}])^2 + (\Gamma_{ww} - \mathbb{E}[\Gamma_{ww}])^2\cdot\tau^{*2} + \|(\boldsymbol{\Gamma}_y - \mathbb{E}[\boldsymbol{\Gamma}_y])\|_2^2\cdot T \\ &\qquad + \|(\boldsymbol{\Gamma}_w - \mathbb{E}[\boldsymbol{\Gamma}_w])\|_2^2\cdot\frac{1}{n}\sum_{i=1}^{n}\|(\tilde{\boldsymbol{Y}}_i - 2\tau^*\boldsymbol{W}_i)\|_2^2 \\ &\qquad + \|(\Gamma_\theta - \mathbb{E}[\Gamma_\theta])\|_2^2\cdot\frac{T}{n}\sum_{i=1}^{n}\|\tilde{\boldsymbol{Y}}_i - \tau^*\boldsymbol{W}_i\|_2^2\Big\}.\end{aligned}$$

By Lemma A.2,

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(\hat{\mathcal{V}}_i' - \mathcal{V}_i)^2\right] = o(1).$$

By Markov's inequality,

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{\mathcal{V}}_i' - \mathcal{V}_i)^2 = o_{\mathbb{P}}(1). \tag{A.17}$$

Putting (A.16) and (A.17) together, we obtain that

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{\mathcal{V}}_i - \mathcal{V}_i)^2 \leq \frac{2}{n}\sum_{i=1}^{n}\{(\hat{\mathcal{V}}_i - \hat{\mathcal{V}}_i')^2 + (\hat{\mathcal{V}}_i' - \mathcal{V}_i)^2\} = o_{\mathbb{P}}(1).$$

□

As in Section 2, we estimate the (conservative) variance of the term $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i])$ as

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\hat{\mathcal{V}}_i - \frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i\right)^2 = \frac{n}{n-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i\right)^2\right\}. \tag{A.18}$$

This yields a Wald-type confidence interval for DATE,

$$\hat{C}_{1-\alpha} = [\hat{\tau} - z_{1-\alpha/2}\hat{\sigma}/\sqrt{n}\mathcal{D}, \hat{\tau} + z_{1-\alpha/2}\hat{\sigma}/\sqrt{n}\mathcal{D}], \tag{A.19}$$

where z_η is the η-th quantile of the standard normal distribution.

Theorem A.5. *Assume that $\bar{\Delta}_\pi\bar{\Delta}_y = o(1/\sqrt{n})$. Under Assumptions A.1, A.2, A.4, and A.5,*

$$\liminf_{n\to\infty}\mathbb{P}\left(\tau^* \in \hat{C}_{1-\alpha}\right) \geq 1-\alpha.$$

Proof. By Theorem A.2, Theorem A.3, Lemma A.5, and Assumption A.5,

$$\frac{\mathcal{D}\cdot\sqrt{n}(\hat{\tau}-\tau^*)}{\sigma} = \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}(\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i]) + o_{\mathbb{P}}(1) \xrightarrow{d} N(0,1) \text{ in Kolmogorov-Smirnov distance,}$$

As a result,

$$\left|\mathbb{P}\left(\left|\frac{\mathcal{D}\cdot\sqrt{n}(\hat{\tau}-\tau^*)}{\sigma}\right| \leq z_{1-\alpha/2}\cdot\frac{\hat{\sigma}}{\sigma}\right) - \left\{2\Phi\left(z_{1-\alpha/2}\cdot\frac{\hat{\sigma}}{\sigma}\right) - 1\right\}\right| = o(1), \tag{A.20}$$

where Φ is the cumulative distribution function of the standard normal distribution. Let

$$\sigma_+^2 = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(\mathcal{V}_i - \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\mathcal{V}_i]\right)^2 = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\mathcal{V}_i^2] - \left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}[\mathcal{V}_i]\right)^2. \tag{A.21}$$

Clearly, σ_+^2 is deterministic and

$$\sigma_+^2 = \sigma^2 + \frac{1}{n}\sum_{i=1}^n \left(\mathbb{E}[\mathcal{V}_i] - \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\mathcal{V}_i]\right)^2 \geq \sigma^2.$$

It remains to show that

$$\left|\frac{n-1}{n}\hat{\sigma}^2 - \sigma_+^2\right| = o_{\mathbb{P}}(1). \tag{A.22}$$

In fact, by Assumption A.5, (A.22) implies that

$$\sqrt{\frac{n-1}{n}}\frac{\hat{\sigma}}{\sigma} \xrightarrow{p} \frac{\sigma_+}{\sigma} \geq 1 \Longrightarrow \frac{\hat{\sigma}}{\sigma} \xrightarrow{p} \frac{\sigma_+}{\sigma} \geq 1.$$

By continuous mapping theorem,

$$2\Phi\left(z_{1-\alpha/2} \cdot \frac{\hat{\sigma}}{\sigma}\right) - 1 \xrightarrow{p} 2\Phi\left(z_{1-\alpha/2} \cdot \frac{\sigma_+}{\sigma}\right) - 1 \geq 1 - \alpha,$$

which completes the proof.

Now we prove (A.22). By Proposition A.2 and Jensen's inequality,

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^n (\mathcal{V}_i^2 - \mathbb{E}[\mathcal{V}_i^2])\right|^{1+\omega/2} \leq \frac{2}{n^{1+\omega/2}}\sum_{i=1}^n \mathbb{E}|\mathcal{V}_i^2 - \mathbb{E}[\mathcal{V}_i^2]|^{1+\omega/2} \\
&\leq \frac{2^{1+\omega/2}}{n^{1+\omega/2}}\sum_{i=1}^n \left(\mathbb{E}[|\mathcal{V}_i|^{2+\omega}] + \mathbb{E}[\mathcal{V}_i^2]^{1+\omega/2}\right) \leq \frac{2^{2+\omega/2}}{n^{1+\omega/2}}\sum_{i=1}^n \mathbb{E}\left[|\mathcal{V}_i|^{2+\omega}\right].
\end{aligned}$$

By (A.14),

$$\mathbb{E}\left|\frac{1}{n}\sum_{i=1}^n (\mathcal{V}_i^2 - \mathbb{E}[\mathcal{V}_i^2])\right|^{1+\omega/2} = o(1).$$

By Markov's inequality,

$$\frac{1}{n}\sum_{i=1}^{n}(\mathcal{V}_i^2 - \mathbb{E}[\mathcal{V}_i^2]) = o_{\mathbb{P}}(1). \tag{A.23}$$

Similarly, we have that

$$\frac{1}{n}\sum_{i=1}^{n}(\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i]) = o_{\mathbb{P}}(1). \tag{A.24}$$

In addition, (A.14) and Hölder's inequality imply that

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\mathcal{V}_i^2] = O(1), \quad \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\mathcal{V}_i] = O(1).$$

As a result,

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{V}_i^2 = O_{\mathbb{P}}(1), \quad \frac{1}{n}\sum_{i=1}^{n}\mathcal{V}_i = O_{\mathbb{P}}(1). \tag{A.25}$$

By Lemma A.6, (A.25), and Cauchy-Schwarz inequality,

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i^2 - \frac{1}{n}\sum_{i=1}^{n}\mathcal{V}_i^2\right| \le \frac{2}{n}\sum_{i=1}^{n}\mathcal{V}_i|\hat{\mathcal{V}}_i - \mathcal{V}_i| + \frac{1}{n}\sum_{i=1}^{n}(\hat{\mathcal{V}}_i - \mathcal{V}_i)^2\\ &\le 2\sqrt{\frac{1}{n}\sum_{i=1}^{n}\mathcal{V}_i^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{\mathcal{V}}_i - \mathcal{V}_i)^2} + \frac{1}{n}\sum_{i=1}^{n}(\hat{\mathcal{V}}_i - \mathcal{V}_i)^2 = o_{\mathbb{P}}(1).\end{aligned} \tag{A.26}$$

Similarly,

$$\left|\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{V}_i\right)^2\right| = o_{\mathbb{P}}(1). \tag{A.27}$$

By (A.23), (A.24), and (A.25),

$$\left| \frac{1}{n} \sum_{i=1}^{n} \mathcal{V}_i^2 - \left(\frac{1}{n} \sum_{i=1}^{n} \mathcal{V}_i \right)^2 - \sigma_+^2 \right| = o_{\mathbb{P}}(1). \tag{A.28}$$

Putting (A.26) - (A.28) together, we complete the proof of (A.22). $\square$

A.4 Inference with deterministic $(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i)$ and dependent assignments across units

Recall Theorem A.4 that

$$\mathcal{D} \cdot \sqrt{n}(\hat{\tau} - \tau^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i]) + o_{\mathbb{P}}(1).$$

This is true even when $(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{X}_i)$ are dependent as long as Assumption A.3 holds. If $\mathcal{V}_i$'s are observable, a valid confidence interval for τ^* can be derived if the distribution of $(1/\sqrt{n}) \sum_{i=1}^{n} (\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i])$ can be approximated. Specifically, assume that

$$\frac{(1/\sqrt{n}) \sum_{i=1}^{n} (\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i])}{\sqrt{(1/n) \mathrm{Var}[\sum_{i=1}^{n} \mathcal{V}_i]}} \stackrel{d}{\to} N(0, 1), \tag{A.29}$$

and there exists a conservative oracle variance estimator $\hat{\sigma}^{*2}$ based on $(\mathcal{V}_1, \ldots, \mathcal{V}_n)$ in the sense that

$$\frac{(1/n) \mathrm{Var}[\sum_{i=1}^{n} \mathcal{V}_i]}{\hat{\sigma}^{*2}} \le 1 + o_{\mathbb{P}}(1). \tag{A.30}$$

Then, $[\hat{\tau} - z_{1-\alpha/2} \hat{\sigma}^* / \sqrt{n} \mathcal{D}, \hat{\tau} + z_{1-\alpha/2} \hat{\sigma}^* / \sqrt{n} \mathcal{D}]$ is an asymptotically valid confidence interval for τ^*. Of course, this interval cannot be computed in practice because $\mathcal{V}_i$ is unobserved due to the unknown quantities including $\mathbb{E}[\Gamma_\theta], \mathbb{E}[\boldsymbol{\Gamma}_w], \mathbb{E}[\boldsymbol{\Gamma}_y], \mathbb{E}[\Gamma_{ww}]$, $\mathbb{E}[\Gamma_{wy}]$, and τ^*. A natural variance estimator can be obtained by replacing $\boldsymbol{\mathcal{V}} \triangleq (\mathcal{V}_1, \ldots, \mathcal{V}_n)$ with $\hat{\boldsymbol{\mathcal{V}}} \triangleq (\hat{\mathcal{V}}_1, \ldots, \hat{\mathcal{V}}_n)$ in $\hat{\sigma}^{*2}$. The following theorem makes this intuition rigorous for generic quadratic oracle variance estimators.

Theorem A.6. *Suppose there exists an oracle variance estimator $\hat{\sigma}^{*2}$ such that* *(i)* $\hat{\sigma}^{*2} = \boldsymbol{\mathcal{V}}^\top \boldsymbol{A}_n \boldsymbol{\mathcal{V}}/n$ *for some positive semidefinite (and potentially random) matrix* $\boldsymbol{A}_n$ *with* $\|\boldsymbol{A}_n\|_{\mathrm{op}} = O_{\mathbb{P}}(1)$*;*

(ii) $\hat{\sigma}^{*2}$ *is conservative in the sense that, for every* η *in a neighborhood of* α*,*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{(1/\sqrt{n})\sum_{i=1}^n (\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i])}{\hat{\sigma}^*}\right| \geq z_{1-\eta/2}\right) \leq \eta;$$

(iii) $1/\hat{\sigma}^{*2} = O_{\mathbb{P}}(1)$.

Let $\hat{\sigma}^2 = \hat{\boldsymbol{\mathcal{V}}}^\top \boldsymbol{A}_n \hat{\boldsymbol{\mathcal{V}}}/n$ *and*

$$\hat{C}_{1-\alpha} = [\hat{\tau} - z_{1-\alpha/2}\hat{\sigma}/\sqrt{n}\mathcal{D}, \hat{\tau} + z_{1-\alpha/2}\hat{\sigma}/\sqrt{n}\mathcal{D}].$$

Under Assumptions A.1, A.2, and A.3 with $q > 1/2$*, if* $\boldsymbol{\Pi}$ *be an solution of the DATE equation (2.14) and* $\bar{\Delta}_\pi \bar{\Delta}_y = o(1/\sqrt{n})$*,*

$$\liminf_{n\to\infty} \mathbb{P}\left(\tau^* \in \hat{C}_{1-\alpha}\right) \geq 1 - \alpha.$$

Proof. By Lemma A.6,

$$\frac{1}{n}\|\hat{\boldsymbol{\mathcal{V}}} - \boldsymbol{\mathcal{V}}\|_2^2 = \frac{1}{n}\sum_{i=1}^n (\hat{\mathcal{V}}_i - \mathcal{V}_i)^2 = o_{\mathbb{P}}(1).$$

Since $\boldsymbol{A}_n$ is positive semidefinite, for any $\epsilon \in (0,1)$,

$$(1-\epsilon)\hat{\sigma}^{*2} - \left(\frac{1}{\epsilon} - 1\right)\frac{1}{n}(\hat{\boldsymbol{\mathcal{V}}} - \boldsymbol{\mathcal{V}})^\top \boldsymbol{A}_n (\hat{\boldsymbol{\mathcal{V}}} - \boldsymbol{\mathcal{V}}) \leq \hat{\sigma}^2 \leq (1+\epsilon)\hat{\sigma}^{*2} + \left(\frac{1}{\epsilon} + 1\right)\frac{1}{n}(\hat{\boldsymbol{\mathcal{V}}} - \boldsymbol{\mathcal{V}})^\top \boldsymbol{A}_n (\hat{\boldsymbol{\mathcal{V}}} - \boldsymbol{\mathcal{V}})$$

Thus, for any $\epsilon \in (0,1)$,

$$\mathbb{P}\left(\hat{\sigma}^2 \notin [(1-\epsilon)\hat{\sigma}^{*2}, (1+\epsilon)\hat{\sigma}^{*2}]\right) = o(1).$$

By condition (iii), the above result implies that

$$\left|\frac{\hat{\sigma}}{\hat{\sigma}^*} - 1\right| = o_{\mathbb{P}}(1). \tag{A.31}$$

By Theorem A.4,

$$\mathcal{D} \cdot \sqrt{n}(\hat{\tau} - \tau^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i]) + o_{\mathbb{P}}(1).$$

It remains to show that

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{(1/\sqrt{n}) \sum_{i=1}^{n} (\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i])}{\hat{\sigma}} \right| \geq z_{1-\alpha/2} \right) \leq \alpha.$$

Let $\eta(\epsilon)$ be the quantity such that $z_{1-\eta(\epsilon)/2} = z_{1-\alpha/2} \cdot (1-\epsilon)$. For any sufficiently small ϵ such that $\eta(\epsilon)$ lies in the neighborhood of α in condition (ii),

$$\begin{aligned}
&\mathbb{P}\left(\left| \frac{(1/\sqrt{n}) \sum_{i=1}^{n} (\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i])}{\hat{\sigma}} \right| \geq z_{1-\alpha/2} \right) \\
&= \mathbb{P}\left(\left| \frac{(1/\sqrt{n}) \sum_{i=1}^{n} (\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i])}{\hat{\sigma}^*} \right| \geq z_{1-\alpha/2} \cdot \frac{\hat{\sigma}}{\hat{\sigma}^*} \right) \\
&\leq \mathbb{P}\left(\left| \frac{(1/\sqrt{n}) \sum_{i=1}^{n} (\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i])}{\hat{\sigma}^*} \right| \geq z_{1-\eta(\epsilon)/2} \right) + \mathbb{P}\left(\frac{\hat{\sigma}}{\hat{\sigma}^*} \leq 1-\epsilon \right).
\end{aligned}$$

By (A.31), when n tends to infinity,

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{(1/\sqrt{n}) \sum_{i=1}^{n} (\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i])}{\hat{\sigma}} \right| \geq z_{1-\alpha/2} \right) \leq \eta(\epsilon).$$

The proof is completed by letting $\epsilon \to 0$ and noting that $\lim_{\epsilon\to 0} \eta(\epsilon) = \alpha$. □

When $\boldsymbol{W}_i$'s are independent,

$$\hat{\sigma}^{*2} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathcal{V}_i - \bar{\mathcal{V}})^2.$$

Thus, $\boldsymbol{A}_n = (n/(n-1))(I_n - \mathbf{1}_n \mathbf{1}_n^T/n)$. Clearly, the condition (i) is satisfied because $\|\boldsymbol{A}_n\|_{\text{op}} = n/(n-1)$. Under the assumptions in Theorem A.5, the condition (ii) is satisfied. Moreover, we have shown that $\hat{\sigma}^{*2}$ converges to $\sigma_+^2 \geq \sigma^2 > 0$, and thus the condition (iii) is satisfied. Therefore, Theorem A.5 can be implied by Theorem A.6.

When $\mathcal{V}_i$'s are observed, the variance estimators are quadratic under nearly all types of dependent assignment mechanisms. With fixed potential outcomes, Theorem A.6 applies to completely randomized experiments [Hoeffding, 1951, Li and Ding, 2017], blocked and matched experiments [Pashley and Miratrix, 2021], two-stage randomized experiments [Ohlsson, 1989], and so on. Below, we prove the results for completely randomized experiments with fixed potential outcomes to illustrate how to apply Theorem A.6. The notation is chosen to mimic Theorem 5 and Proposition 3 in Li and Ding [2017].

Theorem A.7. *Assume that $(Y_{it}(1), Y_{it}(0))$ are fixed, and $\hat{\boldsymbol{\pi}}_i = \boldsymbol{\pi}_i$ as in Section 2 (while $(\hat{m}_{it}, \hat{\nu}_{it})$ are allowed to be non-zero). Consider a completely randomized experiments where the treatment assignments are sampled without replacement from Q possible assignments $\{\boldsymbol{w}_{[1]}, \ldots, \boldsymbol{w}_{[Q]}\}$ with n_q units assigned $\boldsymbol{w}_{[q]}$. Let $\boldsymbol{\Pi}$ be a solution of the DATE equation (2.14) with support $\{\boldsymbol{w}_{[1]}, \ldots, \boldsymbol{w}_{[Q]}\}$, and $\mathcal{V}_i(q)$ be the "potential outcome" for $\mathcal{V}_i$ where (Y_{it}, W_{it}) is replaced by $(Y_{it}(\boldsymbol{w}_{[q],t}), \boldsymbol{w}_{[q],t})$, i.e.,*

$$\mathcal{V}_i(q) = \frac{\boldsymbol{\Pi}(\boldsymbol{w}_{[q]})}{\hat{\boldsymbol{\pi}}_i(\boldsymbol{w}_{[q]})} \Bigg\{ (\mathbb{E}[\Gamma_{wy}] - \tau^* \mathbb{E}[\Gamma_{ww}]) - (\mathbb{E}[\boldsymbol{\Gamma}_y] - \tau^* \mathbb{E}[\boldsymbol{\Gamma}_w])^\top J \boldsymbol{w}_{[q]} + \mathbb{E}[\Gamma_\theta] \boldsymbol{w}_{[q]}^\top J \left(\tilde{\mathbf{Y}}_i(q) - \tau^* \boldsymbol{w}_{[q]} \right) - \mathbb{E}[\boldsymbol{\Gamma}_w]^\top J \left(\tilde{\mathbf{Y}}_i(q) - \tau^* \boldsymbol{w}_{[q]} \right) \Bigg\},$$

and $\tilde{\mathbf{Y}}_i(q) = \big(Y_{i1}(\boldsymbol{w}_{[q],1}) - \hat{m}_{i1} - \boldsymbol{w}_{[q],1}\hat{\nu}_{i1}, \ldots, Y_{iT}(\boldsymbol{w}_{[q],T}) - \hat{m}_{iT} - \boldsymbol{w}_{[q],1}\hat{\nu}_{iT} \big)$. Further, for any $q, r = 1, \ldots, Q$, let

$$S_q^2 = \frac{1}{n-1} \sum_{i=1}^n (\mathcal{V}_i(q) - \bar{\mathcal{V}}(q))^2, \quad S_{qr} = \frac{1}{n-1} \sum_{i=1}^n (\mathcal{V}_i(q) - \bar{\mathcal{V}}(q))(\mathcal{V}_i(r) - \bar{\mathcal{V}}(r)),$$

where $\bar{\mathcal{V}}(q) = (1/n) \sum_{i=1}^n \mathcal{V}_i(q)$. Define the variance estimate $\hat{\sigma}^2$ as

$$\hat{\sigma}^2 = \sum_{q=1}^Q \frac{n_q}{n} s_q^2, \quad \text{where } s_q^2 = \frac{1}{n_q - 1} \sum_{i:\boldsymbol{w}_i = \boldsymbol{w}_{[q]}} (\hat{\mathcal{V}}_i - \hat{\bar{\mathcal{V}}}(q))^2, \ \hat{\bar{\mathcal{V}}}(q) = \frac{1}{n_q} \sum_{i:\boldsymbol{w}_i = \boldsymbol{w}_{[q]}} \hat{\mathcal{V}}_i.$$

Further, define the confidence interval as

$$\hat{C}_{1-\alpha} = [\hat{\tau} - z_{1-\alpha/2}\hat{\sigma}/\sqrt{n}\mathcal{D}, \hat{\tau} + z_{1-\alpha/2}\hat{\sigma}/\sqrt{n}\mathcal{D}].$$

Assume that

(a) $Q = O(1)$ and $n_q/n \to \pi_q$ for some constant $\pi_q > 0$;

(b) for any $q, r = 1, \ldots, Q$, S_q^2 and S_{qr} have limiting values S_q^{*2}, S_{qr}^*;

(c) there exists a constant $c_\tau > 0$ such that $\sum_{q=1}^{Q} \pi_q S_q^{*2} > c_\tau$;

(d) there exists a constant $M < \infty$ such that $\max_{i,q}\{\|\tilde{\boldsymbol{Y}}_i(q)\|_2\} < M$.

Then,

$$\liminf_{n\to\infty} \mathbb{P}\left(\tau^* \in \hat{C}_{1-\alpha}\right) \geq 1 - \alpha.$$

Proof. By definition, for any $i \neq j \in [n]$ and $q \neq r \in [Q]$,

$$\mathbb{P}(\boldsymbol{W}_i = \boldsymbol{w}_{[q]}) = \frac{n_q}{n}, \quad \mathbb{P}(\boldsymbol{W}_i = \boldsymbol{W}_j = \boldsymbol{w}_{[q]}) = \frac{n_q(n_q-1)}{n(n-1)}, \quad \mathbb{P}(\boldsymbol{W}_i = \boldsymbol{w}_{[q]}, \boldsymbol{W}_j = \boldsymbol{w}_{[r]}) = \frac{n_q n_r}{n(n-1)}.$$

For any functions f and g on $[0,1]^T$,

$$\mathbb{E}[f(\boldsymbol{W}_i)] = \sum_{q=1}^{Q} \frac{n_q}{n} f(\boldsymbol{w}_{[q]}), \quad \mathbb{E}[g(\boldsymbol{W}_j)] = \sum_{q=1}^{Q} \frac{n_q}{n} g(\boldsymbol{w}_{[q]}),$$

$$\mathbb{E}[f^2(\boldsymbol{W}_i)] = \sum_{q=1}^{Q} \frac{n_q}{n} f^2(\boldsymbol{w}_{[q]}), \quad \mathbb{E}[g^2(\boldsymbol{W}_j)] = \sum_{q=1}^{Q} \frac{n_q}{n} g^2(\boldsymbol{w}_{[q]}),$$

and

$$\mathbb{E}[f(\boldsymbol{W}_i)g(\boldsymbol{W}_i)] = \sum_{q=1}^{Q} \frac{n_q}{n} f(\boldsymbol{w}_{[q]})g(\boldsymbol{w}_{[q]}),$$

$$\mathbb{E}[f(\boldsymbol{W}_i)g(\boldsymbol{W}_j)] = \sum_{q=1}^{Q} \frac{n_q(n_q-1)}{n(n-1)} f(\boldsymbol{w}_{[q]})g(\boldsymbol{w}_{[q]}) + \sum_{q\neq r} \frac{n_q n_r}{n(n-1)} f(\boldsymbol{w}_{[q]})g(\boldsymbol{w}_{[r]}).$$

As a result, for any $i \neq j$

$$\begin{aligned}
&\mathrm{Cov}(f(\boldsymbol{W}_i), g(\boldsymbol{W}_j)) = \mathbb{E}[f(\boldsymbol{W}_i)g(\boldsymbol{W}_j)] - \mathbb{E}[f(\boldsymbol{W}_i)]\mathbb{E}[g(\boldsymbol{W}_j)] \\
&= \sum_{q=1}^{Q} \left(\frac{n_q(n_q-1)}{n(n-1)} - \frac{n_q^2}{n^2}\right) f(\boldsymbol{w}_{[q]})g(\boldsymbol{w}_{[q]}) + \sum_{q\neq r} \left(\frac{n_q n_r}{n(n-1)} - \frac{n_q n_r}{n^2}\right) f(\boldsymbol{w}_{[q]})g(\boldsymbol{w}_{[r]})
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{q=1}^{Q} -\frac{n_q(n-n_q)}{n^2(n-1)} f(\boldsymbol{w}_{[q]}) g(\boldsymbol{w}_{[q]}) + \sum_{q \neq r} \frac{n_q n_r}{n^2(n-1)} f(\boldsymbol{w}_{[q]}) g(\boldsymbol{w}_{[r]}) \\
&= -\frac{1}{n-1} \sum_{q=1}^{Q} \frac{n_q}{n} f(\boldsymbol{w}_{[q]}) g(\boldsymbol{w}_{[q]}) + \frac{1}{n-1} \left(\sum_{q=1}^{Q} \frac{n_q}{n} f(\boldsymbol{w}_{[q]}) \right) \left(\sum_{q=1}^{Q} \frac{n_q}{n} g(\boldsymbol{w}_{[r]}) \right) \\
&= -\frac{1}{n-1} \left(\mathbb{E}[f(\boldsymbol{W}_i) g(\boldsymbol{W}_i)] - \mathbb{E}[f(\boldsymbol{W}_i)] \mathbb{E}[g(\boldsymbol{W}_i)] \right) \\
&= -\frac{1}{n-1} \operatorname{Cov}(f(\boldsymbol{W}_i), g(\boldsymbol{W}_i))
\end{aligned}
$$

By Cauchy-Schwarz inequality,

$$
\begin{aligned}
&|\operatorname{Cov}(f(\boldsymbol{W}_i), g(\boldsymbol{W}_j))| \leq \frac{1}{n-1} |\operatorname{Cov}(f(\boldsymbol{W}_i), g(\boldsymbol{W}_i))| \\
&\leq \frac{1}{n-1} \sqrt{\operatorname{Var}[f(\boldsymbol{W}_i)] \operatorname{Var}[g(\boldsymbol{W}_i)]} = \frac{1}{n-1} \sqrt{\operatorname{Var}[f(\boldsymbol{W}_i)] \operatorname{Var}[g(\boldsymbol{W}_j)]}.
\end{aligned}
$$

This implies that

$$
\rho_{ij} \leq \frac{1}{n-1} \Longrightarrow \rho_i \leq 2.
$$

It is then clear that Assumption A.3 holds under the condition (d). Further, since $\hat{\boldsymbol{\pi}}_i(\boldsymbol{w}_{[q]}) = \boldsymbol{\pi}_i(\boldsymbol{w}_{[q]}) = n_q/n$, the condition (a) implies Assumption A.2 and that $\bar{\Delta}_\pi \bar{\Delta}_y = 0$. On the other hand, Assumption A.1 holds because $\boldsymbol{W}_i$ is completely randomized. Therefore, it remains to check the condition (i) - (iii) in Theorem A.6 with

$$
\hat{\sigma}^{*2} = \sum_{q=1}^{Q} \frac{n_q}{n} s_q^{*2}, \quad \text{where } s_q^{*2} = \frac{1}{n_q - 1} \sum_{i: \boldsymbol{w}_i = \boldsymbol{w}_{[q]}} (\mathcal{V}_i - \bar{\mathcal{V}}(q))^2, \ \bar{\mathcal{V}}(q) = \frac{1}{n_q} \sum_{i: \boldsymbol{w}_i = \boldsymbol{w}_{[q]}} \mathcal{V}_i.
$$

In this case, $\boldsymbol{A}_n$ is a block-diagonal matrix with

$$
\boldsymbol{A}_{n, \mathcal{I}_q, \mathcal{I}_q} = \frac{n_q}{n_q - 1} \left(I_{n_q} - \frac{\mathbf{1}_{n_q} \mathbf{1}_{n_q}^\top}{n_q} \right),
$$

where $\mathcal{I}_q = \{i : \boldsymbol{W}_i = \boldsymbol{w}_{[q]}\}$. As a result,

$$
\|\boldsymbol{A}_n\|_{\text{op}} = \max_q \frac{n_q}{n_q - 1} = O(1).
$$

Thus, the condition (i) holds. The condition (ii) is implied by Proposition 3 in Li and Ding [2017] and the condition (iii) is implied by the condition (c). The theorem is then implied by Theorem A.6. □

A.5 Inference of RIPW estimators when $\bar{\delta}_\pi \bar{\delta}_y \neq o(1/\sqrt{n})$

In this section, we study the asymptotic inference on the RIPW estimator when one of the models is globally misspecified. Doubly robust inference is hard even for cross-sectional data [e.g. Benkeser et al., 2017]. Here, we focus on a practically relevant case where the researcher fits parametric models for both assignments and outcomes. Specifically, we consider a parametric family $f_{\boldsymbol{\kappa}}(\boldsymbol{w}, \boldsymbol{X}_i)$ for $\boldsymbol{\pi}_i(\boldsymbol{w})$, $g_{t,\boldsymbol{\phi}_t}(\boldsymbol{X}_i)$ for m_{it}, and $h_{t,\boldsymbol{\psi}_t}(\boldsymbol{X}_i)$ for ν_{it}, where $\boldsymbol{\kappa} \in \mathbb{R}^d$, $\boldsymbol{\phi}_t \in \mathbb{R}^{d_{\phi,t}}$, $\boldsymbol{\psi}_t \in \mathbb{R}^{d_{\psi,t}}$ are parameter vectors to estimate. We impose the standard regularity conditions on these parametric families.

Assumption A.6. *(a) $1/f_{\boldsymbol{\kappa}}(\boldsymbol{w}; \boldsymbol{x})$ is bounded away from zero uniformly over $(\boldsymbol{\kappa}, \boldsymbol{w}, \boldsymbol{x})$;*

(b) $\|\nabla^2_{\boldsymbol{\kappa}} f_{\boldsymbol{\kappa}}(\boldsymbol{w}; \boldsymbol{x})\|$ is uniformly bounded over $(\boldsymbol{\kappa}, \boldsymbol{w}, \boldsymbol{x})$;

(c) $\|\nabla^2_{\boldsymbol{\phi}_t} g_{t,\boldsymbol{\phi}_t}(\boldsymbol{x})\| + \|\nabla^2_{\boldsymbol{\psi}_t} h_{t,\boldsymbol{\psi}_t}(\boldsymbol{x})\|$ is uniformly bounded over $(\boldsymbol{\phi}_t, \boldsymbol{\psi}_t, \boldsymbol{x}, t)$.

To ease notation, we denote by $\boldsymbol{\phi}$ (resp. $\boldsymbol{\psi}$) the concatenation of $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_T$ (resp. $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_T$) and by $\boldsymbol{\theta}$ the concatenation of $\boldsymbol{\kappa}, \boldsymbol{\phi}, \boldsymbol{\psi}$. We assume that $\hat{\boldsymbol{\theta}}$ has an asymptotically linear expansion:

Assumption A.7. *For some pseudo parameter $\boldsymbol{\theta}'$,*

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}' + \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{C}_i + o_{\mathbb{P}}(n^{-1/2}),$$

where $\boldsymbol{C}_i$ is a function of $(\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{X}_i)$ that has zero mean and bounded second moment.

When $(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{W}_i)$ are independent or weakly dependent, Assumption A.7 holds under standard regularity conditions [Fan and Yao, 2003, Wooldridge, 2010] with root-n rate. In particular, $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}'\| = O_{\mathbb{P}}(1/\sqrt{n})$.

Let $\boldsymbol{\kappa}', \boldsymbol{\phi}'_t, \boldsymbol{\psi}'_t$ denote the corresponding elements of $\boldsymbol{\theta}'$. Further let

$$\boldsymbol{\pi}'_i(\boldsymbol{w}) = f_{\boldsymbol{\kappa}'}(\boldsymbol{w}, \boldsymbol{X}_i), \quad m'_{it} = g_{t,\boldsymbol{\phi}'_t}(\boldsymbol{X}_i), \quad \nu'_{it} = h_{t,\boldsymbol{\psi}'_t}(\boldsymbol{X}_i). \tag{A.32}$$

We say the assignment model is correctly specified if

$$\boldsymbol{\pi}_i(\boldsymbol{w}) = \boldsymbol{\pi}_i'(\boldsymbol{w}), \tag{A.33}$$

and the outcome model is correctly specified if

$$\boldsymbol{m}_i = \boldsymbol{m}_i', \quad \boldsymbol{\nu}_i = \boldsymbol{\nu}_i'. \tag{A.34}$$

When one of these two models is globally misspecified and the other one is correctly specified,

$$\bar{\delta}_\pi \bar{\delta}_y = O(1/\sqrt{n}).$$

Thus, Theorems 3.2 and 3.3 do not apply. Nevertheless, we prove that the RIPW estimator remains asymptotically linear though an additional term is added to $\mathcal{V}_i$ to account for the estimation uncertainty of $\hat{\boldsymbol{\theta}}$ under the assumption that units are independent. We leave the general dependent case for future research.

Since the result is very complicated, we first define several quantities. Let $\Gamma_\theta', \Gamma_{ww}', \Gamma_{wy}', \boldsymbol{\Gamma}_w', \boldsymbol{\Gamma}_y'$, $\Theta_i', \tilde{Y}_{it}'$ be the counterparts of $\Gamma_\theta, \Gamma_{ww}, \Gamma_{wy}, \boldsymbol{\Gamma}_w, \boldsymbol{\Gamma}_y, \Theta_i, \tilde{Y}_{it}$ with $(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i)$ replaced by $(\boldsymbol{\pi}_i', \boldsymbol{m}_i', \boldsymbol{\nu}_i')$. Further let

$$\boldsymbol{L}_i = \nabla_{\boldsymbol{\kappa}} \log f_{\boldsymbol{\kappa}'}(\boldsymbol{W}_i, X_i), \quad \boldsymbol{G}_i = \begin{bmatrix} \nabla_{\boldsymbol{\phi}_1} g_{1,\boldsymbol{\phi}_1'}(\boldsymbol{X}_i) & 0 & \cdots & 0 \\ 0 & \nabla_{\boldsymbol{\phi}_2} g_{2,\boldsymbol{\phi}_2'}(\boldsymbol{X}_i) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \nabla_{\boldsymbol{\phi}_T} g_{T,\boldsymbol{\phi}_T'}(\boldsymbol{X}_i) \end{bmatrix},$$

and

$$\boldsymbol{H}_i = \begin{bmatrix} \nabla_{\boldsymbol{\psi}_1} h_{1,\boldsymbol{\psi}_1'}(\boldsymbol{X}_i) & 0 & \cdots & 0 \\ 0 & \nabla_{\boldsymbol{\psi}_2} h_{2,\boldsymbol{\psi}_2'}(\boldsymbol{X}_i) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \nabla_{\boldsymbol{\psi}_T} h_{T,\boldsymbol{\psi}_T'}(\boldsymbol{X}_i) \end{bmatrix}.$$

For any vector $\boldsymbol{v}$ that has the same dimension as $\boldsymbol{\theta}$, let $\mathcal{P}_{\boldsymbol{\kappa}}(\boldsymbol{v}), \mathcal{P}_{\boldsymbol{\phi}}(\boldsymbol{v}), \mathcal{P}_{\boldsymbol{\psi}}(\boldsymbol{v})$ denote the subvectors corresponding to the positions of $\boldsymbol{\kappa}, \boldsymbol{\phi}, \boldsymbol{\psi}$ in $\boldsymbol{\theta}$. We define $\boldsymbol{A}$ and $\boldsymbol{B}$ as two vectors such that

$$\begin{aligned}
\mathcal{P}_{\boldsymbol{\kappa}}(\boldsymbol{A}) &= \frac{1}{n}\sum_{i=1}^{n}\Big\{ -\mathbb{E}[\Theta_i'\boldsymbol{L}_i]\mathbb{E}[\Gamma_{wy}'] - \mathbb{E}[\Theta_i'\boldsymbol{L}_i\boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i']\mathbb{E}[\Gamma_\theta'] + \mathbb{E}[\Theta_i'\boldsymbol{L}_i\tilde{\boldsymbol{Y}}_i'^\top J]\mathbb{E}[\boldsymbol{\Gamma}_w'] \\
&\qquad\qquad + \mathbb{E}[\Theta_i'\boldsymbol{L}_i\boldsymbol{W}_i^\top J]\mathbb{E}[\boldsymbol{\Gamma}_y']\Big\} \\
\mathcal{P}_{\boldsymbol{\phi}}(\boldsymbol{A}) &= \frac{1}{n}\sum_{i=1}^{n}\{\mathbb{E}[\Theta_i'\boldsymbol{G}_iJ]\mathbb{E}[\boldsymbol{\Gamma}_w'] - \mathbb{E}[\Theta_i'\boldsymbol{G}_iJ\boldsymbol{W}_i]\mathbb{E}[\Gamma_\theta']\} \\
\mathcal{P}_{\boldsymbol{\psi}}(\boldsymbol{A}) &= \frac{1}{n}\sum_{i=1}^{n}\{\mathbb{E}[\Theta_i'\boldsymbol{H}_i\operatorname{diag}(\boldsymbol{W}_i)J]\mathbb{E}[\boldsymbol{\Gamma}_w'] - \mathbb{E}[\Theta_i'\boldsymbol{H}_i\operatorname{diag}(\boldsymbol{W}_i)J\boldsymbol{W}_i]\mathbb{E}[\Gamma_\theta']\},
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{P}_{\boldsymbol{\kappa}}(\boldsymbol{B}) &= \frac{1}{n}\sum_{i=1}^{n}\big(2\mathbb{E}[\Theta_i'\boldsymbol{L}_i\boldsymbol{W}_i^\top J\boldsymbol{W}_i]\mathbb{E}[\boldsymbol{\Gamma}_w] - \mathbb{E}[\Theta_i'\boldsymbol{L}_i]\mathbb{E}[\Gamma_{ww}'] - \mathbb{E}[\Theta_i'\boldsymbol{L}_i\boldsymbol{W}_i^\top J\boldsymbol{W}_i]\mathbb{E}[\Gamma_\theta']\big) \\
\mathcal{P}_{\boldsymbol{\phi}}(\boldsymbol{B}) &= \mathcal{P}_{\boldsymbol{\psi}}(\boldsymbol{B}) = 0.
\end{aligned}$$

Theorem A.8. *Assume that $(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{W}_i, \boldsymbol{X}_i)$ are independent. Further assume that either (A.33) or (A.34) holds. In the setting of Theorem 3.2, under Assumptions A.6 and A.7,*

$$\mathcal{D}\cdot\sqrt{n}(\hat{\tau}(\boldsymbol{\Pi}) - \tau^*) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathcal{V}_i' + \mathcal{U}_i - \mathbb{E}[\mathcal{V}_i']) + o_{\mathbb{P}}(1/\sqrt{n}),$$

if either the assignment model or the outcome model is correctly specified. Above, $\mathcal{V}_i'$ is defined as in Theorem 3.2 with $(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i)$ replaced by $(\boldsymbol{\pi}_i', \boldsymbol{m}_i', \boldsymbol{\nu}_i')$ and

$$\mathcal{U}_i = \langle \boldsymbol{C}_i, \boldsymbol{A} - \boldsymbol{B}\tau^*\rangle, \quad \mathbb{E}[\mathcal{U}_i] = 0,$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are defined above.

Remark A.1. To make inference, we can replace $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}_i$ by their plug-in estimates. It is straightforward, though tedious, to prove that the plug-in variance estimator is consistent. Importantly, this does not require the researcher to know which model is misspecified apriori. Here we present a proof sketch to illustrate why this is true. First, we can define an oracle RIPW estimate $\hat{\tau}'$ with the model estimates (A.32). Under Assumptions A.6 and A.7, we can derive an asymptotically linear expansion of $\sqrt{n}(\hat{\tau} - \hat{\tau}')$ via the Taylor expansion. These estimates are deterministic and $\bar{\delta}'_\pi \bar{\delta}'_y = 0$ if (A.33) or (A.34) holds, we can apply Theorem A.4 to obtain an asymptotically linear expansion of $\sqrt{n}(\hat{\tau}' - \tau^*)$ where the influence function can be estimated consistently without knowing which model is misspecified. Adding up two asymptotically linear expansions yields our result.

Proof. Write $\hat{\tau}$ for $\hat{\tau}(\mathbf{\Pi})$. Further let

$$\hat{\tau}' = \frac{\mathcal{N}'}{\mathcal{D}'}, \quad \mathcal{N}' = \Gamma'_{wy}\Gamma'_\theta - \mathbf{\Gamma}_w^{'\top}\mathbf{\Gamma}'_y, \quad \mathcal{D}' = \Gamma'_{ww}\Gamma'_\theta - \mathbf{\Gamma}_w^{'\top}\mathbf{\Gamma}'_w.$$

We will show that

$$\mathcal{N} = \mathcal{N}' + \left\langle \boldsymbol{A}, \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}' \right\rangle + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right), \quad \mathcal{D} = \mathcal{D}' + \left\langle \boldsymbol{B}, \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}' \right\rangle + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right). \tag{A.35}$$

Under (A.35),

$$\begin{aligned}
\mathcal{D} \cdot \sqrt{n}(\hat{\tau} - \hat{\tau}') &= \sqrt{n}\left(\mathcal{N} - \mathcal{N}'\frac{\mathcal{D}}{\mathcal{D}'}\right) = \sqrt{n}\left(\mathcal{N} - \mathcal{N}' - \mathcal{N}'\frac{\mathcal{D} - \mathcal{D}'}{\mathcal{D}'}\right) \\
&= \sqrt{n}\left(\mathcal{N} - \mathcal{N}' - \hat{\tau}'(\mathcal{D} - \mathcal{D}')\right) = \left\langle \boldsymbol{A} - \boldsymbol{B}\hat{\tau}', \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}') \right\rangle + o_{\mathbb{P}}(1)
\end{aligned}$$

It is easy to see that the assumptions of Theorem A.4 are implied by independence, Assumption 3.3, and Assumption A.6. Since $\hat{\tau}'$ is the RIPW estimator with deterministic $(\boldsymbol{\pi}_i, \boldsymbol{m}_i, \boldsymbol{\nu}_i)$, Theorem A.4 implies that

$$\hat{\tau}' = \tau^* + o_{\mathbb{P}}(1).$$

By definition and Assumption (A.3), $\|\boldsymbol{A}\| + \|\boldsymbol{B}\| = O_{\mathbb{P}}(1)$. Under Assumption A.7, $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}'\| = O_{\mathbb{P}}(1/\sqrt{n})$. Together with Assumption A.7, it implies

$$\mathcal{D} \cdot \sqrt{n}(\hat{\tau} - \hat{\tau}') = \left\langle \boldsymbol{A} - \boldsymbol{B}\tau^*, \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}') \right\rangle + o_{\mathbb{P}}(1) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathcal{U}_i + o_{\mathbb{P}}(1).$$

By Theorem A.4 again,

$$\mathcal{D} \cdot \sqrt{n}(\hat{\tau}' - \tau^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathcal{V}_i' - \mathbb{E}[\mathcal{V}_i']) + o_{\mathbb{P}}(1).$$

Combining the two pieces yields the desired result.

Now we turn to proving (A.35). By Assumption A.6,

$$\Theta_i - \Theta_i' = \frac{\boldsymbol{\Pi}(\boldsymbol{W}_i)}{f_{\hat{\boldsymbol{\kappa}}}(\boldsymbol{W}_i, \boldsymbol{X}_i)} - \frac{\boldsymbol{\Pi}(\boldsymbol{W}_i)}{f_{\boldsymbol{\kappa}'}(\boldsymbol{W}_i, \boldsymbol{X}_i)} = -\Theta_i' \left\langle \boldsymbol{L}_i, \hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}' \right\rangle + O(1) \cdot \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}'\|^2. \tag{A.36}$$

where the $O(1)$ terms are uniformly bounded across all units. By (A.36),

$$\Gamma_\theta - \Gamma_\theta' = -\left\langle \frac{1}{n} \sum_{i=1}^{n} \Theta_i' \boldsymbol{L}_i, \hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}' \right\rangle + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$$

Similar to Lemma A.2, we can show

$$\frac{1}{n} \sum_{i=1}^{n} \Theta_i' \boldsymbol{L}_i = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\Theta_i' \boldsymbol{L}_i] + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right).$$

Thus,

$$\Gamma_\theta - \Gamma_\theta' = -\left\langle \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\Theta_i' \boldsymbol{L}_i], \hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}' \right\rangle + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right). \tag{A.37}$$

Using the same argument, we can prove that

$$\Gamma_{ww} - \Gamma_{ww}' = -\left\langle \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\Theta_i' \boldsymbol{L}_i \boldsymbol{W}_i^\top J \boldsymbol{W}_i], \hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}' \right\rangle + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right), \tag{A.38}$$

and

$$\boldsymbol{\Gamma}_w - \boldsymbol{\Gamma}_w' = -\left\langle \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\Theta_i' \boldsymbol{L}_i \boldsymbol{W}_i^\top J], \hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}' \right\rangle + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right), \tag{A.39}$$

where we define $\langle A, b\rangle$ to be $A^\top b$ for a matrix A and vector b. In particular,

$$|\Gamma_\theta - \Gamma'_\theta| + |\Gamma_{ww} - \Gamma'_{ww}| + \|\boldsymbol{\Gamma}_w - \boldsymbol{\Gamma}'_w\| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right).$$

Putting (A.37) - (A.39) and Lemma A.2 together, we have

$$\begin{aligned}
&\mathcal{D} - \mathcal{D}' \\
&= \Gamma'_{ww}(\Gamma_\theta - \Gamma'_\theta) + \Gamma'_\theta(\Gamma_{ww} - \Gamma'_{ww}) - 2\boldsymbol{\Gamma}'^{\top}_w(\boldsymbol{\Gamma}_w - \boldsymbol{\Gamma}'_w) + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right) \\
&= \mathbb{E}[\Gamma'_{ww}](\Gamma_\theta - \Gamma'_\theta) + \mathbb{E}[\Gamma'_\theta](\Gamma_{ww} - \Gamma'_{ww}) - 2\mathbb{E}[\boldsymbol{\Gamma}'_w]^\top(\boldsymbol{\Gamma}_w - \boldsymbol{\Gamma}'_w) + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right) \\
&= \left\langle \frac{1}{n}\sum_{i=1}^n \left(2\mathbb{E}[\Theta'_i\boldsymbol{L}_i\boldsymbol{W}_i^\top J]\mathbb{E}[\boldsymbol{\Gamma}_w] - \mathbb{E}[\Theta'_i\boldsymbol{L}_i]\mathbb{E}[\Gamma'_{ww}] - \mathbb{E}[\Theta'_i\boldsymbol{L}_i\boldsymbol{W}_i^\top J\boldsymbol{W}_i]\mathbb{E}[\Gamma'_\theta]\right), \hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}'\right\rangle \\
&\quad + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

By definition of $\boldsymbol{B}$,

$$\mathcal{D} - \mathcal{D}' = \left\langle \boldsymbol{B}, \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}'\right\rangle + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right).$$

This proves the second part of (A.35).

To prove the first part of (A.35), we first recall that

$$\tilde{\boldsymbol{Y}}_i = \boldsymbol{Y}_i - \boldsymbol{m}_i - \operatorname{diag}(\boldsymbol{W}_i)\boldsymbol{\nu}_i, \quad \tilde{\boldsymbol{Y}}'_i = \boldsymbol{Y}_i - \boldsymbol{m}'_i - \operatorname{diag}(\boldsymbol{W}_i)\boldsymbol{\nu}'_i.$$

Similar to (A.36), Assumption A.6 implies

$$\tilde{\boldsymbol{Y}}_i - \tilde{\boldsymbol{Y}}'_i = -\left\langle \boldsymbol{G}_i, \hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right\rangle - \left\langle \boldsymbol{H}_i \operatorname{diag}(\boldsymbol{W}_i), \hat{\boldsymbol{\psi}} - \boldsymbol{\psi}\right\rangle + O(1)\cdot\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2, \tag{A.40}$$

where the $O(1)$ terms are uniformly bounded across all units. Together with (A.36),

$$\begin{aligned}
&\Gamma_{wy} - \Gamma'_{wy} \\
&= \frac{1}{n}\sum_{i=1}^n \Theta'_i\boldsymbol{W}_i^\top J(\tilde{\boldsymbol{Y}}_i - \tilde{\boldsymbol{Y}}'_i) + \frac{1}{n}\sum_{i=1}^n(\Theta_i - \Theta'_i)\boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}'_i + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

$$= -\left\langle \frac{1}{n}\sum_{i=1}^{n} \Theta_i' \boldsymbol{G}_i J \boldsymbol{W}_i, \hat{\boldsymbol{\phi}} - \boldsymbol{\phi}' \right\rangle - \left\langle \frac{1}{n}\sum_{i=1}^{n} \Theta_i' \boldsymbol{H}_i \operatorname{diag}(\boldsymbol{W}_i) J \boldsymbol{W}_i, \hat{\boldsymbol{\psi}} - \boldsymbol{\psi}' \right\rangle$$
$$- \left\langle \frac{1}{n}\sum_{i=1}^{n} \Theta_i' \boldsymbol{L}_i \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i', \hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}' \right\rangle + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right).$$

Similar to (A.37) - (A.39),

$$\Gamma_{wy} - \Gamma_{wy}' = -\left\langle \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\Theta_i' \boldsymbol{G}_i J \boldsymbol{W}_i], \hat{\boldsymbol{\phi}} - \boldsymbol{\phi}' \right\rangle - \left\langle \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\Theta_i' \boldsymbol{H}_i \operatorname{diag}(\boldsymbol{W}_i) J \boldsymbol{W}_i], \hat{\boldsymbol{\psi}} - \boldsymbol{\psi}' \right\rangle$$
$$- \left\langle \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\Theta_i' \boldsymbol{L}_i \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i'], \hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}' \right\rangle + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right). \tag{A.41}$$

Using the same argument, we can show

$$\boldsymbol{\Gamma}_y - \boldsymbol{\Gamma}_y' = -\left\langle \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\Theta_i' \boldsymbol{G}_i J], \hat{\boldsymbol{\phi}} - \boldsymbol{\phi}' \right\rangle - \left\langle \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\Theta_i' \boldsymbol{H}_i \operatorname{diag}(\boldsymbol{W}_i) J], \hat{\boldsymbol{\psi}} - \boldsymbol{\psi}' \right\rangle$$
$$- \left\langle \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\Theta_i' \boldsymbol{L}_i \tilde{\boldsymbol{Y}}_i'^\top J], \hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}' \right\rangle + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right). \tag{A.42}$$

As a result,

$$\mathcal{N} - \mathcal{N}'$$
$$= \Gamma_{wy}'(\Gamma_\theta - \Gamma_\theta') + \Gamma_\theta'(\Gamma_{wy} - \Gamma_{wy}') - \boldsymbol{\Gamma}_w'^\top(\boldsymbol{\Gamma}_y - \boldsymbol{\Gamma}_y') - \boldsymbol{\Gamma}_y'^\top(\boldsymbol{\Gamma}_w - \boldsymbol{\Gamma}_w') + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$$
$$= \mathbb{E}[\Gamma_{wy}'](\Gamma_\theta - \Gamma_\theta') + \mathbb{E}[\Gamma_\theta'](\Gamma_{wy} - \Gamma_{wy}') - \mathbb{E}[\boldsymbol{\Gamma}_w']^\top(\boldsymbol{\Gamma}_y - \boldsymbol{\Gamma}_y') - \mathbb{E}[\boldsymbol{\Gamma}_y']^\top(\boldsymbol{\Gamma}_w - \boldsymbol{\Gamma}_w') + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$$
$$= \left\langle \frac{1}{n}\sum_{i=1}^{n} \Big\{ -\mathbb{E}[\Theta_i' \boldsymbol{L}_i]\mathbb{E}[\Gamma_{wy}'] - \mathbb{E}[\Theta_i' \boldsymbol{L}_i \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i']\mathbb{E}[\Gamma_\theta'] + \mathbb{E}[\Theta_i' \boldsymbol{L}_i \tilde{\boldsymbol{Y}}_i'^\top J]\mathbb{E}[\boldsymbol{\Gamma}_w'] \right.$$
$$\left. + \mathbb{E}[\Theta_i' \boldsymbol{L}_i \boldsymbol{W}_i^\top J]\mathbb{E}[\boldsymbol{\Gamma}_y'] \Big\}, \hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}' \right\rangle$$
$$+ \left\langle \frac{1}{n}\sum_{i=1}^{n} \left\{ \mathbb{E}[\Theta_i' \boldsymbol{G}_i J]\mathbb{E}[\boldsymbol{\Gamma}_w'] - \mathbb{E}[\Theta_i' \boldsymbol{G}_i J \boldsymbol{W}_i]\mathbb{E}[\Gamma_\theta'] \right\}, \hat{\boldsymbol{\phi}} - \boldsymbol{\phi}' \right\rangle$$
$$+ \left\langle \frac{1}{n}\sum_{i=1}^{n} \left\{ \mathbb{E}[\Theta_i' \boldsymbol{H}_i \operatorname{diag}(\boldsymbol{W}_i) J]\mathbb{E}[\boldsymbol{\Gamma}_w'] - \mathbb{E}[\Theta_i' \boldsymbol{H}_i \operatorname{diag}(\boldsymbol{W}_i) J \boldsymbol{W}_i]\mathbb{E}[\Gamma_\theta'] \right\}, \hat{\boldsymbol{\psi}} - \boldsymbol{\psi}' \right\rangle$$

$$+ o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right).$$

By definition of $\boldsymbol{A}$,

$$\mathcal{N} - \mathcal{N}' = \left\langle \boldsymbol{A}, \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}' \right\rangle + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right).$$

$\square$

A.6 Proof of Proposition 2.1: induced weights are non-negative

Proof. Let $\boldsymbol{\omega} = \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}]$. Then

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\mathrm{diag}(\boldsymbol{W})J(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])] \\
&= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\mathrm{diag}(\boldsymbol{W})(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])] - \frac{1}{T}\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\mathrm{diag}(\boldsymbol{W})\mathbf{1}_T\mathbf{1}_T^\top(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])] \\
&= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}] - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\mathrm{diag}(\boldsymbol{W})]\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}] - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[\boldsymbol{W}\left(\frac{\mathbf{1}_T^\top\boldsymbol{W}}{T}\right)\right] + \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}]\frac{\mathbf{1}_T^\top\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}]}{T} \\
&= \boldsymbol{\omega} - \mathrm{diag}(\boldsymbol{\omega})\boldsymbol{\omega} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[\boldsymbol{W}\left(\frac{\mathbf{1}_T^\top\boldsymbol{W}}{T}\right)\right] + \boldsymbol{\omega}\left(\frac{\mathbf{1}_T^\top\boldsymbol{\omega}}{T}\right).
\end{aligned}$$

By (2.18), for any t,

$$\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[\left\|\tilde{\boldsymbol{W}} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\tilde{\boldsymbol{W}}]\right\|_2^2\right]\xi_t = \boldsymbol{\omega}_t - \boldsymbol{\omega}_t^2 - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[\boldsymbol{W}_t\left(\frac{\mathbf{1}_T^\top\boldsymbol{W}}{T}\right)\right] + \boldsymbol{\omega}_t\left(\frac{\mathbf{1}_T^\top\boldsymbol{\omega}}{T}\right). \tag{A.43}$$

Now we consider two scenarios.

1. If $\boldsymbol{\omega}_t \le \mathbf{1}_T^\top\boldsymbol{\omega}/T$,

   $$\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[\boldsymbol{W}_t\left(\frac{\mathbf{1}_T^\top\boldsymbol{W}}{T}\right)\right] \le \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}_t] = \boldsymbol{\omega}_t.$$

   Then the right-hand side of (A.43) is lower bounded by

   $$\boldsymbol{\omega}_t - \boldsymbol{\omega}_t^2 - \boldsymbol{\omega}_t + \boldsymbol{\omega}_t\left(\frac{\mathbf{1}_T^\top\boldsymbol{\omega}}{T}\right) = \boldsymbol{\omega}_t\left(\frac{\mathbf{1}_T^\top\boldsymbol{\omega}}{T} - \boldsymbol{\omega}_t\right) \ge 0.$$

2. If $\boldsymbol{\omega}_t > \mathbf{1}_T^\top \boldsymbol{\omega}/T$,

$$\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[\boldsymbol{W}_t\left(\frac{\mathbf{1}_T^\top \boldsymbol{W}}{T}\right)\right] \leq \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[\frac{\mathbf{1}_T^\top \boldsymbol{W}}{T}\right] = \frac{\mathbf{1}_T^\top \boldsymbol{\omega}}{T}.$$

Then the right-hand side of (A.43) is lower bounded by

$$\boldsymbol{\omega}_t - \boldsymbol{\omega}_t^2 - \frac{\mathbf{1}_T^\top \boldsymbol{\omega}}{T} + \boldsymbol{\omega}_t\left(\frac{\mathbf{1}_T^\top \boldsymbol{\omega}}{T}\right) = \left(\boldsymbol{\omega}_t - \frac{\mathbf{1}_T^\top \boldsymbol{\omega}}{T}\right)(1-\boldsymbol{\omega}_t) \geq 0.$$

□

A.7 Extension to generalized DATE in Remark 2.3

We prove that, under Assumptions A.1 - A.3, $\hat{\tau}(\boldsymbol{\Pi};\zeta) = \tau^*(\xi;\zeta) + o_{\mathbb{P}}(1)$ if, further, $\hat{\boldsymbol{\pi}}_i = \boldsymbol{\pi}_i$ and $n\|\zeta\|_\infty = O(1)$.

To prove consistency, we just need to modify the proofs in Appendix A.3.1 and A.3.2 by redefining Θ_i as

$$\Theta_i = \frac{(n\zeta_i)\boldsymbol{\Pi}(\boldsymbol{W}_i)}{\boldsymbol{\pi}_i(\boldsymbol{W}_i)},$$

and redefining $\Gamma_\theta, \boldsymbol{\Gamma}_w, \Gamma_{ww}, \Gamma_{wy}, \boldsymbol{\Gamma}_y$ correspondingly. First, we can apply the same arguments to show that Lemma A.2, Lemma A.3, and Theorem A.2 continue to hold under the above assumptions. We are left to prove that

$$N_* = 0.$$

As in the proof of Lemma A.4, we assume without loss of generality that $\tau^*(\xi;\zeta) = 0$; otherwise, we replace $Y_{it}(1)$ by $Y_{it}(1) - \tau^*(\xi;\zeta)$ and the resulting $\hat{\tau}$ becomes $\hat{\tau} - \tau^*(\xi;\zeta)$. Note that this reduction relies on the fact that $\sum_{i=1}^n \zeta_i = 1$. Using the same argument, we can show that (A.11) continues to hold with the new definition of Θ_i and other related quantities, i.e.,

$$N_* = \frac{1}{n}\sum_{i=1}^n \left\{\mathbb{E}[\Theta_i J \boldsymbol{W}_i]\mathbb{E}[\Gamma_\theta] - \mathbb{E}[\Theta_i]\mathbb{E}[\boldsymbol{\Gamma}_w]\right\}^\top \mathbb{E}[\tilde{\boldsymbol{Y}}_i(0)]$$

$$+\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{E}[\Theta_i \boldsymbol{W}_i^\top J\,\mathrm{diag}(\boldsymbol{W}_i)]\mathbb{E}[\Gamma_\theta] - \mathbb{E}[\boldsymbol{\Gamma}_w]^\top \mathbb{E}[\Theta_i\,\mathrm{diag}(\boldsymbol{W}_i)]\right\}\tilde{\boldsymbol{\tau}}_i.$$

By (A.12),

$$\mathbb{E}[\Theta_i J\boldsymbol{W}_i] = n\zeta_i \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[J\boldsymbol{W}], \quad \mathbb{E}[\Theta_i] = n\zeta_i,$$

and

$$\mathbb{E}[\Theta_i \boldsymbol{W}_i^\top J\,\mathrm{diag}(\boldsymbol{W}_i)] = n\zeta_i \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W} J\,\mathrm{diag}(\boldsymbol{W})], \quad \mathbb{E}[\Theta_i\,\mathrm{diag}(\boldsymbol{W}_i)] = n\zeta_i \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\mathrm{diag}(\boldsymbol{W})].$$

Thus,

$$\Gamma_\theta = 1, \quad \boldsymbol{\Gamma}_w = \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[J\boldsymbol{W}].$$

Then,

$$\mathbb{E}[\Theta_i J\boldsymbol{W}_i]\mathbb{E}[\Gamma_\theta] - \mathbb{E}[\Theta_i]\mathbb{E}[\boldsymbol{\Gamma}_w] = n\zeta_i \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[J\boldsymbol{W}] - n\zeta_i \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[J\boldsymbol{W}] = 0,$$

and by DATE equation,

$$\begin{aligned}
&\mathbb{E}[\Theta_i \boldsymbol{W}_i^\top J\,\mathrm{diag}(\boldsymbol{W}_i)]\mathbb{E}[\Gamma_\theta] - \mathbb{E}[\boldsymbol{\Gamma}_w]^\top \mathbb{E}[\Theta_i\,\mathrm{diag}(\boldsymbol{W}_i)] \\
&= n\zeta_i \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])^\top J\,\mathrm{diag}(\boldsymbol{W})] \\
&= n\zeta_i \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])^\top J\boldsymbol{W}]\xi^\top.
\end{aligned}$$

Since we assume $\tau^* = 0$,

$$\begin{aligned}
N_* &= \sum_{i=1}^{n} \zeta_i \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])^\top J\boldsymbol{W}]\xi^\top \tilde{\boldsymbol{\tau}}_i \\
&= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])^\top J\boldsymbol{W}]\left(\sum_{i=1}^{n} \zeta_i \xi^\top \tilde{\boldsymbol{\tau}}_i\right) \\
&= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])^\top J\boldsymbol{W}]\tau^* = 0.
\end{aligned}$$

A.8 Proof of Theorem 4.1

Let $\boldsymbol{W}_{i,\text{ex}} = (\boldsymbol{W}_{i,0}, \boldsymbol{W}_{i,-1}, \ldots, \boldsymbol{W}_{i,-p}) \in \mathbb{R}^{T\times(p+1)}$, where $\boldsymbol{W}_{i,0} = \boldsymbol{W}_i$, and $\hat{\boldsymbol{\tau}}_{\text{ex}} = (\hat{\tau}_0, \hat{\tau}_{-1} \ldots, \hat{\tau}_{-p})$ where the entries are defined in (4.2). To derive the non-stochastic formula of $\hat{\boldsymbol{\tau}}_{\text{ex}}$, we can repeat the steps in the proof of Theorem A.1 by replacing $\boldsymbol{W}_i$ with $\boldsymbol{W}_{i,\text{ex}}$ in $\Gamma_{ww}, \boldsymbol{\Gamma}_w$, and Γ_{wy}. This results in

$$\hat{\boldsymbol{\tau}}_{\text{ex}} = \left\{\boldsymbol{\Gamma}_{ww,\text{ex}} - \frac{\boldsymbol{\Gamma}_{w,\text{ex}}^\top \boldsymbol{\Gamma}_{w,\text{ex}}}{\Gamma_\theta}\right\}^{-1} \left\{\boldsymbol{\Gamma}_{wy,\text{ex}} - \frac{\boldsymbol{\Gamma}_{w,\text{ex}}^\top \boldsymbol{\Gamma}_y}{\Gamma_\theta}\right\},$$

where

$$\boldsymbol{\Gamma}_{ww,\text{ex}} = \frac{1}{n}\sum_{i=1}^n \Theta_i \boldsymbol{W}_{i,\text{ex}}^\top J \boldsymbol{W}_{i,\text{ex}} \in \mathbb{R}^{(p+1)\times(p+1)}, \quad \boldsymbol{\Gamma}_{w,\text{ex}} \triangleq \frac{1}{n}\sum_{i=1}^n \Theta_i J \boldsymbol{W}_{i,\text{ex}} \in \mathbb{R}^{T\times(p+1)},$$

and

$$\boldsymbol{\Gamma}_{wy,\text{ex}} = \frac{1}{n}\sum_{i=1}^n \Theta_i \boldsymbol{W}_{i,\text{ex}}^\top J \boldsymbol{Y}_i \in \mathbb{R}^{(p+1)\times T}, \quad \boldsymbol{\Gamma}_y = \frac{1}{n}\sum_{i=1}^n \Theta_i J \boldsymbol{Y}_i \in \mathbb{R}^T.$$

Note that $\boldsymbol{Y}_i = \tilde{\boldsymbol{Y}}_i$ in this case since no regression adjustment is applied.

Following the same steps as in Appendix A.3.1, we can prove that

$$\hat{\boldsymbol{\tau}}_{\text{ex}} \xrightarrow{p} \left\{\mathbb{E}[\boldsymbol{\Gamma}_{ww,\text{ex}}] - \frac{\mathbb{E}[\boldsymbol{\Gamma}_{w,\text{ex}}]^\top \mathbb{E}[\boldsymbol{\Gamma}_{w,\text{ex}}]}{\mathbb{E}[\Gamma_\theta]}\right\}^{-1} \left\{\mathbb{E}[\boldsymbol{\Gamma}_{wy,\text{ex}}] - \frac{\mathbb{E}[\boldsymbol{\Gamma}_{w,\text{ex}}]^\top \mathbb{E}[\boldsymbol{\Gamma}_y]}{\mathbb{E}[\Gamma_\theta]}\right\}. \tag{A.44}$$

Since $\hat{\boldsymbol{\pi}}_i = \boldsymbol{\pi}_i$, by (A.12),

$$\mathbb{E}[\Gamma_\theta] = 1, \quad \mathbb{E}[\boldsymbol{\Gamma}_{ww,\text{ex}}] = \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}_{\text{ex}}^\top J \boldsymbol{W}_{\text{ex}}], \quad \mathbb{E}[\boldsymbol{\Gamma}_{w,\text{ex}}] = \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[J\boldsymbol{W}_{\text{ex}}].$$

Thus,

$$\mathbb{E}[\boldsymbol{\Gamma}_{ww,\text{ex}}] - \frac{\mathbb{E}[\boldsymbol{\Gamma}_{w,\text{ex}}]^\top \mathbb{E}[\boldsymbol{\Gamma}_{w,\text{ex}}]}{\mathbb{E}[\Gamma_\theta]} = \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[(\boldsymbol{W}_{\text{ex}} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}_{\text{ex}}])^\top J(\boldsymbol{W}_{\text{ex}} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}_{\text{ex}}])] \tag{A.45}$$

By assumption (4.1),

$$\boldsymbol{Y}_i = \boldsymbol{Y}_i(\mathbf{0}_{p+1}) + \sum_{\ell=0}^{p} \boldsymbol{W}_{i,-\ell}\tau_{i,-\ell} = \boldsymbol{Y}_i(\mathbf{0}_{p+1}) + \boldsymbol{W}_{i,\mathrm{ex}}\boldsymbol{\tau}_{i,\mathrm{ex}},$$

where

$$\boldsymbol{\tau}_{i,\mathrm{ex}} = (\tau_{i,0}, \tau_{i,-1}, \ldots, \tau_{i,-p}) \in \mathbb{R}^{p+1}.$$

Then

$$\begin{aligned}
\mathbb{E}[\boldsymbol{\Gamma}_{wy,\mathrm{ex}}] &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\Theta_i \boldsymbol{W}_{i,\mathrm{ex}}^{\top} J \boldsymbol{Y}_i] = \frac{1}{n}\sum_{i=1}^{n} \left\{\mathbb{E}[\Theta_i \boldsymbol{W}_{i,\mathrm{ex}}^{\top} J \boldsymbol{Y}_i(\mathbf{0}_{p+1})] + \mathbb{E}[\Theta_i \boldsymbol{W}_{i,\mathrm{ex}}^{\top} J \boldsymbol{W}_{i,\mathrm{ex}}]\boldsymbol{\tau}_{i,\mathrm{ex}}\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n} \left\{\mathbb{E}[\Theta_i \boldsymbol{W}_{i,\mathrm{ex}}^{\top} J]\mathbb{E}[\boldsymbol{Y}_i(\mathbf{0}_{p+1})] + \mathbb{E}[\Theta_i \boldsymbol{W}_{i,\mathrm{ex}}^{\top} J \boldsymbol{W}_{i,\mathrm{ex}}]\boldsymbol{\tau}_{i,\mathrm{ex}}\right\} \\
&= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}_{\mathrm{ex}}^{\top} J] \left\{\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[J\boldsymbol{Y}_i(\mathbf{0}_{p+1})]\right\} + \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}_{\mathrm{ex}}^{\top} J \boldsymbol{W}_{\mathrm{ex}}] \left\{\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\tau}_{i,\mathrm{ex}}\right\},
\end{aligned}$$

where the second last line uses the assumption that $\boldsymbol{\pi}_i(\boldsymbol{w}) = \mathbb{P}(\boldsymbol{W}_i = \boldsymbol{w} \mid \boldsymbol{Y}_i(\mathbf{0}_{p+1}))$ and the last line is a result of (A.12) and the fact that $J^2 = J$. Similarly,

$$\mathbb{E}[\boldsymbol{\Gamma}_y] = \frac{1}{n}\sum_{i=1}^{n} \left\{\mathbb{E}[J\boldsymbol{Y}_i(\mathbf{0}_{p+1})] + \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[J\boldsymbol{W}_{\mathrm{ex}}]\boldsymbol{\tau}_{i,\mathrm{ex}}\right\}.$$

Thus,

$$\mathbb{E}[\boldsymbol{\Gamma}_{wy,\mathrm{ex}}] - \frac{\mathbb{E}[\boldsymbol{\Gamma}_{w,\mathrm{ex}}]^{\top}\mathbb{E}[\boldsymbol{\Gamma}_y]}{\mathbb{E}[\Gamma_\theta]} = \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[(\boldsymbol{W}_{\mathrm{ex}} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}_{\mathrm{ex}}])^{\top} J (\boldsymbol{W}_{\mathrm{ex}} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}_{\mathrm{ex}}])] \left\{\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\tau}_{i,\mathrm{ex}}\right\}. \tag{A.46}$$

Combining (A.44), (A.45), and (A.46) together, we obtain that

$$\hat{\boldsymbol{\tau}}_{\mathrm{ex}} \xrightarrow{p} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\tau}_{i,\mathrm{ex}}.$$

A.9 Miscellaneous

Proposition A.1. *[Petrov [1975], p. 112, Theorem 5] Let $X_1, X_2, \ldots, X_n$ be independent random variables such that $\mathbb{E}[X_j] = 0$, for all j. Assume also $\mathbb{E}[X_j^2 g(X_j)] < \infty$ for some function g that is non-negative, even, and non-decreasing in the interval $x > 0$, with $x/g(x)$ being non-decreasing for $x > 0$. Write $B_n = \sum_j \mathrm{Var}[X_j]$. Then,*

$$d_K\left(\mathcal{L}\left(\frac{1}{\sqrt{B_n}}\sum_{j=1}^{n} X_j\right), N(0,1)\right) \leq \frac{A}{B_n g(\sqrt{B_n})}\sum_{j=1}^{n}\mathbb{E}\left[X_j^2 g(X_j)\right],$$

where A is a universal constant, $\mathcal{L}(\cdot)$ denotes the probability law, d_K denotes the Kolmogorov-Smirnov distance (i.e., the ℓ_∞-norm of the difference of CDFs)

Proposition A.2 (Theorem 2 of von Bahr and Esseen [1965])**.** *Let $\{Z_i\}_{i=1,\ldots,n}$ be independent mean-zero random variables. Then for any $a \in [0, 1)$,*

$$\mathbb{E}\left|\sum_{i=1}^{n} Z_i\right|^{1+a} \leq 2\sum_{i=1}^{n}\mathbb{E}|Z_i|^{1+a}.$$

B Inference with cross-fitted model estimates

B.1 Cross-fitted RIPW estimator and main result

We split the data into K almost equal-sized folds with $\mathcal{I}_k$ denoting the index sets of the k-th fold and $|\mathcal{I}_k| \in \{\lfloor n/K \rfloor, \lceil n/K \rceil\}$. For each $i \in \mathcal{I}_k$, we estimate $(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i)$ using $\{(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{W}_i) : i \notin \mathcal{I}_k\}$. When $\{(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{W}_i) : i \in [n]\}$ are independent, it is obvious that

$$\{(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i) : i \in \mathcal{I}_k\} \perp\!\!\!\perp \{(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{W}_i) : i \in \mathcal{I}_k\}.$$

We assume that

$$\frac{1}{T}\sum_{t=1}^{T}\hat{m}_{it} = \frac{1}{|\mathcal{I}_k|}\sum_{i\in\mathcal{I}_k}\hat{m}_{it} = 0, \quad \forall i \in \mathcal{I}_k, \quad t = 1, \ldots, T,$$

and

$$\frac{1}{|\mathcal{I}_k|}\sum_{i\in\mathcal{I}_k}\xi'\hat{\boldsymbol{\nu}}_i = 0. \tag{B.1}$$

Otherwise, we apply the transformation (3.1) and (3.2) in each fold to enforce the above.

For valid inference, we need an additional assumption on the stability of the estimates.

Assumption B.1. *There exist functions $\{\boldsymbol{\pi}_i' : i \in [n]\}$ which satisfy Assumption A.2 and vectors $\{(\boldsymbol{m}_i', \boldsymbol{\nu}_i') : i \in [n]\}$ which satisfy Assumption A.3, such that*

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{E}[(\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) - \boldsymbol{\pi}_i'(\boldsymbol{W}_i))^2] + \mathbb{E}[\|\hat{\boldsymbol{m}}_i - \boldsymbol{m}_i'\|_2^2] + \mathbb{E}[\|\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i'\|_2^2]\right\} = O(n^{-r}) \tag{B.2}$$

for some $r > 0$. Furthermore,

$$\boldsymbol{\pi}_i' = \boldsymbol{\pi}_i \textit{ for all } i, \quad \textit{or} \quad (\boldsymbol{m}_i', \boldsymbol{\nu}_i') = (\boldsymbol{m}_i, \boldsymbol{\nu}_i) \textit{ for all } i. \tag{B.3}$$

The condition (B.2) states that the estimates need to be asymptotically deterministic given the covariates. This is a very mild assumption. For example, when $\hat{\boldsymbol{\pi}}_i$ is estimated from a parametric model $\{f(\boldsymbol{X}_i;\boldsymbol{\theta}) : \boldsymbol{\theta} \in \mathbb{R}^d\}$ as $f(\boldsymbol{X}_i;\hat{\boldsymbol{\theta}})$, under standard regularity conditions, $\hat{\boldsymbol{\theta}}$ converges to a limit $\boldsymbol{\theta}_0$ even if the model is misspecified. As a result, $\hat{\boldsymbol{\pi}}_i$ converges to $\boldsymbol{\pi}_i' = f(\boldsymbol{X}_i;\boldsymbol{\theta}_0)$. Under certain smoothness assumption, the estimates converge in the standard parametric rate and thus (B.2) holds with $r = 1$. On the other hand, in the settings of Section 2, (B.2) is always satisfied with $\boldsymbol{\pi}_i' = \boldsymbol{\pi}_i$ and $\boldsymbol{m}_i' = \boldsymbol{\nu}_i = \mathbf{0}_T$. More generally, if $\bar{\delta}_\pi^2 + \bar{\delta}_y^2 = O(n^{-r})$, it is also satisfied with $\boldsymbol{\pi}_i' = \boldsymbol{\pi}_i$ and $(\boldsymbol{m}_i', \boldsymbol{\nu}_i') = (\boldsymbol{m}_i, \boldsymbol{\nu}_i)$. A similar assumption was considered for cross-sectional data by Chernozhukov et al. [2020].

The condition (B.3) allows one of the treatment and outcome models to be inconsistently estimated. This covers the settings in Section 2 where the outcome model does not need to be consistently estimated. It also covers the classical model-based inference in which case the assignment model can be arbitrarily misspecified.

Theorem B.1. *Assume that $\{(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{W}_i) : i \in [n]\}$ are independent. Let $\{(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i) : i \in [n]\}$ be estimates obtained from K-fold cross-fitting where $K = O(1)$. Under Assumptions A.1, A.2, A.4, and B.1,* *(i)* $\hat{\tau}(\mathbf{\Pi}) - \tau^*(\xi) = o_{\mathbb{P}}(1)$ *if* $\bar{\delta}_\pi \bar{\delta}_y = o(1)$*;*

(ii) Let $\hat{C}_{1-\alpha}$ *be the same confidence interval as in Theorem 3.3. Then*

$$\liminf_{n\to\infty} \mathbb{P}\left(\tau^*(\xi) \in \hat{C}_{1-\alpha}\right) \geq 1 - \alpha$$

if (a) $\bar{\delta}_\pi \bar{\delta}_y = o(1/\sqrt{n})$*, (b) Assumption B.1 holds with* $r > 1/2$*, and (c) (2.24) holds if* $(\Theta_i, \boldsymbol{Y}_i)$ *are replaced by* $(\mathbf{\Pi}(\boldsymbol{W}_i)/\boldsymbol{\pi}_i'(\boldsymbol{W}_i), \boldsymbol{Y}_i - \boldsymbol{m}_i' - \mathrm{diag}(\boldsymbol{W}_i)\boldsymbol{\nu}_i')$ *in the definition of* $\mathcal{V}_i$*.*

The proof of Theorem B.1 is quite involved because our cross-fitted estimator is non-standard. The standard cross-fitting [Chernozhukov et al., 2017] would compute $\hat{\tau}_k(\mathbf{\Pi})$ on $\mathcal{I}_k$ with $\{(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i) : i \in \mathcal{I}_k\}$ and then take $\hat{\tau}(\mathbf{\Pi})$ as the average of $\{\hat{\tau}_k(\mathbf{\Pi}) : k \in [K]\}$. Under the assumptions of Theorem 3.2, it is straightforward to show each $\hat{\tau}_k(\mathbf{\Pi})$ is asymptotically linear and hence their average $\hat{\tau}(\mathbf{\Pi})$. In contrast, our estimator only cross-fitted the nuisance parameters $\{(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i) : i \in [n]\}$ but compute $\hat{\tau}(\mathbf{\Pi})$ using the whole dataset. While it is theoretically convenient to deal with the standard cross-fitting estimator, the standard version would fit weighted TWFE regressions on merely n/K units which would cause instability when n is moderate as in many economic applications. For this reason, we opt for our version to max out the sample size for computing $\hat{\tau}(\mathbf{\Pi})$, even though the technical proofs are lengthier.

B.2 De-randomization

Cross-fitting involves random data splits which introduce operational variation into the final estimate. We propose a de-randomization procedure that mitigates this source of unnecessary uncertainty by averaging over multiple splits. In particular, we consider B independent splits and add a superscript (b) to denote the quantities involved in the b-th split.

In the proof presented in the next subsection, we will show in (B.14) that, for each given data split,

$$\mathcal{D}^{(b)} \cdot \sqrt{n}(\hat{\tau}^{(b)} - \tau^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathcal{V}_i' - \mathbb{E}[\mathcal{V}_i']) + o_{\mathbb{P}}(1),$$

Note that $\mathcal{V}_i'$ does not depend on b. We define the de-randomized cross-fitted RIPW estimate as

$$\hat{\tau} = \frac{\sum_{b=1}^{B} \mathcal{D}^{(b)} \hat{\tau}^{(b)}}{\sum_{b=1}^{B} \mathcal{D}^{(b)}}. \tag{B.4}$$

Then, when $B = O(1)$,

$$\left(\frac{1}{B} \sum_{b=1}^{B} \mathcal{D}^{(b)} \right) \cdot \sqrt{n}(\hat{\tau} - \tau^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathcal{V}_i' - \mathbb{E}[\mathcal{V}_i']) + o_{\mathbb{P}}(1).$$

Furthermore, in Section B.3.2 we show that

$$\frac{1}{n} \sum_{i=1}^{n} (\mathcal{V}_i' - \hat{\mathcal{V}}_i^{(b)})^2 = o_{\mathbb{P}}(1).$$

Denote by $\bar{\hat{\mathcal{V}}}_i$ the average influence function:

$$\bar{\hat{\mathcal{V}}}_i = \frac{1}{B} \sum_{b=1}^{B} \hat{\mathcal{V}}_i^{(b)}.$$

Then,

$$\frac{1}{n} \sum_{i=1}^{n} (\mathcal{V}_i' - \bar{\hat{\mathcal{V}}}_i)^2 = o_{\mathbb{P}}(1).$$

Therefore, we can estimate the variance of $\hat{\tau}$ by the sample variance of $\bar{\hat{\mathcal{V}}}_1, \ldots, \bar{\hat{\mathcal{V}}}_n$. This justifies the confidence interval stated in Algorithm 1.

B.3 Proof of Theorem B.1

For convenience, we assume that $m = n/K$ is an integer. All proofs in this subsection can be easily extended to the general case. Without loss of generality, we can assume that

$$\mathbb{E}[\bar{\Delta}_\pi^2] = \bar{\delta}_\pi^2 = \Omega(n^{-r}), \quad \mathbb{E}[\bar{\Delta}_y^2] = \bar{\delta}_y^2 = \Omega(n^{-r}), \tag{B.5}$$

where $a_n = \Omega(b_n)$ iff $b_n = O(a_n)$. Otherwise, we can replace $(\boldsymbol{\pi}'_i, \boldsymbol{m}'_i, \boldsymbol{\nu}'_i)$ by $(\boldsymbol{\pi}_i, \boldsymbol{m}_i, \boldsymbol{\nu}_i)$ without decreasing r.

We use a superscript (k) to denote the corresponding quantity in fold k, i.e.,

$$\Gamma_\theta^{(k)} \triangleq \frac{1}{m}\sum_{i\in\mathcal{I}_k}\Theta_i, \quad \Gamma_{ww}^{(k)} \triangleq \frac{1}{m}\sum_{i\in\mathcal{I}_k}\Theta_i \boldsymbol{W}_i^\top J \boldsymbol{W}_i, \quad \Gamma_{wy}^{(k)} \triangleq \frac{1}{m}\sum_{i\in\mathcal{I}_k}\Theta_i \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i,$$

$$\boldsymbol{\Gamma}_w^{(k)} \triangleq \frac{1}{m}\sum_{i\in\mathcal{I}_k}\Theta_i J \boldsymbol{W}_i, \quad \boldsymbol{\Gamma}_y^{(k)} \triangleq \frac{1}{m}\sum_{i\in\mathcal{I}_k}\Theta_i J \tilde{\boldsymbol{Y}}_i.$$

As in the proof of Theorem A.3, we assume $\tau^* = 0$ without loss of generality. Let $\left(\Gamma'_{wy}, \Gamma'_\theta, \boldsymbol{\Gamma}'_w, \boldsymbol{\Gamma}'_y\right)$ and $(\Theta'_i, \tilde{\boldsymbol{Y}}'_i, \tilde{\boldsymbol{\tau}}'_i)$ be the counterpart of $(\Gamma_{wy}, \Gamma_\theta, \boldsymbol{\Gamma}_w, \boldsymbol{\Gamma}_y)$ and $(\Theta_i, \tilde{\boldsymbol{Y}}_i, \tilde{\boldsymbol{\tau}}_i)$ with $(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i)$ replaced by $(\boldsymbol{\pi}'_i, \boldsymbol{m}'_i, \boldsymbol{\nu}'_i)$. We first claim that

$$\Gamma_{wy}\Gamma_\theta - \boldsymbol{\Gamma}_w^\top \boldsymbol{\Gamma}_y - \left\{\Gamma'_{wy}\Gamma'_\theta - \boldsymbol{\Gamma}_w'^\top \boldsymbol{\Gamma}'_y\right\} = O_{\mathbb{P}}\left(n^{-\min\{r,(r'+1)/2\}} + \sqrt{\mathbb{E}[\bar{\Delta}_\pi^2]}\cdot\sqrt{\mathbb{E}[\bar{\Delta}_y^2]}\right), \tag{B.6}$$

where $r' = r\omega/(2+\omega)$. The proof of (B.6) is relegated to the end. Here we prove the rest of the theorem under (B.6).

Note that $\Gamma'_{wy}\Gamma'_\theta - \boldsymbol{\Gamma}_w'^\top \boldsymbol{\Gamma}'_y$ is the numerator of $\hat{\tau}$ when $\{(\boldsymbol{\pi}'_i, \boldsymbol{m}'_i, \boldsymbol{\nu}'_i) : i = 1, \ldots, n\}$ are used as the estimates. Let

$$\delta'_{\pi i} = \sqrt{\mathbb{E}[(\boldsymbol{\pi}'_i(\boldsymbol{W}_i) - \boldsymbol{\pi}_i(\boldsymbol{W}_i))^2]}, \quad \delta'_{yi} = \sqrt{\mathbb{E}[\|\boldsymbol{m}'_i - \boldsymbol{m}_i\|_2^2] + \mathbb{E}[\|\boldsymbol{\nu}'_i - \boldsymbol{\nu}_i\|_2^2]}, \tag{B.7}$$

and

$$\bar{\delta}'_\pi = \sqrt{\frac{1}{n}\sum_{i=1}^n \delta'^2_{\pi i}}, \quad \bar{\delta}'_y = \sqrt{\frac{1}{n}\sum_{i=1}^n \delta'^2_{yi}}. \tag{B.8}$$

By Assumption B.1 and (B.5),

$$\bar{\delta}'^2_\pi = \frac{1}{n}\sum_{i=1}^n \delta'^2_{\pi i}$$

$$\leq \frac{2}{n}\sum_{i=1}^{n}\left\{\mathbb{E}[\Delta_{\pi i}^2]+\mathbb{E}[(\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i)-\boldsymbol{\pi}_i'(\boldsymbol{W}_i))^2]\right\}$$

$$= O\left(\mathbb{E}[\bar{\Delta}_\pi^2]+n^{-r}\right)=O\left(\mathbb{E}[\bar{\Delta}_\pi^2]\right). \tag{B.9}$$

Similarly,

$$\bar{\delta}_y'^2=\frac{1}{n}\sum_{i=1}^{n}\delta_{yi}'^2=O\left(\mathbb{E}[\bar{\Delta}_y^2]\right). \tag{B.10}$$

As a result,

$$\bar{\delta}_\pi'\bar{\delta}_y'=O\left(\sqrt{\mathbb{E}[\bar{\Delta}_\pi^2]}\cdot\sqrt{\mathbb{E}[\bar{\Delta}_y^2]}\right).$$

Note that Assumption A.4 implies Assumption A.3 with $q=1$. By Theorem A.2 and Theorem A.3,

$$\Gamma_{wy}'\Gamma_\theta'-\boldsymbol{\Gamma}_w'^\top\boldsymbol{\Gamma}_y'=\frac{1}{n}\sum_{i=1}^{n}(\mathcal{V}_i'-\mathbb{E}[\mathcal{V}_i'])+O_{\mathbb{P}}\left(\sqrt{\mathbb{E}[\bar{\Delta}_\pi^2]}\cdot\sqrt{\mathbb{E}[\bar{\Delta}_y^2]}\right)+o_{\mathbb{P}}(1/\sqrt{n}) \tag{B.11}$$

$$=O_{\mathbb{P}}\left(\sqrt{\mathbb{E}[\bar{\Delta}_\pi^2]}\cdot\sqrt{\mathbb{E}[\bar{\Delta}_y^2]}\right)+o_{\mathbb{P}}(1), \tag{B.12}$$

where

$$\mathcal{V}_i'=\Theta_i'\left\{\mathbb{E}[\Gamma_{wy}']-\mathbb{E}[\boldsymbol{\Gamma}_y']^\top J\boldsymbol{W}_i+\mathbb{E}[\Gamma_\theta']\boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i'-\mathbb{E}[\boldsymbol{\Gamma}_w']^\top J\tilde{\boldsymbol{Y}}_i'\right\}.$$

On the other hand, by (B.6),

$$\mathcal{D}(\hat{\tau}-\tau^*)=\Gamma_{wy}\Gamma_\theta-\boldsymbol{\Gamma}_w^\top\boldsymbol{\Gamma}_y=\Gamma_{wy}'\Gamma_\theta'-\boldsymbol{\Gamma}_w'^\top\boldsymbol{\Gamma}_y'+O_{\mathbb{P}}\left(n^{-\min\{r,(r'+1)/2\}}+\sqrt{\mathbb{E}[\bar{\Delta}_\pi^2]}\cdot\sqrt{\mathbb{E}[\bar{\Delta}_y^2]}\right). \tag{B.13}$$

When $\sqrt{\mathbb{E}[\bar{\Delta}_\pi^2]}\cdot\sqrt{\mathbb{E}[\bar{\Delta}_y^2]}=o(1)$, (B.12) and (B.13) imply that

$$\mathcal{D}(\hat{\tau}-\tau^*)=o_{\mathbb{P}}(1).$$

The consistency then follows from Lemma A.3.

When $\sqrt{\mathbb{E}[\bar{\Delta}_\pi^2]} \cdot \sqrt{\mathbb{E}[\bar{\Delta}_y^2]} = o(1/\sqrt{n})$ and $r > 1/2$, (B.12) and (B.13) imply that

$$\mathcal{D} \cdot \sqrt{n}(\hat{\tau} - \tau^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathcal{V}_i' - \mathbb{E}[\mathcal{V}_i']) + o_{\mathbb{P}}(1). \tag{B.14}$$

Let $\hat{\mathcal{V}}_i'$ denote the plug-in estimate of $\mathcal{V}_i'$ assuming that $(\boldsymbol{\pi}_i', \boldsymbol{m}_i', \boldsymbol{\nu}_i')$ is known, i.e.,

$$\hat{\mathcal{V}}_i' = \Theta_i' \Big\{ \Gamma_{wy}' - \boldsymbol{\Gamma}_y'^{\top} J \boldsymbol{W}_i + \Gamma_\theta' \boldsymbol{W}_i^{\top} J \tilde{\boldsymbol{Y}}_i' - \boldsymbol{\Gamma}_w'^{\top} J \tilde{\boldsymbol{Y}}_i' \Big\}. \tag{B.15}$$

By Lemma A.5, under Assumption A.5 (with $(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i) = (\boldsymbol{\pi}_i', \boldsymbol{m}_i', \boldsymbol{\nu}_i')$),

$$\frac{\mathcal{D} \cdot \sqrt{n}(\hat{\tau} - \tau^*)}{\sigma'} \xrightarrow{d} N(0,1) \text{ in Kolmogorov-Smirnov distance,}$$

where

$$\sigma'^2 = \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}(\mathcal{V}_i') \geq v_0.$$

Similar to (A.21), define

$$\sigma_+'^2 = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left(\mathcal{V}_i' - \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\mathcal{V}_i'] \right)^2 = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\mathcal{V}_i'^2] - \left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\mathcal{V}_i'] \right)^2.$$

Obviously, $\sigma_+'^2 \geq \sigma'^2$. Furthermore, define an oracle variance estimate $\hat{\sigma}'^2$ as

$$\hat{\sigma}'^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(\hat{\mathcal{V}}_i' - \frac{1}{n} \sum_{i=1}^{n} \hat{\mathcal{V}}_i' \right)^2 = \frac{n}{n-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{\mathcal{V}}_i'^2 - \left(\frac{1}{n} \sum_{i=1}^{n} \hat{\mathcal{V}}_i' \right)^2 \right\}.$$

Recalling (A.18) that

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(\hat{\mathcal{V}}_i - \frac{1}{n} \sum_{i=1}^{n} \hat{\mathcal{V}}_i \right)^2 = \frac{n}{n-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{\mathcal{V}}_i^2 - \left(\frac{1}{n} \sum_{i=1}^{n} \hat{\mathcal{V}}_i \right)^2 \right\}.$$

Similar to (A.22) in Theorem A.5, it remains to prove that

$$|\hat{\sigma}^2 - \sigma_+^{\prime 2}| = o_{\mathbb{P}}(1).$$

Using the same arguments as in Theorem A.5, we can prove that

$$|\hat{\sigma}^{\prime 2} - \sigma_+^{\prime 2}| = o_{\mathbb{P}}(1).$$

Therefore, the proof will be completed if

$$|\hat{\sigma}^2 - \hat{\sigma}^{\prime 2}| = o_{\mathbb{P}}(1). \tag{B.16}$$

We present the proof of (B.16) in the end.

B.3.1 Proof of (B.6)

Let $\left(\Gamma_{wy}^{\prime(k)}, \Gamma_{\theta}^{\prime(k)}, \boldsymbol{\Gamma}_{w}^{\prime(k)}, \boldsymbol{\Gamma}_{y}^{\prime(k)}\right)$ be the counterpart of $(\Gamma_{wy}^{(k)}, \Gamma_{\theta}^{(k)}, \boldsymbol{\Gamma}_{w}^{(k)}, \boldsymbol{\Gamma}_{y}^{(k)})$ with $(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i)$ replaced by $(\boldsymbol{\pi}_i', \boldsymbol{m}_i', \boldsymbol{\nu}_i')$. Since the proof is lengthy, we decompose it into seven steps.

Step 1 By triangle inequality and Cauchy-Schwarz inequality,

$$\begin{aligned}
&|\Gamma_{wy} - \Gamma_{wy}'| \\
&\le \frac{1}{n}\sum_{i=1}^n |\Theta_i \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i - \Theta_i' \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i'| \\
&\le \frac{1}{n}\sum_{i=1}^n \left|\Theta_i \boldsymbol{W}_i^\top J(\hat{\boldsymbol{m}}_i - \boldsymbol{m}_i')\right| + \frac{1}{n}\sum_{i=1}^n \left|\Theta_i \boldsymbol{W}_i^\top J \operatorname{diag}(\boldsymbol{W}_i)(\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i')\right| \\
&\quad + \frac{1}{n}\sum_{i=1}^n \left|(\Theta_i - \Theta_i')\boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i'\right| \\
&\le \sqrt{\left(\frac{1}{n}\sum_{i=1}^n \|\Theta_i \boldsymbol{W}_i^\top J\|_2^2\right)\left(\frac{1}{n}\sum_{i=1}^n \|\hat{\boldsymbol{m}}_i - \boldsymbol{m}_i'\|_2^2\right)} \\
&\quad + \sqrt{\left(\frac{1}{n}\sum_{i=1}^n \|\Theta_i \boldsymbol{W}_i^\top J \operatorname{diag}(\boldsymbol{W}_i)\|_2^2\right)\left(\frac{1}{n}\sum_{i=1}^n \|\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i'\|_2^2\right)}
\end{aligned}$$

$$+\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}\|(\Theta_i-\Theta_i')\boldsymbol{W}_i^\top J\|_2^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}\|\tilde{\boldsymbol{Y}}_i'\|_2^2\right)}.$$

By Assumption A.5 and Hölder's inequality,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\|\tilde{\boldsymbol{Y}}_i'\|_2^2] \le \left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\|\tilde{\boldsymbol{Y}}_i'\|_2^{2+\omega}]\right)^{2/(2+\omega)} = O(1).$$

By Markov's inequality,

$$\frac{1}{n}\sum_{i=1}^{n}\|\tilde{\boldsymbol{Y}}_i'\|_2^2 = O_{\mathbb{P}}(1). \tag{B.17}$$

By Assumption A.2 and the boundedness of $\|\boldsymbol{W}_i J\|_2$ and $\|\boldsymbol{W}_i J\,\mathrm{diag}(\boldsymbol{W}_i)\|_2$,

$$\frac{1}{n}\sum_{i=1}^{n}\|\Theta_i\boldsymbol{W}_i^\top J\|_2^2 = O(1), \quad \frac{1}{n}\sum_{i=1}^{n}\|\Theta_i\boldsymbol{W}_i^\top J\,\mathrm{diag}(\boldsymbol{W}_i)\|_2^2 = O(1),$$

and, further, by Markov's inequality,

$$\frac{1}{n}\sum_{i=1}^{n}\|(\Theta_i-\Theta_i')\boldsymbol{W}_i^\top J\|_2^2 = O_{\mathbb{P}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i)-\boldsymbol{\pi}_i'(\boldsymbol{W}_i))^2]\right).$$

Putting pieces together and using Assumption B.1, we arrive at

$$|\Gamma_{wy}-\Gamma_{wy}'| = O_{\mathbb{P}}(n^{-r/2}).$$

Similarly, we can prove that

$$|\Gamma_{wy}-\Gamma_{wy}'| + |\Gamma_\theta-\Gamma_\theta'| + \|\boldsymbol{\Gamma}_w-\boldsymbol{\Gamma}_w'\|_2 + \|\boldsymbol{\Gamma}_y-\boldsymbol{\Gamma}_y'\|_2 = O_{\mathbb{P}}(n^{-r/2}). \tag{B.18}$$

As a consequence,

$$\left|(\Gamma_{wy}-\Gamma_{wy}')(\Gamma_\theta-\Gamma_\theta') - (\boldsymbol{\Gamma}_w-\boldsymbol{\Gamma}_w')^\top(\boldsymbol{\Gamma}_y-\boldsymbol{\Gamma}_y')\right| = O_{\mathbb{P}}(n^{-r}). \tag{B.19}$$

Step 2 Note that Assumption A.4 implies Assumption A.3 with $q = 1$. By Lemma A.2,

$$\left|\Gamma'_\theta - \mathbb{E}[\Gamma'_\theta]\right| + \left|\Gamma'_{wy} - \mathbb{E}[\Gamma'_{wy}]\right| + \left\|\boldsymbol{\Gamma}'_w - \mathbb{E}[\boldsymbol{\Gamma}'_w]\right\|_2 + \left\|\boldsymbol{\Gamma}'_y - \mathbb{E}[\boldsymbol{\Gamma}'_y]\right\|_2 = O_{\mathbb{P}}\left(n^{-1/2}\right).$$

By (B.18), we have

$$\begin{aligned}\Big|&(\Gamma_{wy} - \Gamma'_{wy})(\Gamma'_\theta - \mathbb{E}[\Gamma'_\theta]) + (\Gamma'_{wy} - \mathbb{E}[\Gamma'_{wy}])(\Gamma_\theta - \Gamma'_\theta) \\ &- (\boldsymbol{\Gamma}_w - \boldsymbol{\Gamma}'_w)^\top(\boldsymbol{\Gamma}'_y - \mathbb{E}[\boldsymbol{\Gamma}'_y]) - (\boldsymbol{\Gamma}'_w - \mathbb{E}[\boldsymbol{\Gamma}'_w])^\top(\boldsymbol{\Gamma}_y - \boldsymbol{\Gamma}'_y)\Big| = O_{\mathbb{P}}(n^{-(r+1)/2}). \end{aligned} \tag{B.20}$$

Step 3 Note that

$$\Gamma_{wy} - \Gamma'_{wy} = \frac{1}{K}\sum_{k=1}^{K}\left(\Gamma_{wy}^{(k)} - \Gamma_{wy}^{\prime(k)}\right).$$

For each k,

$$\Gamma_{wy}^{(k)} - \Gamma_{wy}^{\prime(k)} = \frac{1}{m}\sum_{i\in\mathcal{I}_k}(\Theta_i \boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i - \Theta'_i \boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}'_i).$$

Under Assumption A.4, the summands are independent conditional on $\mathcal{D}_{-[k]} \triangleq \{(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{X}_i) : i \notin \mathcal{I}_k\}$. Let $\mathbb{E}^{(k)}$ and $\mathrm{Var}^{(k)}$ denote the expectation and variance conditional on $\mathcal{D}_{-[k]}$. By Chebyshev's inequality,

$$\begin{aligned}&\left(\Gamma_{wy}^{(k)} - \Gamma_{wy}^{\prime(k)} - \mathbb{E}^{(k)}[\Gamma_{wy}^{(k)} - \Gamma_{wy}^{\prime(k)}]\right)^2 \\ &= O_{\mathbb{P}}\left(\frac{1}{m^2}\sum_{i\in\mathcal{I}_k}\mathrm{Var}^{(k)}\left(\Theta_i \boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i - \Theta'_i \boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}'_i\right)\right) \\ &\overset{(i)}{=} O_{\mathbb{P}}\left(\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}^{(k)}\left(\Theta_i \boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i - \Theta'_i \boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}'_i\right)^2\right) \\ &\overset{(ii)}{=} O_{\mathbb{P}}\left(\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}\left(\Theta_i \boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i - \Theta'_i \boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}'_i\right)^2\right),\end{aligned} \tag{B.21}$$

where (i) follows from $K = O(1)$ and (ii) applies Markov's inequality. By Jensen's inequality and Cauchy-Schwarz inequality,

$$\begin{aligned}
&\mathbb{E}\left(\Theta_i \boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i - \Theta_i' \boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i'\right)^2 \\
&\le 3\Big\{\mathbb{E}\left(\Theta_i \boldsymbol{W}_i^\top J(\hat{\boldsymbol{m}}_i - \boldsymbol{m}_i')\right)^2 + \mathbb{E}\left(\Theta_i \boldsymbol{W}_i^\top J\,\mathrm{diag}(\boldsymbol{W}_i)(\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i')\right)^2 \\
&\qquad + \mathbb{E}\left((\Theta_i - \Theta_i')\boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i'\right)^2\Big\} \\
&\le 3\Big\{\mathbb{E}\left[\|\Theta_i \boldsymbol{W}_i^\top J\|_2^2 \cdot \|\hat{\boldsymbol{m}}_i - \boldsymbol{m}_i'\|_2^2\right] + \mathbb{E}\left[\|\Theta_i \boldsymbol{W}_i^\top J\,\mathrm{diag}(\boldsymbol{W}_i)\|_2^2 \cdot \|\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i'\|_2^2\right] \\
&\qquad + \mathbb{E}\left[(\Theta_i - \Theta_i')^2(\boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i')^2\right]\Big\} \\
&\le C\left\{\mathbb{E}\left[(\hat{\boldsymbol{m}}_i - \boldsymbol{m}_i')^2\right] + \mathbb{E}\left[(\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i')^2\right] + \mathbb{E}\left[(\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) - \boldsymbol{\pi}_i'(\boldsymbol{W}_i))^2(\tilde{\boldsymbol{Y}}_i')^2\right]\right\},
\end{aligned} \tag{B.22}$$

where C is a constant that only depends on c_π and T. The second term can be bounded by

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[(\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) - \boldsymbol{\pi}_i'(\boldsymbol{W}_i))^2(\tilde{\boldsymbol{Y}}_i')^2\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n (\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) - \boldsymbol{\pi}_i'(\boldsymbol{W}_i))^2(\tilde{\boldsymbol{Y}}_i')^2\right] \\
&\overset{(i)}{\le} \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^n (\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) - \boldsymbol{\pi}_i'(\boldsymbol{W}_i))^{2(1+2/\omega)}\right)^{\omega/(2+\omega)}\left(\frac{1}{n}\sum_{i=1}^n (\tilde{\boldsymbol{Y}}_i')^{2+\omega}\right)^{2/(2+\omega)}\right] \\
&\overset{(ii)}{\le} \left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[(\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) - \boldsymbol{\pi}_i'(\boldsymbol{W}_i))^{2(1+2/\omega)}\right]\right)^{\omega/(2+\omega)}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[(\tilde{\boldsymbol{Y}}_i')^{2+\omega}\right]\right)^{2/(2+\omega)} \\
&\overset{(iii)}{\le} \left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}(\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) - \boldsymbol{\pi}_i'(\boldsymbol{W}_i))^2\right)^{\omega/(2+\omega)}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[(\tilde{\boldsymbol{Y}}_i')^{2+\omega}\right]\right)^{2/(2+\omega)},
\end{aligned}$$

where (i) applies the Hölder's inequality for sums, (ii) applies the Hölder's inequality that $\mathbb{E}[XY] \le \mathbb{E}[X^{(2+\omega)/\omega}]^{\omega/(2+\omega)}\mathbb{E}[Y^{(2+\omega)/2}]^{2/(2+\omega)}$, and (iii) uses the fact that $|\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) - \boldsymbol{\pi}_i'(\boldsymbol{W}_i)| \le 1$. By Assumptions A.5 and B.1,

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[(\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) - \boldsymbol{\pi}_i'(\boldsymbol{W}_i))^2(\tilde{\boldsymbol{Y}}_i')^2\right] = O\left(n^{-r\omega/(2+\omega)}\right) = O\left(n^{-r'}\right) \tag{B.23}$$

(B.22) and (B.23) together imply that

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\Theta_i\boldsymbol{W}_i^{\top}J\tilde{\boldsymbol{Y}}_i-\Theta_i'\boldsymbol{W}_i^{\top}J\tilde{\boldsymbol{Y}}_i'\right)^2=O\left(n^{-r'}\right). \tag{B.24}$$

By (B.21), for each k,

$$\Gamma_{wy}^{(k)}-\Gamma_{wy}^{\prime(k)}-\mathbb{E}^{(k)}[\Gamma_{wy}^{(k)}-\Gamma_{wy}^{\prime(k)}]=O_{\mathbb{P}}\left(n^{-(r'+1)/2}\right).$$

Since $K=O(1)$, it implies that

$$\left|\Gamma_{wy}-\Gamma_{wy}'-\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}^{(k)}[\Gamma_{wy}^{(k)}-\Gamma_{wy}^{\prime(k)}]\right|=O_{\mathbb{P}}\left(n^{-(r'+1)/2}\right).$$

Similarly, we have

$$\begin{aligned}&\left|\Gamma_{\theta}-\Gamma_{\theta}'-\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}^{(k)}[\Gamma_{\theta}^{(k)}-\Gamma_{\theta}^{\prime(k)}]\right|+\left\|\boldsymbol{\Gamma}_{w}-\boldsymbol{\Gamma}_{w}'-\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}^{(k)}[\boldsymbol{\Gamma}_{w}^{(k)}-\boldsymbol{\Gamma}_{w}^{\prime(k)}]\right\|_2\\&+\left\|\boldsymbol{\Gamma}_{y}-\boldsymbol{\Gamma}_{y}'-\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}^{(k)}[\boldsymbol{\Gamma}_{y}^{(k)}-\boldsymbol{\Gamma}_{y}^{\prime(k)}]\right\|_2=O_{\mathbb{P}}\left(n^{-(r'+1)/2}\right).\end{aligned}$$

By Lemma A.2,

$$|\mathbb{E}[\Gamma_{\theta}']|+|\mathbb{E}[\Gamma_{wy}']|+\|\mathbb{E}[\boldsymbol{\Gamma}_{w}']\|_2+\|\mathbb{E}[\boldsymbol{\Gamma}_{y}']\|_2=O(1).$$

Therefore,

$$\begin{aligned}&\Bigg|\left(\Gamma_{wy}-\Gamma_{wy}'-\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}^{(k)}[\Gamma_{wy}^{(k)}-\Gamma_{wy}^{\prime(k)}]\right)\mathbb{E}[\Gamma_{\theta}']\\&+\mathbb{E}[\Gamma_{wy}']\left(\Gamma_{\theta}-\Gamma_{\theta}'-\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}^{(k)}[\Gamma_{\theta}^{(k)}-\Gamma_{\theta}^{\prime(k)}]\right)\\&-\left(\boldsymbol{\Gamma}_{w}-\boldsymbol{\Gamma}_{w}'-\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}^{(k)}[\boldsymbol{\Gamma}_{w}^{(k)}-\boldsymbol{\Gamma}_{w}^{\prime(k)}]\right)^{\top}\mathbb{E}[\boldsymbol{\Gamma}_{y}']\end{aligned}$$

$$
\begin{aligned}
&- \mathbb{E}[\boldsymbol{\Gamma}'_w]^\top \left(\boldsymbol{\Gamma}_y - \boldsymbol{\Gamma}'_y - \frac{1}{K}\sum_{k=1}^K \mathbb{E}^{(k)}[\boldsymbol{\Gamma}_y^{(k)} - \boldsymbol{\Gamma}_y'^{(k)}] \right) \Bigg| \\
&= O_{\mathbb{P}}(n^{-(r'+1)/2}).
\end{aligned} \tag{B.25}
$$

Step 4 Note that $\mathbb{E}[\Gamma'_{wy}]\mathbb{E}[\Gamma'_\theta] - \mathbb{E}[\boldsymbol{\Gamma}'_w]^\top \mathbb{E}[\boldsymbol{\Gamma}'_y]$ is the limit of $\mathcal{D} \cdot \sqrt{n}(\hat{\tau} - \tau^*)$ when $\{(\boldsymbol{\pi}'_i, \boldsymbol{m}'_i, \boldsymbol{\nu}'_i) : i = 1, \ldots, n\}$ are plugged in as the estimates. Under Assumption B.1, either $\boldsymbol{\pi}'_i = \boldsymbol{\pi}_i$ for all $i \in [n]$ or $(\boldsymbol{m}'_i, \boldsymbol{\nu}'_i) = (\boldsymbol{m}_i, \boldsymbol{\nu}_i)$ for all $i \in [n]$. Then, by Lemma A.4,

$$
\mathbb{E}[\Gamma'_{wy}]\mathbb{E}[\Gamma'_\theta] - \mathbb{E}[\boldsymbol{\Gamma}'_w]^\top \mathbb{E}[\boldsymbol{\Gamma}'_y] = 0. \tag{B.26}
$$

Step 5 We shall prove that

$$
\left| \frac{1}{K}\sum_{k=1}^K \mathbb{E}^{(k)}[\Gamma_{wy}^{(k)}]\mathbb{E}[\Gamma'_\theta] - \mathbb{E}[\boldsymbol{\Gamma}'_w]^\top \mathbb{E}^{(k)}[\boldsymbol{\Gamma}_y^{(k)}] \right| = O\left(\sqrt{\mathbb{E}[\bar{\Delta}_\pi^2]} \cdot \sqrt{\mathbb{E}[\bar{\Delta}_y^2]} \right). \tag{B.27}
$$

By definition, we can write

$$
\Delta_{\pi i} = \sqrt{\mathbb{E}^{(k)}[(\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) - \boldsymbol{\pi}_i(\boldsymbol{W}_i))^2]}, \quad \Delta_{yi} = \sqrt{\mathbb{E}^{(k)}[\|\hat{\boldsymbol{m}}_i - \boldsymbol{m}_i\|_2^2] + \mathbb{E}^{(k)}[\|\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i\|_2^2]}, \quad \forall i \in \mathcal{I}_k.
$$

By Assumption A.1 and A.4,

$$
\begin{aligned}
\mathbb{E}^{(k)}[\Gamma_{wy}^{(k)}] &= \frac{1}{m}\sum_{i\in\mathcal{I}_k} \mathbb{E}^{(k)}[\Theta_i \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i] \\
&= \frac{1}{m}\sum_{i\in\mathcal{I}_k} \mathbb{E}^{(k)}[\Theta_i \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i(0)] + \frac{1}{m}\sum_{i\in\mathcal{I}_k} \mathbb{E}^{(k)}[\Theta_i \boldsymbol{W}_i^\top J \operatorname{diag}(\boldsymbol{W}_i)\tilde{\boldsymbol{\tau}}_i] \\
&= \frac{1}{m}\sum_{i\in\mathcal{I}_k} \mathbb{E}^{(k)}[\Theta_i \boldsymbol{W}_i^\top J]\mathbb{E}^{(k)}[\tilde{\boldsymbol{Y}}_i(0)] + \frac{1}{m}\sum_{i\in\mathcal{I}_k} \mathbb{E}^{(k)}[\Theta_i \boldsymbol{W}_i^\top J \operatorname{diag}(\boldsymbol{W}_i)]\mathbb{E}^{(k)}[\tilde{\boldsymbol{\tau}}_i].
\end{aligned}
$$

Similarly,

$$
\mathbb{E}^{(k)}[\boldsymbol{\Gamma}_y^{(k)}] = \frac{1}{m}\sum_{i\in\mathcal{I}_k} \mathbb{E}^{(k)}[\Theta_i J]\mathbb{E}^{(k)}[\tilde{\boldsymbol{Y}}_i(0)] + \frac{1}{m}\sum_{i\in\mathcal{I}_k} \mathbb{E}^{(k)}[\Theta_i J \operatorname{diag}(\boldsymbol{W}_i)]\mathbb{E}^{(k)}[\tilde{\boldsymbol{\tau}}_i].
$$

Putting the pieces together and using the fact that $\mathbb{E}[\boldsymbol{\Gamma}'_w]^\top J = \mathbb{E}[\boldsymbol{\Gamma}'_w]^\top, \mathbb{E}^{(k)}[\boldsymbol{\Gamma}_w]^\top J =$

$\mathbb{E}^{(k)}[\boldsymbol{\Gamma}_w]^\top$,

$$
\begin{aligned}
&\mathbb{E}^{(k)}[\Gamma_{wy}^{(k)}]\mathbb{E}[\Gamma'_\theta] - \mathbb{E}[\boldsymbol{\Gamma}'_w]^\top \mathbb{E}^{(k)}[\boldsymbol{\Gamma}_y^{(k)}] \\
&= \frac{1}{m}\sum_{i\in\mathcal{I}_k}\Big\{\mathbb{E}^{(k)}[\Theta_i \boldsymbol{W}_i^\top J\operatorname{diag}(\boldsymbol{W}_i)]\mathbb{E}[\Gamma'_\theta] - \mathbb{E}[\boldsymbol{\Gamma}'_w]^\top \mathbb{E}^{(k)}[\Theta_i J\operatorname{diag}(\boldsymbol{W}_i)]\Big\}\mathbb{E}^{(k)}[\tilde{\boldsymbol{\tau}}_i]. \\
&\quad + \frac{1}{m}\sum_{i\in\mathcal{I}_k}\Big\{\mathbb{E}^{(k)}[\Theta_i \boldsymbol{W}_i^\top J]\mathbb{E}[\Gamma'_\theta] - \mathbb{E}[\boldsymbol{\Gamma}'_w]^\top \mathbb{E}^{(k)}[\Theta_i]\Big\}\mathbb{E}^{(k)}[\tilde{\boldsymbol{Y}}_i(0)] \\
&\triangleq \frac{1}{m}\sum_{i\in\mathcal{I}_k}\boldsymbol{a}_{i1}^\top \mathbb{E}^{(k)}[\tilde{\boldsymbol{\tau}}_i] + \frac{1}{m}\sum_{i\in\mathcal{I}_k}\boldsymbol{a}_{i2}^\top \mathbb{E}^{(k)}[\tilde{\boldsymbol{Y}}_i(0)]
\end{aligned}
\tag{B.28}
$$

As in the proof of Theorem A.3. Let

$$
\Theta_i^* = \frac{\boldsymbol{\Pi}(\boldsymbol{W}_i)}{\boldsymbol{\pi}_i(\boldsymbol{W}_i)}, \quad \tilde{\boldsymbol{Y}}_i^* = \boldsymbol{Y}_i - \boldsymbol{m}_i - \operatorname{diag}(\boldsymbol{W}_i)\boldsymbol{\nu}_i,
$$

and $(\Gamma_\theta^*, \boldsymbol{\Gamma}_w^*)$ be the counterpart of $(\Gamma_\theta, \boldsymbol{\Gamma}_w)$ with $(\Theta_i, \tilde{\boldsymbol{Y}}_i)$ replaced by $(\Theta_i^*, \tilde{\boldsymbol{Y}}_i^*)$. Recalling (A.13) on page 63, there exists a constant C_1 that only depends on c_π and T such that

$$
\begin{aligned}
&\left|\mathbb{E}^{(k)}[(\Theta_i - \Theta_i^*)\boldsymbol{W}_i^\top J\operatorname{diag}(\boldsymbol{W}_i)]\right| + \left\|\mathbb{E}^{(k)}[(\Theta_i - \Theta_i^*)J\boldsymbol{W}_i]\right\|_2 + \left|\mathbb{E}^{(k)}[\Theta_i - \Theta_i^*]\right| \\
&\quad + \left\|\mathbb{E}^{(k)}[(\Theta_i - \Theta_i^*)J\operatorname{diag}(\boldsymbol{W}_i)]\right\|_{\mathrm{op}} \le C_1\Delta_{\pi i}.
\end{aligned}
\tag{B.29}
$$

where $\delta'_{\pi i}$ and $\bar{\delta}'_\pi$ are defined in (B.7) and (B.8), respectively. Then,

$$
\begin{aligned}
&\left|\mathbb{E}^{(k)}[\Theta_i \boldsymbol{W}_i^\top J\operatorname{diag}(\boldsymbol{W}_i)]\mathbb{E}[\Gamma'_\theta] - \left(\mathbb{E}[\Theta_i^* \boldsymbol{W}_i^\top J\operatorname{diag}(\boldsymbol{W}_i)]\mathbb{E}[\Gamma_\theta^*]\right)\right| \\
&\le \left|\mathbb{E}^{(k)}[\Theta_i \boldsymbol{W}_i^\top J\operatorname{diag}(\boldsymbol{W}_i)] - \mathbb{E}[\Theta_i^* \boldsymbol{W}_i^\top J\operatorname{diag}(\boldsymbol{W}_i)]\right| \cdot \mathbb{E}[\Gamma'_\theta] \\
&\quad + \mathbb{E}[\Theta_i^* \boldsymbol{W}_i^\top J\operatorname{diag}(\boldsymbol{W}_i)] \cdot |(\mathbb{E}[\Gamma'_\theta] - \mathbb{E}[\Gamma_\theta^*])| \\
&\le C_1(\mathbb{E}[\Gamma'_\theta]\cdot\Delta_{\pi i} + \mathbb{E}[\Theta_i^* \boldsymbol{W}_i^\top J\operatorname{diag}(\boldsymbol{W}_i)]\cdot\bar{\delta}'_\pi).
\end{aligned}
$$

Note that $\mathbb{E}[\Theta_i^* \boldsymbol{W}_i^\top J\operatorname{diag}(\boldsymbol{W}_i)] = \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}^\top J\operatorname{diag}(\boldsymbol{W})] \le T$ is a constant. By Assumption A.2, $\mathbb{E}[\Gamma_\theta] \le 1/c_\pi$. Thus,

$$
\left\|\boldsymbol{a}_{i1} - \left(\mathbb{E}[\Theta_i^* \boldsymbol{W}_i^\top J\operatorname{diag}(\boldsymbol{W}_i)]\mathbb{E}[\Gamma_\theta^*] - \mathbb{E}[\boldsymbol{\Gamma}_w^*]^\top \mathbb{E}[\Theta_i^* J\operatorname{diag}(\boldsymbol{W}_i)]\right)\right\|_2
$$

$$\leq C_2(\Delta_{\pi i} + \bar{\delta}'_\pi), \tag{B.30}$$

for some constant C_2 that only depends on c_π and T. Let

$$\boldsymbol{a}^*_{i1} = \mathbb{E}[\Theta^*_i \boldsymbol{W}_i^\top J \operatorname{diag}(\boldsymbol{W}_i)]\mathbb{E}[\Gamma^*_\theta] - \mathbb{E}[\boldsymbol{\Gamma}^*_w]^\top \mathbb{E}[\Theta^*_i J \operatorname{diag}(\boldsymbol{W}_i)].$$

Since we assume $\tau^* = 0$, $\|\mathbb{E}^{(k)}[\tilde{\boldsymbol{\tau}}_i]\|_2 = \|\mathbb{E}^{(k)}[\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i]\|_2 \leq \Delta_{yi}$,

$$\left| \frac{1}{m} \sum_{i \in \mathcal{I}_k} \boldsymbol{a}_{i1}^\top \mathbb{E}^{(k)}[\tilde{\boldsymbol{\tau}}_i] - \frac{1}{m} \sum_{i \in \mathcal{I}_k} \boldsymbol{a}_{i1}^{*\top} \mathbb{E}^{(k)}[\tilde{\boldsymbol{\tau}}_i] \right| \leq \frac{C_3}{m} \sum_{i \in \mathcal{I}_k} (\Delta_{\pi i} + \bar{\Delta}'_\pi)\Delta_{yi}. \tag{B.31}$$

On the other hand, by definition of Θ^*_i,

$$\mathbb{E}[\Theta^*_i \boldsymbol{W}_i^\top J \operatorname{diag}(\boldsymbol{W}_i)] = \mathbb{E}_{\boldsymbol{W} \sim \boldsymbol{\Pi}}[\boldsymbol{W}^\top J \operatorname{diag}(\boldsymbol{W})], \quad \mathbb{E}[\Theta^*_i J \operatorname{diag}(\boldsymbol{W}_i)] = \mathbb{E}_{\boldsymbol{W} \sim \boldsymbol{\Pi}}[J \operatorname{diag}(\boldsymbol{W})],$$

and

$$\mathbb{E}[\Gamma^*_\theta] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\Theta^*_i] = 1, \quad \mathbb{E}[\boldsymbol{\Gamma}^*_w] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\Theta^*_i J \boldsymbol{W}_i] = \mathbb{E}_{\boldsymbol{W} \sim \boldsymbol{\Pi}}[J\boldsymbol{W}].$$

Thus,

$$\begin{aligned}
\boldsymbol{a}^{*\top}_{i1} &= \mathbb{E}_{\boldsymbol{W} \sim \boldsymbol{\Pi}}[\boldsymbol{W}^\top J \operatorname{diag}(\boldsymbol{W})] - \mathbb{E}_{\boldsymbol{W} \sim \boldsymbol{\Pi}}[J\boldsymbol{W}]^\top \mathbb{E}_{\boldsymbol{W} \sim \boldsymbol{\Pi}}[J \operatorname{diag}(\boldsymbol{W}_i)] \\
&= \mathbb{E}_{\boldsymbol{W} \sim \boldsymbol{\Pi}}[(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W} \sim \boldsymbol{\Pi}}[\boldsymbol{W}])^\top J \operatorname{diag}(\boldsymbol{W})].
\end{aligned}$$

By the DATE equation,

$$\boldsymbol{a}^{*\top}_{i1} = \mathbb{E}_{\boldsymbol{W} \sim \boldsymbol{\Pi}}[(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W} \sim \boldsymbol{\Pi}}[\boldsymbol{W}])^\top J\boldsymbol{W}]\xi^\top.$$

As a consequence,

$$\frac{1}{m} \sum_{i \in \mathcal{I}_k} \boldsymbol{a}^{*\top}_{i1} \mathbb{E}^{(k)}[\tilde{\boldsymbol{\tau}}_i] = \mathbb{E}_{\boldsymbol{W} \sim \boldsymbol{\Pi}}[(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W} \sim \boldsymbol{\Pi}}[\boldsymbol{W}])^\top J\boldsymbol{W}] \cdot \left(\frac{1}{m} \sum_{i \in \mathcal{I}_k} \xi^\top \mathbb{E}^{(k)}[\tilde{\boldsymbol{\tau}}_i] \right). \tag{B.32}$$

Now we turn to the second and third terms of (B.28). Similar to (B.30), we can show that

$$\|\boldsymbol{a}_{i2}\|_2 = \|\boldsymbol{a}_{i2} - (\mathbb{E}[\boldsymbol{\Gamma}_w^*]^\top \mathbb{E}[\Gamma_\theta^*] - \mathbb{E}[\boldsymbol{\Gamma}_w^*]^\top \mathbb{E}[\Gamma_\theta^*])\|_2 \le C_3(\Delta_{\pi i} + \bar{\delta}'_\pi),$$

for some constant C_3 that only depends on c_π and T. By (A.3), $\mathbb{E}^{(k)}[\|\tilde{\boldsymbol{Y}}_i(0)\|_2] \le \Delta_{yi}$. Therefore,

$$\left|\frac{1}{m}\sum_{i\in\mathcal{I}_k} \boldsymbol{a}_{i2}^\top \mathbb{E}^{(k)}[\tilde{\boldsymbol{Y}}_i(0)]\right| \le \frac{C_3}{m}\sum_{i\in\mathcal{I}_k}(\Delta_{\pi i} + \bar{\Delta}'_\pi)\Delta_{yi}. \tag{B.33}$$

Putting (B.28), (B.31), (B.32), and (B.33) together, we arrive at

$$\begin{aligned}
&\left|\mathbb{E}^{(k)}[\Gamma_{wy}^{(k)}]\mathbb{E}[\Gamma'_\theta] - \mathbb{E}[\boldsymbol{\Gamma}'_w]^\top \mathbb{E}[\boldsymbol{\Gamma}_y^{(k)}] - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])^\top J \boldsymbol{W}] \cdot \left(\frac{1}{m}\sum_{i\in\mathcal{I}_k} \xi^\top \mathbb{E}^{(k)}[\tilde{\boldsymbol{\tau}}_i]\right)\right| \\
&\le \frac{C_4}{m}\sum_{i\in\mathcal{I}_k}(\Delta_{\pi i} + \bar{\delta}'_\pi)\Delta_{yi},
\end{aligned}$$

for some constant C_4 that only depends on c_π and T. Since $\tau^* = 0$,

$$\begin{aligned}
&\frac{1}{K}\sum_{k=1}^K \left(\frac{1}{m}\sum_{i\in\mathcal{I}_k} \xi^\top \mathbb{E}^{(k)}[\tilde{\boldsymbol{\tau}}_i]\right) = \frac{1}{K}\sum_{k=1}^K \left(\frac{1}{m}\sum_{i\in\mathcal{I}_k} \xi^\top (\boldsymbol{\tau}_i - \mathbb{E}^{(k)}[\hat{\boldsymbol{\nu}}_i])\right) \\
&= \tau^* - \frac{1}{K}\sum_{k=1}^K \left(\frac{1}{m}\sum_{i\in\mathcal{I}_k} \xi^\top \mathbb{E}^{(k)}[\hat{\boldsymbol{\nu}}_i]\right) = 0,
\end{aligned}$$

where the last step uses (B.1). Therefore, averaging over k and marginalizing over $\mathcal{D}_{-k}$ yields that

$$\left|\frac{1}{K}\sum_{k=1}^K \mathbb{E}^{(k)}[\Gamma_{wy}]\mathbb{E}[\Gamma'_\theta] - \mathbb{E}[\boldsymbol{\Gamma}'_w]^\top \mathbb{E}^{(k)}[\boldsymbol{\Gamma}_y^{(k)}]\right| = O_{\mathbb{P}}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[(\Delta_{\pi i} + \bar{\delta}'_\pi)\Delta_{yi}\right]\right).$$

By Cauchy-Schwarz inequality,

$$\left|\frac{1}{K}\sum_{k=1}^K \mathbb{E}^{(k)}[\Gamma_{wy}]\mathbb{E}[\Gamma'_\theta] - \mathbb{E}[\boldsymbol{\Gamma}'_w]^\top \mathbb{E}^{(k)}[\boldsymbol{\Gamma}_y^{(k)}]\right|$$

$$
\begin{aligned}
&= O_{\mathbb{P}}\left(\frac{1}{n}\sum_{i=1}^{n}\sqrt{\mathbb{E}\left[(\Delta_{\pi i}+\bar{\delta}'_{\pi})^2\right]}\sqrt{\mathbb{E}\left[\Delta_{yi}^2\right]}\right) \\
&= O_{\mathbb{P}}\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(\Delta_{\pi i}+\bar{\delta}'_{\pi})^2\right]}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\Delta_{yi}^2\right]}\right) \\
&= O_{\mathbb{P}}\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{E}[\Delta_{\pi i}^2]+\bar{\delta}'^2_{\pi}\right)}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\Delta_{yi}^2]}\right) \\
&= O_{\mathbb{P}}\left(\sqrt{\mathbb{E}[\bar{\Delta}_{\pi}^2]+\bar{\delta}'^2_{\pi}}\cdot\sqrt{\mathbb{E}[\bar{\Delta}_{y}^2]}\right).
\end{aligned}
$$

Therefore, (B.27) is proved by (B.9) and (B.10) on page 95.

Step 6 Next, we shall prove that

$$
\left|\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[\Gamma'_{wy}]\mathbb{E}^{(k)}[\Gamma_{\theta}^{(k)}-\Gamma_{\theta}^{'(k)}]-\mathbb{E}^{(k)}[\boldsymbol{\Gamma}_{w}^{(k)}-\boldsymbol{\Gamma}_{w}^{'(k)}]^{\top}\mathbb{E}[\boldsymbol{\Gamma}'_{y}]\right| = O\left(\sqrt{\mathbb{E}[\bar{\Delta}_{\pi}^2]}\cdot\sqrt{\mathbb{E}[\bar{\Delta}_{y}^2]}\right). \tag{B.34}
$$

Using the same argument as (B.29), we can show that

$$
\|\mathbb{E}[(\Theta'_i-\Theta_i)J\boldsymbol{W}_i]\|_2+|\mathbb{E}[\Theta'_i-\Theta_i]|\leq C_1(\delta'_{\pi i}+\Delta_{\pi i}).
$$

Averaging over $i\in\mathcal{I}_k$, we obtain that

$$
|\mathbb{E}^{(k)}[\Gamma_{\theta}^{(k)}]-\mathbb{E}^{(k)}[\Gamma_{\theta}^{'(k)}]|+\|\mathbb{E}^{(k)}[\boldsymbol{\Gamma}_{w}^{(k)}]-\mathbb{E}^{(k)}[\boldsymbol{\Gamma}_{w}^{'(k)}]\|_2\leq C_1(\bar{\delta}'_{\pi}+\bar{\Delta}_{\pi})=O\left(\bar{\Delta}_{\pi}\right), \tag{B.35}
$$

where the last step uses (B.5). On the other hand,

$$
\begin{aligned}
\mathbb{E}[\Gamma'_{wy}] &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\Theta'_i\boldsymbol{W}_i^{\top}J\tilde{\boldsymbol{Y}}'_i] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\Theta'_i\boldsymbol{W}_i^{\top}J\tilde{\boldsymbol{Y}}'_i(0)]+\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\Theta'_i\boldsymbol{W}_i^{\top}J\,\mathrm{diag}(\boldsymbol{W}_i)\tilde{\boldsymbol{\tau}}'_i] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\Theta'_i\boldsymbol{W}_i^{\top}J]\mathbb{E}[\tilde{\boldsymbol{Y}}'_i(0)]+\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\Theta'_i\boldsymbol{W}_i^{\top}J\,\mathrm{diag}(\boldsymbol{W}_i)]\tilde{\boldsymbol{\tau}}'_i.
\end{aligned}
$$

Note that

$$\|\mathbb{E}[\Theta_i' \boldsymbol{W}_i^\top J]\|_2 \le \frac{\sqrt{T}}{c_\pi}, \quad \|\mathbb{E}[\Theta_i' \boldsymbol{W}_i^\top J \operatorname{diag}(\boldsymbol{W}_i)]\|_2 \le \frac{T}{c_\pi}, \quad \|\mathbb{E}[\tilde{\boldsymbol{Y}}_i'(0)]\|_2 + \|\tilde{\boldsymbol{\tau}}_i'\|_2 \le \delta_{yi}'.$$

As a result, there exists a constant C_5 that only depends on c_π and T such that

$$|\mathbb{E}[\Gamma_{wy}']| \le \frac{C_5}{n}\sum_{i=1}^n \delta_{yi}' = O\left(\mathbb{E}[\bar{\Delta}_y^2]\right).$$

Similarly,

$$\|\mathbb{E}[\boldsymbol{\Gamma}_y']\|_2 = O\left(\mathbb{E}[\bar{\Delta}_y^2]\right).$$

Together with (B.35), we prove (B.34).

Step 7 Consider the following decompositions:

$$\begin{aligned}
&\Gamma_{wy}\Gamma_\theta - \Gamma_{wy}'\Gamma_\theta' \\
&= (\Gamma_{wy} - \Gamma_{wy}')(\Gamma_\theta - \Gamma_\theta') \\
&+ (\Gamma_{wy} - \Gamma_{wy}')(\Gamma_\theta' - \mathbb{E}[\Gamma_\theta']) + (\Gamma_{wy}' - \mathbb{E}[\Gamma_{wy}'])(\Gamma_\theta - \Gamma_\theta') \\
&+ \left(\Gamma_{wy} - \Gamma_{wy}' - \frac{1}{K}\sum_{k=1}^K \mathbb{E}^{(k)}[\Gamma_{wy}^{(k)} - \Gamma_{wy}^{\prime(k)}]\right)\mathbb{E}[\Gamma_\theta'] \\
&+ \mathbb{E}[\Gamma_{wy}']\left(\Gamma_\theta - \Gamma_\theta' - \frac{1}{K}\sum_{k=1}^K \mathbb{E}^{(k)}[\Gamma_\theta^{(k)} - \Gamma_\theta^{\prime(k)}]\right) \\
&- \frac{1}{K}\left(\sum_{k=1}^K \mathbb{E}^{(k)}[\Gamma_{wy}^{\prime(k)}]\right)\cdot\mathbb{E}[\Gamma_\theta'] \\
&+ \frac{1}{K}\left(\sum_{k=1}^K \mathbb{E}^{(k)}[\Gamma_{wy}^{(k)}]\right)\cdot\mathbb{E}[\Gamma_\theta'] \\
&+ \mathbb{E}[\Gamma_{wy}']\cdot\frac{1}{K}\left(\sum_{k=1}^K \mathbb{E}^{(k)}[\Gamma_\theta^{(k)} - \Gamma_\theta^{\prime(k)}]\right),
\end{aligned}$$

and

$$
\begin{aligned}
&\mathbf{\Gamma}_w^\top \mathbf{\Gamma}_y - \mathbf{\Gamma}_w'^\top \mathbf{\Gamma}_y' \\
&= (\mathbf{\Gamma}_w - \mathbf{\Gamma}_w')^\top (\mathbf{\Gamma}_y - \mathbf{\Gamma}_y') \\
&+ (\mathbf{\Gamma}_w - \mathbf{\Gamma}_w')^\top (\mathbf{\Gamma}_y' - \mathbb{E}[\mathbf{\Gamma}_y']) + (\mathbf{\Gamma}_w' - \mathbb{E}[\mathbf{\Gamma}_w'])^\top (\mathbf{\Gamma}_y - \mathbf{\Gamma}_y') \\
&+ \left(\mathbf{\Gamma}_w - \mathbf{\Gamma}_w' - \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}^{(k)}[\mathbf{\Gamma}_w^{(k)} - \mathbf{\Gamma}_w'^{(k)}] \right)^\top \mathbb{E}[\mathbf{\Gamma}_y'] \\
&+ \mathbb{E}[\mathbf{\Gamma}_w']^\top \left(\mathbf{\Gamma}_y - \mathbf{\Gamma}_y' - \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}^{(k)}[\mathbf{\Gamma}_y^{(k)} - \mathbf{\Gamma}_y'^{(k)}] \right) \\
&- \mathbb{E}[\mathbf{\Gamma}_w']^\top \frac{1}{K} \left(\sum_{k=1}^{K} \mathbb{E}^{(k)}[\mathbf{\Gamma}_y'^{(k)}] \right) \\
&+ \mathbb{E}[\mathbf{\Gamma}_w']^\top \frac{1}{K} \left(\sum_{k=1}^{K} \mathbb{E}^{(k)}[\mathbf{\Gamma}_y^{(k)}] \right) \\
&+ \frac{1}{K} \left(\sum_{k=1}^{K} \mathbb{E}^{(k)}[\mathbf{\Gamma}_w^{(k)} - \mathbf{\Gamma}_w'^{(k)}] \right)^\top \mathbb{E}[\mathbf{\Gamma}_y'].
\end{aligned}
$$

Since $(\boldsymbol{\pi}_i', \boldsymbol{m}_i', \boldsymbol{\tau}_i')$ are deterministic,

$$
\frac{1}{K} \left(\sum_{k=1}^{K} \mathbb{E}^{(k)}[\Gamma_{wy}'^{(k)}] \right) \cdot \mathbb{E}[\Gamma_\theta'] = \frac{1}{K} \left(\sum_{k=1}^{K} \mathbb{E}[\Gamma_{wy}'^{(k)}] \right) \cdot \mathbb{E}[\Gamma_\theta'] = \mathbb{E}[\Gamma_{wy}'] \mathbb{E}[\Gamma_\theta'],
$$

and

$$
\mathbb{E}[\mathbf{\Gamma}_w']^\top \frac{1}{K} \left(\sum_{k=1}^{K} \mathbb{E}^{(k)}[\mathbf{\Gamma}_y'^{(k)}] \right) = \mathbb{E}[\mathbf{\Gamma}_w']^\top \frac{1}{K} \left(\sum_{k=1}^{K} \mathbb{E}[\mathbf{\Gamma}_y'^{(k)}] \right) = \mathbb{E}[\mathbf{\Gamma}_w']^\top \mathbb{E}[\mathbf{\Gamma}_y'].
$$

By (B.19), (B.20), (B.25), (B.26), (B.27), (B.34), and triangle inequality,

$$
\Gamma_{wy}\Gamma_\theta - \mathbf{\Gamma}_w^\top \mathbf{\Gamma}_y - \left\{ \Gamma_{wy}'\Gamma_\theta' - \mathbf{\Gamma}_w'^\top \mathbf{\Gamma}_y' \right\} = O_{\mathbb{P}}\left(n^{-r} + n^{-(r+1)/2} + n^{-(r'+1)/2} + \sqrt{\mathbb{E}[\bar{\Delta}_\pi^2]} \cdot \sqrt{\mathbb{E}[\bar{\Delta}_y^2]} \right).
$$

The proof of (B.6) is then completed.

B.3.2 Proof of (B.16)

Let

$$\hat{\mathcal{V}}_i'' = \Theta_i'\Big\{\Gamma_{wy} - \boldsymbol{\Gamma}_y^\top J\boldsymbol{W}_i + \Gamma_\theta \boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i' - \boldsymbol{\Gamma}_w^\top J\tilde{\boldsymbol{Y}}_i'\Big\}. \tag{B.36}$$

Recalling the definition of $\hat{\mathcal{V}}_i'$ in (B.15) on page 96,

$$\begin{aligned}
|\hat{\mathcal{V}}_i' - \hat{\mathcal{V}}_i''| &\le |\Gamma_{wy} - \Gamma_{wy}'| \cdot \Theta_i' + \|\boldsymbol{\Gamma}_y - \boldsymbol{\Gamma}_y'\|_2 \cdot \|\Theta_i' J\boldsymbol{W}_i\|_2 \\
&\qquad + |\Gamma_\theta - \Gamma_\theta'| \cdot |\Theta_i'\boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i'| + \|\boldsymbol{\Gamma}_w - \boldsymbol{\Gamma}_w'\|_2 \cdot \|\Theta_i' J\tilde{\boldsymbol{Y}}_i'\|_2 \\
&\quad \le \left\{|\Gamma_{wy} - \Gamma_{wy}'| + \|\boldsymbol{\Gamma}_y - \boldsymbol{\Gamma}_y'\|_2 + |\Gamma_\theta - \Gamma_\theta'| + \|\boldsymbol{\Gamma}_w - \boldsymbol{\Gamma}_w'\|_2\right\} \\
&\qquad \cdot \Big\{\Theta_i' + \|\Theta_i' J\boldsymbol{W}_i\|_2 + |\Theta_i'\boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i'| + \|\Theta_i' J\tilde{\boldsymbol{Y}}_i'\|_2\Big\}
\end{aligned}$$

By Jensen's inequality and Cauchy-Schwarz inequality,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n (\hat{\mathcal{V}}_i' - \hat{\mathcal{V}}_i'')^2 \\
&\le 4\left\{|\Gamma_{wy} - \Gamma_{wy}'| + \|\boldsymbol{\Gamma}_y - \boldsymbol{\Gamma}_y'\|_2 + |\Gamma_\theta - \Gamma_\theta'| + \|\boldsymbol{\Gamma}_w - \boldsymbol{\Gamma}_w'\|_2\right\}^2 \\
&\quad \cdot \frac{1}{n}\sum_{i=1}^n \Big\{\Theta_i'^2 + \|\Theta_i' J\boldsymbol{W}_i\|_2^2 + |\Theta_i'\boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i'|^2 + \|\Theta_i' J\tilde{\boldsymbol{Y}}_i'\|_2^2\Big\} \\
&\le \frac{8T}{c_\pi^2}\left\{|\Gamma_{wy} - \Gamma_{wy}'|^2 + \|\boldsymbol{\Gamma}_y - \boldsymbol{\Gamma}_y'\|_2^2 + |\Gamma_\theta - \Gamma_\theta'|^2 + \|\boldsymbol{\Gamma}_w - \boldsymbol{\Gamma}_w'\|_2^2\right\} \cdot \frac{1}{n}\sum_{i=1}^n \Big\{1 + \|\tilde{\boldsymbol{Y}}_i'\|_2^2\Big\},
\end{aligned}$$

where the last inequality uses Assumption A.2. By (B.17) and (B.18) on page 98,

$$\frac{1}{n}\sum_{i=1}^n (\hat{\mathcal{V}}_i' - \hat{\mathcal{V}}_i'')^2 = O_{\mathbb{P}}(n^{-r}) = o_{\mathbb{P}}(1). \tag{B.37}$$

On the other hand, recalling the definition of $\hat{\mathcal{V}}_i$ in (A.15) on page 67,

$$\begin{aligned}
|\hat{\mathcal{V}}_i'' - \hat{\mathcal{V}}_i| &\le |\Gamma_{wy}| \cdot |\Theta_i' - \Theta_i| + \|\boldsymbol{\Gamma}_y\|_2 \cdot \|(\Theta_i' - \Theta_i)J\boldsymbol{W}_i\|_2 \\
&\qquad + |\Gamma_\theta| \cdot |\Theta_i'\boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i' - \Theta_i\boldsymbol{W}_i^\top J\tilde{\boldsymbol{Y}}_i| + \|\boldsymbol{\Gamma}_w\|_2 \cdot \|\Theta_i' J\tilde{\boldsymbol{Y}}_i' - \Theta_i J\tilde{\boldsymbol{Y}}_i\|_2
\end{aligned}$$

$$
\begin{aligned}
&+ \Theta_i \hat{\tau} \left\{ |\Gamma_{ww}| + |\boldsymbol{\Gamma}_w^\top J \boldsymbol{W}_i| + |\Gamma_\theta \boldsymbol{W}_i^\top J \boldsymbol{W}_i| + |\boldsymbol{\Gamma}_w^\top J \boldsymbol{W}_i| \right\} \\
\leq& \left\{ |\Gamma_{wy}| + \|\boldsymbol{\Gamma}_y\|_2 + |\Gamma_\theta| + \|\boldsymbol{\Gamma}_w\|_2 \right\} \cdot \Bigg\{ |\Theta_i' - \Theta_i| + \|(\Theta_i' - \Theta_i) J \boldsymbol{W}_i\|_2 \\
&+ |\Theta_i' \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i' - \Theta_i \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i| + \|\Theta_i' J \tilde{\boldsymbol{Y}}_i' - \Theta_i J \tilde{\boldsymbol{Y}}_i\|_2 \Bigg\} \\
&+ \Theta_i |\hat{\tau}| \left\{ |\Gamma_{ww}| + |\boldsymbol{\Gamma}_w^\top J \boldsymbol{W}_i| + |\Gamma_\theta \boldsymbol{W}_i^\top J \boldsymbol{W}_i| + |\boldsymbol{\Gamma}_w^\top J \boldsymbol{W}_i| \right\}.
\end{aligned}
$$

Since $\|J\boldsymbol{W}_i\|_2 \leq \sqrt{T}$,

$$
\|(\Theta_i' - \Theta_i) J \boldsymbol{W}_i\|_2 \leq \sqrt{T} |\Theta_i' - \Theta_i|.
$$

By triangle inequality,

$$
\begin{aligned}
&|\Theta_i' \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i' - \Theta_i \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i| \\
&\leq |\Theta_i \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i' - \Theta_i \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i| + |\Theta_i' \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i' - \Theta_i \boldsymbol{W}_i^\top J \tilde{\boldsymbol{Y}}_i'| \\
&\leq \frac{\sqrt{T}}{c_\pi} \|\tilde{\boldsymbol{Y}}`_i - \tilde{\boldsymbol{Y}}_i\|_2 + \sqrt{T} \|\tilde{\boldsymbol{Y}}_i'\|_2 \cdot |\Theta_i' - \Theta_i| \\
&= \frac{\sqrt{T}}{c_\pi} \|\hat{\boldsymbol{m}}_i - \boldsymbol{m}_i'\|_2 + \sqrt{T} \|\tilde{\boldsymbol{Y}}_i'\|_2 \cdot |\Theta_i' - \Theta_i|.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\|\Theta_i' J \tilde{\boldsymbol{Y}}_i' - \Theta_i J \tilde{\boldsymbol{Y}}_i\|_2 \\
&\leq \|\Theta_i J \tilde{\boldsymbol{Y}}_i' - \Theta_i J \tilde{\boldsymbol{Y}}_i\|_2 + \|\Theta_i' J \tilde{\boldsymbol{Y}}_i' - \Theta_i J \tilde{\boldsymbol{Y}}_i'\|_2 \\
&\leq \frac{1}{c_\pi} \|\tilde{\boldsymbol{Y}}_i' - \tilde{\boldsymbol{Y}}_i\|_2 + \|\tilde{\boldsymbol{Y}}_i'\|_2 \cdot |\Theta_i' - \Theta_i| \\
&= \frac{1}{c_\pi} \|\hat{\boldsymbol{m}}_i - \boldsymbol{m}_i'\|_2 + \|\tilde{\boldsymbol{Y}}_i'\|_2 \cdot |\Theta_i' - \Theta_i|
\end{aligned}
$$

Putting pieces together, we have that

$$
\begin{aligned}
&(\hat{\mathcal{V}}_i'' - \hat{\mathcal{V}}_i)^2 \\
&\leq C \left\{ |\Gamma_{wy}| + \|\boldsymbol{\Gamma}_y\|_2 + |\Gamma_\theta| + \|\boldsymbol{\Gamma}_w\|_2 \right\}^2 \left\{ |\Theta_i' - \Theta_i|^2 \cdot (1 + \|\tilde{\boldsymbol{Y}}_i'\|_2^2) + \|\hat{\boldsymbol{m}}_i - \boldsymbol{m}_i'\|_2^2 \right\}
\end{aligned}
$$

$$+ C|\hat{\tau}| \left\{ |\Gamma_{ww}| + |\boldsymbol{\Gamma}_w^\top J \boldsymbol{W}_i| + |\Gamma_\theta \boldsymbol{W}_i^\top J \boldsymbol{W}_i| + |\boldsymbol{\Gamma}_w^\top J \boldsymbol{W}_i| \right\},$$

for some constant C that only depends on c_π and T. By Lemma A.2 and Markov's inequality,

$$|\Gamma_{wy}| + \|\boldsymbol{\Gamma}_y\|_2 + |\Gamma_\theta| + \|\boldsymbol{\Gamma}_w\|_2 = O_{\mathbb{P}}(1), \quad |\Gamma_{ww}| + |\boldsymbol{\Gamma}_w^\top J \boldsymbol{W}_i| + |\Gamma_\theta \boldsymbol{W}_i^\top J \boldsymbol{W}_i| + |\boldsymbol{\Gamma}_w^\top J \boldsymbol{W}_i| = O(1).$$

By the first part of the theorem,

$$|\hat{\tau}| = o_{\mathbb{P}}(1).$$

Therefore,

$$\frac{1}{n}\sum_{i=1}^n (\hat{\mathcal{V}}_i'' - \hat{\mathcal{V}}_i)^2 = O_{\mathbb{P}}\left(\frac{1}{n}\sum_{i=1}^n \left\{ |\Theta_i' - \Theta_i|^2 \cdot (1 + \|\tilde{\boldsymbol{Y}}_i'\|_2^2) + \|\hat{\boldsymbol{m}}_i - \boldsymbol{m}_i'\|_2^2 \right\}\right) + o_{\mathbb{P}}(1). \tag{B.38}$$

By Assumption A.2 and B.1,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \mathbb{E}[|\Theta_i' - \Theta_i|^2] &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\frac{\boldsymbol{\Pi}(\boldsymbol{W}_i)^2}{\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i)^2 \boldsymbol{\pi}_i'(\boldsymbol{W}_i)^2} |\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) - \boldsymbol{\pi}_i'(\boldsymbol{W}_i)|^2\right] \\ &\le \frac{1}{c_\pi^2}\sum_{i=1}^n \mathbb{E}[(\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) - \boldsymbol{\pi}_i'(\boldsymbol{W}_i))^2] = O(n^{-r}) = o(1).\end{aligned}$$

By Assumption B.1,

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\|\hat{\boldsymbol{m}}_i - \boldsymbol{m}_i'\|_2^2\right] = O(n^{-r}) = o(1).$$

By Markov's inequality, we obtain that

$$\frac{1}{n}\sum_{i=1}^n |\Theta_i' - \Theta_i|^2 + \frac{1}{n}\sum_{i=1}^n \|\hat{\boldsymbol{m}}_i - \boldsymbol{m}_i'\|_2^2 = o_{\mathbb{P}}(1). \tag{B.39}$$

By Hölder's inequality,

$$\frac{1}{n}\sum_{i=1}^{n}|\Theta_i'-\Theta_i|^2\cdot\|\tilde{\boldsymbol{Y}}_i'\|_2^2\leq\left(\frac{1}{n}\sum_{i=1}^{n}|\Theta_i'-\Theta_i|^{2(1+2/\omega)}\right)^{\omega/(2+\omega)}\left(\frac{1}{n}\sum_{i=1}^{n}\|\tilde{\boldsymbol{Y}}_i'\|_2^{2+\omega}\right)^{2/(2+\omega)}.$$

By Markov's inequality and Assumption A.4,

$$\frac{1}{n}\sum_{i=1}^{n}\|\tilde{\boldsymbol{Y}}_i'\|_2^{2+\omega}=O_{\mathbb{P}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\|\tilde{\boldsymbol{Y}}_i'\|_2^{2+\omega}]\right)=O_{\mathbb{P}}(1).$$

By Assumption A.2,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[|\Theta_i'-\Theta_i|^{2(1+2/\omega)}\right]\\
&=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\frac{\boldsymbol{\Pi}(\boldsymbol{W}_i)^{2(1+2/\omega)}}{\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i)^{2(1+2/\omega)}\boldsymbol{\pi}_i'(\boldsymbol{W}_i)^{2(1+2/\omega)}}|\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i)-\boldsymbol{\pi}_i'(\boldsymbol{W}_i)|^{2(1+2/\omega)}\right]\\
&\leq\frac{1}{c_\pi^{4(1+2/\omega)}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i)-\boldsymbol{\pi}_i'(\boldsymbol{W}_i))^{2(1+2/\omega)}\right]\\
&\overset{(i)}{\leq}\frac{1}{c_\pi^{4(1+2/\omega)}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i)-\boldsymbol{\pi}_i'(\boldsymbol{W}_i))^2\right]\\
&=O\left(n^{-r}\right)=o(1),
\end{aligned}$$

where (i) uses the fact that $|\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i)-\boldsymbol{\pi}_i'(\boldsymbol{W}_i)|\leq 1$. Thus, by Markov's inequality,

$$\frac{1}{n}\sum_{i=1}^{n}|\Theta_i'-\Theta_i|^2\cdot\|\tilde{\boldsymbol{Y}}_i'\|_2^2=o_{\mathbb{P}}(1).\tag{B.40}$$

Putting (B.38), (B.39), and (B.40) together, we conclude that

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{\mathcal{V}}_i''-\hat{\mathcal{V}}_i)^2=o_{\mathbb{P}}(1).\tag{B.41}$$

By Jensen's inequality, (B.37), and (B.41),

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{\mathcal{V}}_i' - \hat{\mathcal{V}}_i)^2 \leq \frac{2}{n}\sum_{i=1}^{n}(\hat{\mathcal{V}}_i' - \hat{\mathcal{V}}_i'')^2 + \frac{2}{n}\sum_{i=1}^{n}(\hat{\mathcal{V}}_i'' - \hat{\mathcal{V}}_i)^2 = o_{\mathbb{P}}(1). \tag{B.42}$$

By Lemma A.2, it is easy to see that

$$\left|\frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i\right|^2 \leq \frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i^2 = O_{\mathbb{P}}(1). \tag{B.43}$$

As a result,

$$\left|\frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i' - \frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i\right| \leq \frac{1}{n}\sum_{i=1}^{n}|\hat{\mathcal{V}}_i' - \hat{\mathcal{V}}_i| \leq \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{\mathcal{V}}_i' - \hat{\mathcal{V}}_i)^2} = o_{\mathbb{P}}(1).$$

Together with (B.43), it implies that

$$\left|\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i'\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i\right)^2\right| = o_{\mathbb{P}}(1).$$

On the other hand, by triangle inequality, Cauchy-Schwarz inequality, and (B.43),

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i'^2 - \frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i^2\right| \leq \frac{2}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i(\hat{\mathcal{V}}_i' - \hat{\mathcal{V}}_i) + \frac{1}{n}\sum_{i=1}^{n}(\hat{\mathcal{V}}_i' - \hat{\mathcal{V}}_i)^2 \\
&\leq 2\sqrt{\frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{\mathcal{V}}_i' - \hat{\mathcal{V}}_i)^2} + \frac{1}{n}\sum_{i=1}^{n}(\hat{\mathcal{V}}_i' - \hat{\mathcal{V}}_i)^2 = o_{\mathbb{P}}(1).
\end{aligned}$$

Therefore,

$$|\hat{\sigma}^2 - \hat{\sigma}'^2| \leq \left|\frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i^2 - \frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i'^2\right| + \left|\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{V}}_i'\right)^2\right| = o_{\mathbb{P}}(1).$$

C Solutions of the DATE equation

C.1 The case of two periods

When there are two periods, the DATE equation only involves four variables $\mathbf{\Pi}(0,0), \mathbf{\Pi}(0,1), \mathbf{\Pi}(1,0), \mathbf{\Pi}(1,1)$. Through some tedious algebra presented in Appendix C.4.1, we can show that the DATE equation can be simplified into the following equation:

$$\{\mathbf{\Pi}(1,1)-\mathbf{\Pi}(0,0)\}\{\mathbf{\Pi}(1,0)-\mathbf{\Pi}(0,1)\} = (\xi_1-\xi_2)\left\{(\mathbf{\Pi}(1,0)-\mathbf{\Pi}(0,1))^2 - (\mathbf{\Pi}(1,0)+\mathbf{\Pi}(0,1))\right\}. \tag{C.1}$$

C.1.1 Difference-in-difference designs

In the setting of difference-in-difference (DiD), $(0,0)$ and $(0,1)$ are the only two possible treatment assignments. As a result, we should set the support of the reshaped distribution to be $\mathbb{S}^* = \{(0,0),(0,1)\}$. Then (C.1) reduces to

$$\mathbf{\Pi}(0,0)\mathbf{\Pi}(0,1) = (\xi_1-\xi_2)(\mathbf{\Pi}(0,1)^2 - \mathbf{\Pi}(0,1)) = (\xi_2-\xi_1)\mathbf{\Pi}(0,0)\mathbf{\Pi}(0,1).$$

It has a solution only when $\xi_2 - \xi_1 = 1$, i.e. $(\xi_1, \xi_2) = (0,1)$ and hence $\tau^*(\xi) = \tau_2$, in which case any reshaped distribution $\mathbf{\Pi}$ with $\mathbf{\Pi}(0,0), \mathbf{\Pi}(0,1) > 0$ is a solution. This is not surprising because for DiD, no unit is treated in the first period and thus τ_1 is unidentifiable. Nonetheless, τ_2 is an informative causal estimand in the literature of DiD. This implies that the RIPW estimator with any $\mathbf{\Pi}$ with $\mathbf{\Pi}(0,0), \mathbf{\Pi}(0,1) > 0$ and $\mathbf{\Pi}(0,0)+\mathbf{\Pi}(0,1) = 1$ yields a doubly robust DiD estimator.

C.1.2 Cross-over designs

For a two-period cross-over design, $(0,1)$ and $(1,0)$ are the only two possible treatment assignments. Since the support of $\mathbf{\Pi}$ must contain at least two elements, it has to be $\mathbb{S}^* = \{(1,0),(0,1)\}$. Then DATE equation reduces to

$$0 = (\xi_1-\xi_2)\left\{(\mathbf{\Pi}(1,0)-\mathbf{\Pi}(0,1))^2 - (\mathbf{\Pi}(1,0)+\mathbf{\Pi}(0,1))\right\}.$$

When $\xi_1 \neq \xi_2$, it implies that

$$0 = (\mathbf{\Pi}(1,0) - \mathbf{\Pi}(0,1))^2 - (\mathbf{\Pi}(1,0) + \mathbf{\Pi}(0,1)) = (\mathbf{\Pi}(1,0) - \mathbf{\Pi}(0,1))^2 - 1.$$

It never holds since $\mathbf{\Pi}(1,0), \mathbf{\Pi}(0,1) > 0$. By contrast, when $\xi_1 = \xi_2 = 1/2$, any $\mathbf{\Pi}$ with support $(1,0)$ and $(0,1)$ is a solution.

C.1.3 Estimating equally-weighted DATE for general designs

When $\xi_1 = \xi_2 = 1/2$, the DATE equation reduces to

$$\{\mathbf{\Pi}(1,1) - \mathbf{\Pi}(0,0)\}\{\mathbf{\Pi}(1,0) - \mathbf{\Pi}(0,1)\} = 0 \Longleftrightarrow \mathbf{\Pi}(1,1) = \mathbf{\Pi}(0,0) \text{ or } \mathbf{\Pi}(1,0) = \mathbf{\Pi}(0,1).$$

If $\mathbb{S}^* = \{(1,1),(0,0),(1,0),(0,1)\}$ in Assumption 3.2, that is, when all combinations of treatments are possible, the solutions are

$$\begin{aligned} &(\mathbf{\Pi}(1,1), \mathbf{\Pi}(0,0), \mathbf{\Pi}(0,1), \mathbf{\Pi}(1,0)) = (a, a, b, 1-2a-b), \quad a > 0, 2a+b < 1 \\ \text{or } &(\mathbf{\Pi}(1,1), \mathbf{\Pi}(0,0), \mathbf{\Pi}(0,1), \mathbf{\Pi}(1,0)) = (a, 1-a-2b, b, b), \quad b > 0, a+2b < 1. \end{aligned}$$

The uniform distribution on $\mathbb{S}^*$ is a solution, implying that the IPW weights deliver the average effect in this case. If $\mathbb{S}^* = \{(1,1),(0,0),(0,1)\}$ (staggered adoption), we cannot make $\mathbf{\Pi}(1,0)$ and $\mathbf{\Pi}(0,1)$ equal since the former must be zero while the latter must be positive. Therefore, the solutions can be characterized as

$$(\mathbf{\Pi}(1,1), \mathbf{\Pi}(0,0), \mathbf{\Pi}(0,1)) = (a, a, 1-2a), \quad a \in (0, 1/2). \tag{C.2}$$

Again, the uniform distribution on $\mathbb{S}^*$ is a solution. However, we will show in the next section that the uniform distribution is not a solution for staggered adoption designs with $T \geq 3$.

C.2 Staggered adoption with multiple periods

For staggered adoption designs, $\boldsymbol{\pi}_i$ is supported on

$$\mathcal{W}_T^{\text{sta}} \triangleq \{\boldsymbol{w} : \boldsymbol{w}_1 = \ldots = \boldsymbol{w}_i = 0, \boldsymbol{w}_{i+1} = \ldots = \boldsymbol{w}_T = 1 \text{ for some } i = 0, 1, \ldots, T\}.$$

For notational convenience, we denote by $\boldsymbol{w}_{(j)}$ the vector in $\mathcal{W}_T^{\text{sta}}$ with j entries equal to 1 for $j = 0, 1, \ldots, T$. Thus, the support $\mathbb{S}^*$ of $\boldsymbol{\Pi}$ must be a subset of $\mathcal{W}_T^{\text{sta}}$. For general weights, the DATE equation is a quadratic system with complicated structures. Nonetheless, when $\xi_1 = \ldots = \xi_T = 1/T$, the solution set is an union of segments on the T-dimensional simplex with closed-form expressions. We focus on the equally-weighted DATE in this section.

Theorem C.1. *Let $\mathbb{S}^* = \{\boldsymbol{w}_{(0)}, \boldsymbol{w}_{(j_1)}, \ldots, \boldsymbol{w}_{(j_r)}, \boldsymbol{w}_{(T)}\}$ with $1 \le j_1 < \ldots < j_r \le T-1$. Then the set of solutions of the DATE equation with support $\mathbb{S}^*$ is characterized by the following linear system:*

$$\begin{cases} \boldsymbol{\Pi}(\boldsymbol{w}_{(T)}) = \frac{T-j_r}{T} - \boldsymbol{\Pi}(\boldsymbol{w}_{(j_r)}) + \frac{1}{T}\sum_{k=1}^{r} j_k \boldsymbol{\Pi}(\boldsymbol{w}_{(j_k)}) \\ \boldsymbol{\Pi}(\boldsymbol{w}_{(j_{k+1})}) + \boldsymbol{\Pi}(\boldsymbol{w}_{(j_k)}) = \frac{j_{k+1}-j_k}{T}, \quad k = 1, \ldots, r-1 \\ \boldsymbol{\Pi}(\boldsymbol{w}_{(0)}) = 1 - \boldsymbol{\Pi}(\boldsymbol{w}_{(T)}) - \sum_{k=1}^{r} \boldsymbol{\Pi}(\boldsymbol{w}_{(j_k)}) \\ \boldsymbol{\Pi}(\boldsymbol{w}) > 0 \textit{ iff } \boldsymbol{w} \in \mathbb{S}^* \end{cases} \tag{C.3}$$

Furthermore, the solution set of (C.3) is either an empty set or a 1-dimensional segment in the form of $\{\lambda\boldsymbol{\Pi}^{(1)} + (1-\lambda)\boldsymbol{\Pi}^{(2)} : \lambda \in (0,1)\}$ for some distributions $\boldsymbol{\Pi}^{(1)}$ and $\boldsymbol{\Pi}^{(2)}$.

The proof of Theorem C.1 is presented in Appendix C.4.2. In the following corollary, we show that the solution set with $\mathbb{S}^* = \mathcal{W}_T^{\text{sta}}$ is always non-empty with nice explicit expressions.

Corollary C.1. *When $\mathbb{S}^* = \mathcal{W}_T^{\text{sta}}$, the solution set of (C.3) is $\{\lambda\boldsymbol{\Pi}^{(1)} + (1-\lambda)\boldsymbol{\Pi}^{(2)} : \lambda \in (0,1)\}$ where*

- *if T is odd,*

$$\boldsymbol{\Pi}^{(1)}(\boldsymbol{w}_{(T)}) = \frac{(T+1)^2}{4T^2}, \;\; \boldsymbol{\Pi}^{(1)}(\boldsymbol{w}_{(0)}) = \frac{T^2-1}{4T^2}, \;\; \boldsymbol{\Pi}^{(1)}(\boldsymbol{w}_j) = \frac{I(j \textit{ is odd})}{T}, \; j = 1, \ldots, T-1,$$

$$\textit{and } \; \boldsymbol{\Pi}^{(2)}(\boldsymbol{w}_{(j)}) = \boldsymbol{\Pi}^{(1)}(\boldsymbol{w}_{(T-j)}), \; j = 0, \ldots, T;$$

- *if T is even,*

$$\mathbf{\Pi}^{(1)}(\boldsymbol{w}_{(T)}) = \mathbf{\Pi}^{(1)}(\boldsymbol{w}_{(0)}) = \frac{1}{4}, \ \ \mathbf{\Pi}^{(1)}(\boldsymbol{w}_j) = \frac{I(j \ \textit{is odd})}{T}, \ j = 1, \ldots, T-1,$$

$$\textit{and} \ \ \mathbf{\Pi}^{(2)}(\boldsymbol{w}_{(T)}) = \mathbf{\Pi}^{(2)}(\boldsymbol{w}_{(0)}) = \frac{T+2}{4T}, \ \ \mathbf{\Pi}^{(2)}(\boldsymbol{w}_j) = \frac{I(j \ \textit{is even})}{T}, \ j = 1, \ldots, T-1.$$

In particular, when $T = 3$ and $\mathbb{S}^* = \mathcal{W}_T^{\mathrm{sta}}$, the solution set is

$$\left\{(\mathbf{\Pi}(\boldsymbol{w}_{(0)}), \mathbf{\Pi}(\boldsymbol{w}_{(1)}), \mathbf{\Pi}(\boldsymbol{w}_{(2)}), \mathbf{\Pi}(\boldsymbol{w}_{(3)}) = \lambda\left(\frac{2}{9}, \frac{1}{3}, 0, \frac{4}{9}\right) + (1-\lambda)\left(\frac{4}{9}, 0, \frac{1}{3}, \frac{2}{9}\right) : \lambda \in (0,1)\right\}. \tag{C.4}$$

Clearly, the uniform distribution on $\mathbb{S}^*$ is excluded. Thus, although the RIPW estimator with a uniform reshaped distribution is inconsistent, the non-uniform distribution $(1/3, 1/6, 1/6, 1/3)$, namely the midpoint of the solution set, induces a consistent RIPW estimator. For general T, it is easy to see that the midpoint is

$$\mathbf{\Pi}(\boldsymbol{w}_{(T)}) = \mathbf{\Pi}(\boldsymbol{w}_{(0)}) = \frac{T+1}{4T}, \ \ \mathbf{\Pi}(\boldsymbol{w}_{(j)}) = \frac{1}{2T}, \ j = 1, \ldots, T-1. \tag{C.5}$$

This distribution uniformly assigns probabilities on the subset $\{\boldsymbol{w}_{(1)}, \ldots, \boldsymbol{w}_{(T-1)}\}$ while puts a large mass on $\{\boldsymbol{w}_{(0)}, \boldsymbol{w}_{(T)}\}$. Intuitively, the asymmetry is driven by the special roles of $\boldsymbol{w}_{(0)}$ and $\boldsymbol{w}_{(T)}$: the former provides the only control group for period T while the latter provides the only treated group for period 1.

Corollary C.1 offers a unified recipe for the reshaped distribution when the positivity Assumption 3.2 holds for all possible assignments. In some applications, certain assignment never or rarely occurs and we are forced to restrict the support of $\mathbf{\Pi}$ into a smaller subset $\mathbb{S}^*$. To start with, we provide a detailed account of the case $T = 3$. When $j_1 = 1, j_2 = 2$, (C.4) shows that $\mathbf{\Pi}(\boldsymbol{w}_{(0)}), \mathbf{\Pi}(\boldsymbol{w}_{(3)}) > 0$, and thus $\mathbb{S}^*$ must be $\mathcal{W}_3^{\mathrm{sta}}$ and cannot be $\{\boldsymbol{w}_{(1)}, \boldsymbol{w}_{(2)}\}$, $\{\boldsymbol{w}_{(0)}, \boldsymbol{w}_{(1)}, \boldsymbol{w}_{(2)}\}$, or $\{\boldsymbol{w}_{(1)}, \boldsymbol{w}_{(2)}, \boldsymbol{w}_{(3)}\}$. When $j_1 = 1, r = 1$, via some tedious algebra, the solution set of (C.3) is

$$\left\{(\mathbf{\Pi}(\boldsymbol{w}_{(0)}), \mathbf{\Pi}(\boldsymbol{w}_{(1)}), \mathbf{\Pi}(\boldsymbol{w}_{(2)}), \mathbf{\Pi}(\boldsymbol{w}_{(3)}) = \lambda\,(0,1,0,0) + (1-\lambda)\left(\frac{1}{3}, 0, 0, \frac{2}{3}\right) : \lambda \in (0,1)\right\}. \tag{C.6}$$

Thus, $\{\boldsymbol{w}_{(0)}, \boldsymbol{w}_{(1)}, \boldsymbol{w}_{(3)}\}$ is the only support with $j_1 = 1, r = 1$ that induces a non-empty solution set of (C.3). Similarly, we can show that the only support with $j_2 = 1, r = 1$ that induces a non-empty solution set as

$$\left\{(\mathbf{\Pi}(\boldsymbol{w}_{(0)}), \mathbf{\Pi}(\boldsymbol{w}_{(1)}), \mathbf{\Pi}(\boldsymbol{w}_{(2)}), \mathbf{\Pi}(\boldsymbol{w}_{(3)}) = \lambda\,(0,0,1,0) + (1-\lambda)\left(\frac{2}{3}, 0, 0, \frac{1}{3}\right) : \lambda \in (0,1)\right\}. \quad \text{(C.7)}$$

In sum, $\mathcal{W}_T^{\text{sta}}, \mathcal{W}_T^{\text{sta}} \setminus \{\boldsymbol{w}_{(1)}\}, \mathcal{W}_T^{\text{sta}} \setminus \{\boldsymbol{w}_{(2)}\}$ are the only three supports with non-empty solution sets, characterized by (C.4), (C.6), and (C.7), respectively.

For $T = 3$, $\{j_1, \ldots, j_r\}$ can be any non-empty subset of $\{1, 2\}$. Via some tedious algebra, we can show that this continues to be true for $T = 4$. However, this no longer holds for $T \geq 5$. For instance, if $\{j_1, \ldots, j_r\} = \{1, 2, 4, 5\}$, the second equation of (C.3) implies that

$$\mathbf{\Pi}(\boldsymbol{w}_{(1)}) + \mathbf{\Pi}(\boldsymbol{w}_{(2)}) = \mathbf{\Pi}(\boldsymbol{w}_{(4)}) + \mathbf{\Pi}(\boldsymbol{w}_{(5)}) = \frac{1}{T}, \quad \mathbf{\Pi}(\boldsymbol{w}_{(2)}) + \mathbf{\Pi}(\boldsymbol{w}_{(4)}) = \frac{2}{T}.$$

Under the support constraint, the first two equations imply that $\mathbf{\Pi}(\boldsymbol{w}_{(2)}), \mathbf{\Pi}(\boldsymbol{w}_{(4)}) < 1/T$, contradicting with the third equation. Nonetheless, the contradiction can be resolved if any of these four elements is discarded. If this is the case in practice, we can discard the element that is believed to be the least likely assignment.

C.3 Other designs

In many applications, the treatment can be switched on and off at different periods for a single unit. In general, a design is characterized by a collection of possible assignments $\mathbb{S}_{\text{design}}$. If any subset $\mathbb{S}^* \subset \mathbb{S}_{\text{design}}$ yields a non-empty solution set of the DATE equation, we can derive a doubly robust estimator of the DATE. In this section, we consider several designs with more than two periods which are not staggered adoption designs.

First we consider transient designs with zero or one period being treated and with each period being treated with a non-zero chance, i.e.,

$$\mathcal{W}_{T,1}^{\text{tra}} = \left\{\boldsymbol{w} \in \{0,1\}^T : \sum_{t=1}^{T} \boldsymbol{w}_t \leq 1\right\}.$$

For notational convenience, we denote by $\tilde{\boldsymbol{w}}_{(0)}$ the never-treated assignment and $\tilde{\boldsymbol{w}}_{(j)}$ the assignment with only j-th period treated. The above design can be encountered, for example, when the treatment is a natural disaster. The following theorem characterizes all solutions of the DATE equation for any ξ.

Theorem C.2. *When $\mathbb{S}^* = \mathcal{W}^{\text{tra}}_{T,1}$, $\mathbf{\Pi}$ is a solution of the DATE equation iff there exists $b > 0$ such that*

$$\mathbf{\Pi}(\tilde{\boldsymbol{w}}_{(t)}) \left\{ 1 - \mathbf{\Pi}(\tilde{\boldsymbol{w}}_{(t)}) - \frac{\mathbf{\Pi}(\tilde{\boldsymbol{w}}_{(0)})}{T} \right\} = \xi_t b, \quad \forall t \in [T].$$

In particular, when $\xi_t = 1/T$ for every t, Theorem C.2 implies that $\mathbf{\Pi} \sim \text{Unif}(\mathcal{W}^{\text{tra}}_{T,1})$ is a solution. In fact, for any given $\mathbf{\Pi}(\tilde{\boldsymbol{w}}_0) \in (0, 1)$, $\mathbf{\Pi}$ is a solution if

$$\mathbf{\Pi}(\cdot \mid \boldsymbol{W} \neq \tilde{\boldsymbol{w}}_{(0)}) \sim \text{Unif}(\{\tilde{\boldsymbol{w}}_{(1)}, \dots, \tilde{\boldsymbol{w}}_{(T)}\}),$$

where $\boldsymbol{W}$ denotes a generic random vector drawn from $\mathbf{\Pi}$. The above decomposition can be used to construct solutions for more general transient designs:

$$\mathcal{W}^{\text{tra}}_{T,k} = \left\{ \boldsymbol{w} \in \{0, 1\}^T : \sum_{t=1}^{T} \boldsymbol{w}_t \leq k \right\}.$$

This design is common in marketing experiments where, for example, k is the maximal number of coupons given to a user and each user can receive coupons in any combination of up to k time periods.

Theorem C.3. *When $\mathbb{S}^* = \mathcal{W}^{\text{tra}}_{T}$, $\mathbf{\Pi}$ is a solution of the DATE equation with $\xi_t = 1/T$ ($t = 1, \dots, T$), if*

$$\mathbf{\Pi}\left(\cdot \mid \sum_{t=1}^{T} \boldsymbol{W}_t = k'\right) \sim \text{Unif}(\mathcal{W}^{\text{tra}}_{T,k'} \setminus \mathcal{W}^{\text{tra}}_{T,k'-1}), \quad k' = 1, \dots, k,$$

C.4 Proofs

For notational convenience, denote by $h(\mathbf{\Pi}) = (h_1(\mathbf{\Pi}), \dots, h_T(\mathbf{\Pi}))$ the left-hand side of the DATE equation. We start by a simple but useful observation that, for any $\mathbf{\Pi}$,

$$\mathbf{1}_T^\top h(\mathbf{\Pi}) = \mathbb{E}_{\boldsymbol{W} \sim \mathbf{\Pi}} \left[(\mathbf{1}_T^\top \operatorname{diag}(\boldsymbol{W}) - \mathbf{1}_T^\top \xi \boldsymbol{W}^\top) J(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W} \sim \mathbf{\Pi}}[\boldsymbol{W}]) \right]$$

$$= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[(\boldsymbol{W}^\top - \boldsymbol{W}^\top)J(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])\right] = 0. \tag{C.8}$$

Thus, there is at least one redundant equation and for any matrix $V \in \mathbb{R}^{T\times(T-1)}$ with $V^\top \mathbf{1}_T = 0$,

$$h(\boldsymbol{\Pi}) = 0 \Longleftrightarrow V^\top h(\boldsymbol{\Pi}) = 0. \tag{C.9}$$

C.4.1 Proof of equation (C.1)

Set $V = (1, -1)^\top$ in (C.9). Then

$$V^\top h(\boldsymbol{\Pi}) = 0 \Longleftrightarrow h_1(\boldsymbol{\Pi}) - h_2(\boldsymbol{\Pi}) = 0.$$

As a result,

$$\begin{aligned}
0 &= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[((W_1, -W_2) - (\xi_1 - \xi_2)(W_1, W_2))\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} W_1 - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_1] \\ W_2 - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_2] \end{bmatrix}\right] \\
&= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[(W_1 + W_2 - (\xi_1 - \xi_2)(W_1 - W_2))(W_1 - W_2 - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}(W_1 - W_2))] \\
&= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_1^2 - W_2^2 - (\xi_1 - \xi_2)(W_1 - W_2)^2] \\
&\quad - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_1 + W_2 - (\xi_1 - \xi_2)(W_1 - W_2)]\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}(W_1 - W_2) \\
&= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_1 - W_2 - (\xi_1 - \xi_2)(W_1 - W_2)^2] \\
&\quad - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_1 + W_2 - (\xi_1 - \xi_2)(W_1 - W_2)]\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}(W_1 - W_2) \\
&= (\boldsymbol{\Pi}(1,0) - \boldsymbol{\Pi}(0,1)) - (\xi_1 - \xi_2)(\boldsymbol{\Pi}(1,0) + \boldsymbol{\Pi}(0,1)) \\
&\quad - \{\boldsymbol{\Pi}(1,0) + \boldsymbol{\Pi}(0,1) + 2\boldsymbol{\Pi}(1,1) - (\xi_1 - \xi_2)(\boldsymbol{\Pi}(1,0) - \boldsymbol{\Pi}(0,1))\}\{\boldsymbol{\Pi}(1,0) - \boldsymbol{\Pi}(0,1)\} \\
&= (\boldsymbol{\Pi}(1,0) - \boldsymbol{\Pi}(0,1)) - (\xi_1 - \xi_2)(\boldsymbol{\Pi}(1,0) + \boldsymbol{\Pi}(0,1)) \\
&\quad - \{1 + \boldsymbol{\Pi}(1,1) - \boldsymbol{\Pi}(0,0) - (\xi_1 - \xi_2)(\boldsymbol{\Pi}(1,0) - \boldsymbol{\Pi}(0,1))\}\{\boldsymbol{\Pi}(1,0) - \boldsymbol{\Pi}(0,1)\}.
\end{aligned}$$

Rearranging the terms yields

$$\{\boldsymbol{\Pi}(1,1) - \boldsymbol{\Pi}(0,0)\}\{\boldsymbol{\Pi}(1,0) - \boldsymbol{\Pi}(0,1)\} = (\xi_1 - \xi_2)\left\{(\boldsymbol{\Pi}(1,0) - \boldsymbol{\Pi}(0,1))^2 - (\boldsymbol{\Pi}(1,0) + \boldsymbol{\Pi}(0,1))\right\}.$$

$$\tag{C.10}$$

C.4.2 Proof of Theorem C.1

Let $\boldsymbol{e}_j$ denote the j-th canonical basis in $\mathbb{R}^T$. Then

$$h_j(\boldsymbol{\Pi}) = \boldsymbol{e}_j^\top \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[(\mathrm{diag}(\boldsymbol{W}) - \xi \boldsymbol{W}^\top) J(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])\right].$$

We can decompose $h_j(\boldsymbol{\Pi})$ into $h_{j1}(\boldsymbol{\Pi}) - \xi_j h_2(\boldsymbol{\Pi})$ where

$$h_{j1}(\boldsymbol{\Pi}) = \boldsymbol{e}_j^\top \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[\mathrm{diag}(\boldsymbol{W}) J(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])\right], \quad h_2(\boldsymbol{\Pi}) = \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[\boldsymbol{W}^\top J(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])\right].$$

Then

$$\begin{aligned}
h_{j1}(\boldsymbol{\Pi}) &= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[W_j \boldsymbol{e}_j^\top J(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])\right] \\
&= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[W_j \boldsymbol{e}_j^\top J\boldsymbol{W}\right] - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[W_j \boldsymbol{e}_j^\top J\right] \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}] \\
&= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[W_j\left(W_j - \frac{\mathbf{1}_T^\top \boldsymbol{W}}{T}\right)\right] - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_j]\, \boldsymbol{e}_j^\top J \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}] \\
&= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[W_j\left(W_j - \frac{\mathbf{1}_T^\top \boldsymbol{W}}{T}\right)\right] - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_j]\, \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[W_j - \frac{\mathbf{1}_T^\top \boldsymbol{W}}{T}\right] \\
&= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_j] - (\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_j])^2 + \frac{\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_j]\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\mathbf{1}_T^\top \boldsymbol{W}]}{T} - \frac{\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_j(\mathbf{1}_T^\top \boldsymbol{W})]}{T},
\end{aligned}$$

where the last equality follows from the fact that $W_j^2 = W_j$. By (C.9), it is equivalent to find $\boldsymbol{\Pi}$ satisfying

$$\Delta h_j(\boldsymbol{\Pi}) = h_{j+1}(\boldsymbol{\Pi}) - h_j(\boldsymbol{\Pi}) = 0, \quad j = 1, 2, \ldots, T-1.$$

In this case, $\xi_{j+1} = \xi_j$ for any j, and thus,

$$h_{(j+1)1}(\boldsymbol{\Pi}) - h_{j1}(\boldsymbol{\Pi}) = 0, \quad j = 1, 2, \ldots, T-1. \tag{C.11}$$

By definition,

$$W_{j+1} - W_j = I(\boldsymbol{W} = \boldsymbol{w}_{(T-j)}). \tag{C.12}$$

As a consequence, we have

$$\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_{j+1}] - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_j] = \boldsymbol{\Pi}(\boldsymbol{w}_{(T-j)}),$$

$$(\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_{j+1}])^2 - (\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_j])^2 = \boldsymbol{\Pi}(\boldsymbol{w}_{(T-j)})^2 + 2\boldsymbol{\Pi}(\boldsymbol{w}_{(T-j)})\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_j],$$

and

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_{j+1}(\mathbf{1}_T^\top \boldsymbol{W})] - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_j(\mathbf{1}_T^\top \boldsymbol{W})] \\
&= \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[I(\boldsymbol{W} = \boldsymbol{w}_{(T-j)})(\mathbf{1}_T^\top \boldsymbol{w}_{(T-j)})] = (T-j)\boldsymbol{\Pi}(\boldsymbol{w}_{(T-j)}).
\end{aligned}$$

As a result,

$$\begin{aligned}
&h_{(j+1)1}(\boldsymbol{\Pi}) - h_{j1}(\boldsymbol{\Pi}) \\
&= \boldsymbol{\Pi}(\boldsymbol{w}_{(T-j)}) \left\{ 1 - \boldsymbol{\Pi}(\boldsymbol{w}_{(T-j)}) - 2\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_j] + \frac{\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\mathbf{1}_T^\top \boldsymbol{W}]}{T} - \frac{T-j}{T} \right\} \\
&= \boldsymbol{\Pi}(\boldsymbol{w}_{(T-j)}) \left\{ \frac{j}{T} - \boldsymbol{\Pi}(\boldsymbol{w}_{(T-j)}) - 2\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_j] + \frac{\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\mathbf{1}_T^\top \boldsymbol{W}]}{T} \right\}.
\end{aligned} \tag{C.13}$$

Let

$$g_j(\boldsymbol{\Pi}) = \frac{T-j}{T} - \boldsymbol{\Pi}(\boldsymbol{w}_{(j)}) - 2\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_{T-j}] + \frac{\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\mathbf{1}_T^\top \boldsymbol{W}]}{T}. \tag{C.14}$$

Thus, (C.11) can be reformulated as

$$\boldsymbol{\Pi}(\boldsymbol{w}_{(j)}) = 0 \ \text{ or } \ g_j(\boldsymbol{\Pi}) = 0, \quad j = 1, 2, \ldots, T-1. \tag{C.15}$$

Since $\mathbb{S}^* = \{\boldsymbol{w}_{(0)}, \boldsymbol{w}_{(j_1)}, \ldots, \boldsymbol{w}_{(j_r)}, \boldsymbol{w}_{(T)}\}$, $\boldsymbol{\Pi}(\boldsymbol{w}_{(j_k)}) > 0$ for each $k = 1, \ldots, r$. As a result, (C.15) is equivalent to

$$g_{j_r}(\boldsymbol{\Pi}) = 0, \quad g_{j_k}(\boldsymbol{\Pi}) - g_{j_{k+1}}(\boldsymbol{\Pi}) = 0, \quad k = 1, \ldots, r-1. \tag{C.16}$$

Note that

$$W_{T-j_k} = 1 \Longleftrightarrow \boldsymbol{W} \in \{\boldsymbol{w}_{(j_k+1)}, \ldots, \boldsymbol{w}_{(T)}\}.$$

The first equation is equivalent to

$$\begin{aligned} &\frac{T-j_r}{T} - \boldsymbol{\Pi}(\boldsymbol{w}_{(j_r)}) - 2\boldsymbol{\Pi}(\boldsymbol{w}_{(T)}) + \frac{1}{T}\left(\sum_{k=1}^{r} j_k \boldsymbol{\Pi}(\boldsymbol{w}_{(j_k)}) + T\boldsymbol{\Pi}(\boldsymbol{w}_{(T)})\right) = 0 \\ &\Longleftrightarrow \boldsymbol{\Pi}(\boldsymbol{w}_{(T)}) = \frac{T-j_r}{T} - \boldsymbol{\Pi}(\boldsymbol{w}_{(j_r)}) + \frac{1}{T}\sum_{k=1}^{r} j_k \boldsymbol{\Pi}(\boldsymbol{w}_{(j_k)}). \end{aligned} \tag{C.17}$$

By (C.12),

$$\begin{aligned} \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_{T-j_k}] - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[W_{T-j_{k+1}}] &= \mathbb{P}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left(\boldsymbol{W} \in \{\boldsymbol{w}_{(j_k+1)}, \boldsymbol{w}_{(j_k+2)}, \ldots, \boldsymbol{w}_{(j_{k+1})}\}\right) \\ &= \mathbb{P}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left(\boldsymbol{W} = \boldsymbol{w}_{(j_{k+1})}\right) = \boldsymbol{\Pi}(\boldsymbol{w}_{(j_{k+1})}). \end{aligned}$$

Therefore, the second equation of (C.15) can be simplified to

$$\boldsymbol{\Pi}(\boldsymbol{w}_{(j_{k+1})}) + \boldsymbol{\Pi}(\boldsymbol{w}_{(j_k)}) = \frac{j_{k+1} - j_k}{T}, \quad k = 1, \ldots, r-1. \tag{C.18}$$

Finally the simplex constraint determines $\boldsymbol{\Pi}(\tilde{\boldsymbol{w}}_{(0)})$ as

$$\boldsymbol{\Pi}(\boldsymbol{w}_{(0)}) = 1 - \boldsymbol{\Pi}(\boldsymbol{w}_{(T)}) - \sum_{k=1}^{r} \boldsymbol{\Pi}(\boldsymbol{w}_{(j_k)}). \tag{C.19}$$

Clearly, $\boldsymbol{\Pi}(\boldsymbol{w}_{(j_1)})$ determines all other $\boldsymbol{\Pi}(\boldsymbol{W}_{(j_k)})$'s. Therefore, the solution set of (C.17) - (C.19) is a one-dimensional linear subspace. The solution set of the DATE equation is empty if it has no intersection with the set $\{\boldsymbol{\Pi} : \boldsymbol{\Pi}(\boldsymbol{w}_{(j_k)}) > 0, r = 1, \ldots, r\}$; otherwise, it must be a segment which can be characterized as $\{\lambda\boldsymbol{\Pi}^{(1)} + (1-\lambda)\boldsymbol{\Pi}^{(2)} : \lambda \in (0,1)\}$.

C.4.3 Proof of Theorem C.2

Let $\boldsymbol{\eta} = (\boldsymbol{\Pi}(\tilde{\boldsymbol{w}}_{(1)}), \ldots, \boldsymbol{\Pi}(\tilde{\boldsymbol{w}}_{(T)})) \in \mathbb{R}^T$. Then the DATE equation can be equivalently formulated as

$$\sum_{j=1}^{T} \left(\operatorname{diag}(\tilde{\boldsymbol{w}}_{(j)}) - \xi \tilde{\boldsymbol{w}}_{(j)}^{\top}\right) J(\tilde{\boldsymbol{w}}_{(j)} - \boldsymbol{\eta})\eta_j = 0.$$

Since $\tilde{\boldsymbol{w}}_{(j)} = \boldsymbol{e}_j$, $\operatorname{diag}(\tilde{\boldsymbol{w}}_{(j)}) = \boldsymbol{e}_j \boldsymbol{e}_j^{\top}$ and we can reformulate the above equation as

$$\sum_{j=1}^{T} \left(\boldsymbol{e}_j - \xi\right) \boldsymbol{e}_j^{\top} J(\boldsymbol{e}_j - \boldsymbol{\eta})\eta_j = 0 \Longleftrightarrow \sum_{j=1}^{T} f_j(\boldsymbol{\eta})\boldsymbol{e}_j = \left\{\sum_{j=1}^{T} f_j(\boldsymbol{\eta})\right\}\xi.$$

where $f_j(\boldsymbol{\eta}) = \boldsymbol{e}_j^{\top} J(\boldsymbol{e}_j - \boldsymbol{\eta})\eta_j$. It can be equivalently formulated as an equation on $\boldsymbol{\eta}$ and a scalar b:

$$\sum_{j=1}^{T} f_j(\boldsymbol{\eta})\boldsymbol{e}_j = b\xi. \tag{C.20}$$

This is because for any $\boldsymbol{\eta}$ that satisfies (C.20), multiplying $\mathbf{1}_T^{\top}$ on both sides implies that

$$b = b(\xi^{\top}\mathbf{1}_T) = \sum_{j=1}^{T} f_j(\boldsymbol{\eta}).$$

Taking the j-th entry of both sides, (C.20) yields that

$$f_j(\boldsymbol{\eta}) = \xi_j b. \tag{C.21}$$

By definition,

$$f_j(\boldsymbol{\eta}) = \eta_j \left(\boldsymbol{e}_j^{\top} J \boldsymbol{e}_j - \boldsymbol{e}_j^{\top} J \boldsymbol{\eta}\right) = \eta_j \left(1 - \frac{1}{T} - \eta_j + \frac{1}{T}\sum_{j=1}^{T} \eta_j\right).$$

Since $\mathbf{\Pi}$ should be supported on $\{\tilde{\boldsymbol{w}}_{(0)}, \tilde{\boldsymbol{w}}_{(1)}, \ldots, \tilde{\boldsymbol{w}}_{(T)}\}$,

$$\sum_{j=1}^{T} \eta_j = \sum_{j=1}^{T} \mathbf{\Pi}(\tilde{\boldsymbol{w}}_{(j)}) = 1 - \mathbf{\Pi}(\tilde{\boldsymbol{w}}_{(0)}).$$

Therefore, (C.21) is equivalent to

$$\mathbf{\Pi}(\tilde{\boldsymbol{w}}_{(j)}) \left(1 - \mathbf{\Pi}(\tilde{\boldsymbol{w}}_{(j)}) - \frac{\mathbf{\Pi}(\tilde{\boldsymbol{w}}_{(0)})}{T}\right) = \xi_j b.$$

C.4.4 Proof of Theorem C.3

Let $\|\boldsymbol{w}\|_1$ be the L_1 norm of $\boldsymbol{w}$, i.e., $\|\boldsymbol{w}\|_1 = \sum_{i=1}^{n} w_i$. For given $\mathbf{\Pi}$ such that

$$\mathbf{\Pi}\left(\cdot \mid \|\boldsymbol{w}\|_1 = k'\right) \sim \mathrm{Unif}(\mathcal{W}_{T,k'}^{\mathrm{tra}} \setminus \mathcal{W}_{T,k'-1}^{\mathrm{tra}}), \quad k' = 1, \ldots, k,$$

By symmetry,

$$\mathbb{E}_{\boldsymbol{W} \sim \mathbf{\Pi}}[\boldsymbol{W} \mid \|\boldsymbol{W}\|_1] = \frac{\|\boldsymbol{W}\|_1}{T} \mathbf{1}_T.$$

By the iterated law of expectation,

$$\mathbb{E}_{\boldsymbol{W} \sim \mathbf{\Pi}}[\boldsymbol{W}] = \mathbb{E}_{\|\boldsymbol{W}\|_1}[\mathbb{E}_{\boldsymbol{W} \sim \mathbf{\Pi}}[\boldsymbol{W} \mid \|\boldsymbol{W}\|_1]] = \frac{\mathbb{E}_{\boldsymbol{W} \sim \mathbf{\Pi}}[\|\boldsymbol{W}\|_1]}{T} \mathbf{1}_T.$$

Since $J\mathbf{1}_T = 0$, the DATE equation with $\xi = \mathbf{1}_T / T$ reduces to

$$\mathbb{E}_{\boldsymbol{W} \sim \mathbf{\Pi}}\left[\left(\mathrm{diag}(\boldsymbol{W}) - \frac{\mathbf{1}_T}{T} \boldsymbol{W}^\top\right) J\boldsymbol{W}\right] = 0.$$

We will prove the following stronger claim:

$$\mathbb{E}_{\boldsymbol{W} \sim \mathbf{\Pi}}\left[\left(\mathrm{diag}(\boldsymbol{W}) - \frac{\mathbf{1}_T}{T} \boldsymbol{W}^\top\right) J\boldsymbol{W} \mid \|\boldsymbol{W}\|_1 = k'\right] = 0, \quad \forall k' = 1, \ldots, k.$$

Conditional on $\|\boldsymbol{W}\|_1 = k'$,

$$J\boldsymbol{W} = \boldsymbol{W} - \frac{k'}{T} \mathbf{1}_T, \quad \mathrm{diag}(\boldsymbol{W})\boldsymbol{W} = \boldsymbol{W}, \quad \boldsymbol{W}^\top \boldsymbol{W} = \boldsymbol{W}^\top \mathbf{1}_T = k'$$

Thus,

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[\left(\operatorname{diag}(\boldsymbol{W})-\frac{\mathbf{1}_T}{T}\boldsymbol{W}^\top\right)J\boldsymbol{W}\mid\|\boldsymbol{W}\|_1=k'\right]\\
&=\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[\left(\operatorname{diag}(\boldsymbol{W})-\frac{\mathbf{1}_T}{T}\boldsymbol{W}^\top\right)\left(\boldsymbol{W}-\frac{k'}{T}\mathbf{1}_T\right)\mid\|\boldsymbol{W}\|_1=k'\right]\\
&=\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[\boldsymbol{W}-\frac{k'}{T}\boldsymbol{W}-\frac{k'\mathbf{1}_T}{T}+\frac{k'^2\mathbf{1}_T}{T^2}\mid\|\boldsymbol{W}\|_1=k'\right]\\
&=0.
\end{aligned}
$$

C.5 A general solver via nonlinear programming

For a general design $\mathbb{S}_{\text{design}}=\{\check{\boldsymbol{w}}_{(1)},\ldots,\check{\boldsymbol{w}}_{(K)}\}$, the DATE equation can be formulated as a quadratic system. The j-th equation of DATE equation is

$$
\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}\left[(e_j\boldsymbol{W}_j-\boldsymbol{W}\xi_j)^TJ(\boldsymbol{W}-\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}])\right]=0, \tag{C.22}
$$

Let $\boldsymbol{p}=(\boldsymbol{\Pi}(\check{\boldsymbol{w}}_{(1)}),\ldots,\boldsymbol{\Pi}(\check{\boldsymbol{w}}_{(K)}))\in\mathbb{R}^T$, $A=(\check{\boldsymbol{w}}_{(1)},\ldots,\check{\boldsymbol{w}}_{(K)})\in\mathbb{R}^{T\times K}$, $B^{(j)}=(B_1^{(j)},\ldots,B_K^{(j)})\in\mathbb{R}^{T\times K}$, and $\boldsymbol{b}^{(j)}=(b_1^{(j)},\ldots,b_K^{(j)})^\top\in\mathbb{R}^K$, where

$$
B_k^{(j)}=J(\boldsymbol{e}_j\check{\boldsymbol{w}}_{(k),j}-\check{\boldsymbol{w}}_{(k)}\xi_j)\in\mathbb{R}^T,\quad b_k^{(j)}=\check{\boldsymbol{w}}_{(k)}^\top B_k^{(j)}\in\mathbb{R}.
$$

It is easy to see that $B^{(j)}=J(\boldsymbol{e}_j\boldsymbol{e}_j^\top-\xi_jI)A$ and $\boldsymbol{b}^{(j)}=\operatorname{diag}(A^\top B^{(j)})$. Then (C.22) can be reformulated as

$$
\boldsymbol{p}^\top\boldsymbol{b}^{(j)}-\boldsymbol{p}^\top(A^\top B^{(j)})\boldsymbol{p}=0.
$$

As a result, the DATE equation has a solution iff the minimal value of the following optimization problem is 0:

$$
\min\sum_{j=1}^T\{\boldsymbol{p}^\top\boldsymbol{b}^{(j)}-\boldsymbol{p}^\top(A^\top B^{(j)})\boldsymbol{p}\}^2,\quad\text{s.t., }\boldsymbol{p}^\top\mathbf{1}=1,\boldsymbol{p}\geq 0. \tag{C.23}
$$

We can optimize (C.23) via the standard BFGS algorithm, with the uniform distribution being the initial value. When the minimal value with a given initial value is bounded away from zero, we will try other randomly generated initial values to ensure a thorough search. If none of the initial values yields a zero objective, we claim that the DATE equation has no solution. Note that (C.23) is a nonconvex problem, the BFGS algorithm is not guaranteed to find the global minimum. Therefore, it should be viewed as an attempt to find a solution of the DATE equation instead of a trustable solver.

On the other hand, when the DATE equation has multiple solutions, it is unclear which solution can be found. In principle, we can add different constraints or regularizers to (C.23) in order to obtain a "well-behaved" solution. For instance, it is reasonable to find the most dispersed reshaped function to maximize the sample efficiency. For this purpose, we can find the solution that maximizes $\min_k \mathbf{\Pi}(\check{\boldsymbol{w}}_{(k)})$. This can be achieved by replacing the constraint $\boldsymbol{p} \geq 0$ in (C.23) by $\boldsymbol{p} \geq c\mathbf{1}$ and find the largest c for which the minimal value is zero by a binary search.

D Aggregated AIPW estimator is not doubly robust in the presence of fixed effects

We are not aware of other doubly robust estimators for DATE when the treatment and outcome models are defined as in our paper. In the absence of dynamic treatment effects, it is tempting to treat each period as a cross-sectional data, estimate the time-specific ATE τ_t by an aggregated AIPW estimator, and aggregate these estimates. To the best of our knowledge, this estimator has not been proposed in the literature. However, perhaps surprisingly, we show in this section that the aggregated AIPW estimator is not doubly robust because of the fixed effect terms in the outcome model.

Specifically, for time period t, the AIPW estimator for τ_t is defined as

$$\hat{\tau}_t = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(Y_{it} - \hat{\mathbb{E}}[Y_{it}(1) \mid \boldsymbol{X}_i])W_{it}}{\hat{\mathbb{P}}(W_{it} = 1 \mid \boldsymbol{X}_i)} - \frac{(Y_{it} - \hat{\mathbb{E}}[Y_{it}(0) \mid \boldsymbol{X}_i])(1 - W_{it})}{\hat{\mathbb{P}}(W_{it} = 0 \mid \boldsymbol{X}_i)} + \hat{\mathbb{E}}[Y_{it}(1) \mid \boldsymbol{X}_i] - \hat{\mathbb{E}}[Y_{it}(0) \mid \boldsymbol{X}_i]\right).$$

Then the aggregated AIPW estimator is defined as

$$\hat{\tau}_{\text{AIPW}} = \frac{1}{T}\sum_{t=1}^{T}\hat{\tau}_t.$$

It is known that $\hat{\tau}_t$ is doubly robust in the sense that $\hat{\tau}_t$ is consistent if either $\hat{\mathbb{P}}(W_{it} = 1)$ or $(\hat{\mathbb{E}}[Y_{it}(1) \mid X_i], \hat{\mathbb{E}}[Y_{it}(0) \mid X_i])$ is consistent for all i and t. Importantly, the requirement on the outcome model for the AIPW estimator is strictly stronger than that for the RIPW estimator; the former requires both m_{it} and the fixed effects to be consistently estimated while the latter only requires m_{it} to be consistent. It turns out that the extra requirement leads to tricky problems of the AIPW estimator.

To demonstrate the failure of the AIPW estimator, we only consider the case with sample size $n = 1000$ and a constant treatment effect to highlight that the failure is not driven by small samples or effect heterogeneity. In particular, we consider a standard TWFE model

$$Y_{it}(0) = \alpha_i + \lambda_t + m_{it} + \epsilon_{it}, \quad m_{it} = X_i\beta_t, \quad \tau_{it} = \tau,$$

where $\sum_{i=1}^{n}\alpha_i = \sum_{t=1}^{T}\lambda_t = 0$. The other details are the same as Section 5.1.

Both the RIPW and the aggregated AIPW estimators require estimates of the treatment and outcome models. First, we consider a wrong and a correct treatment model:

- (Wrong treatment model): set $\hat{\boldsymbol{\pi}}_i(\boldsymbol{w}) = |\{j : \boldsymbol{W}_j = \boldsymbol{w}\}|/n$, i.e., the empirical distribution of $\boldsymbol{W}_i$'s that ignores the covariate;
- (Correct treatment model): set $\hat{\boldsymbol{\pi}}_i(\boldsymbol{w}) = |\{j : \boldsymbol{W}_j = \boldsymbol{w}, X_j = X_i\}|/|\{j : X_j = X_i\}|$, i.e., the empirical distribution of $\boldsymbol{W}_i$'s stratified by the covariate.

With a large sample, $\hat{\boldsymbol{\pi}}_i$ in the second setting is a consistent estimator of $\boldsymbol{\pi}_i$. For the aggregated AIPW estimator, we use the marginal distributions of $\hat{\boldsymbol{\pi}}_i$ as the estimates of marginal propensity scores. Similarly, we consider a wrong and a correct outcome model:

- (Wrong outcome model): $\hat{m}_{it} = 0$ for every i and t;
- (Correct outcome model): run unweighted TWFE regression adjusting for interaction between X_i and time fixed effects, i.e., $X_i I(t = t')$ for each $t' = 1, \ldots, t$, and set $\hat{m}_{it} = X_i\hat{\beta}_t$.

With a large sample, the standard theory implies the consistency of $\hat{\beta}_t$, and hence $\hat{m}_{it} \approx m_{it}$. Unlike the RIPW estimator, the aggregated AIPW estimator requires the estimate of full conditional expectations of potential outcomes, instead of merely $\hat{m}_{it}$. In this case, a reasonble estimate of the outcome model can be formulated as

$$\hat{\mathbb{E}}[Y_{it}(0) \mid X_i] = \hat{\alpha}_i + \hat{\lambda}_t + X_i \hat{\beta}_t, \quad \hat{\mathbb{E}}[Y_{it}(1) \mid X_i] = \hat{\mathbb{E}}[Y_{it}(0) \mid X_i] + \hat{\tau}.$$

For short panels with $T = O(1)$, the time fixed effects λ_t's can be estimated via the standard TWFE regression, which are known to be consistent. However, there is no way to consistently estimate the unit fixed effect α_i since only T samples $Y_{i1}, \ldots, Y_{iT}$ can be used for estimation. The central question is how to estimate α_i for the aggregated AIPW estimator. Here we consider three strategies:

(1) using the plug-in estimate of α_i's, even if they are inconsistent;

(2) pretending that α_i does not exist and setting $\hat{\alpha}_i = 0$;

Note that the first strategy cannot be used with cross-fitting because it is impossible to estimate α_i without using the i-th sample.

We then consider all four combinations of outcome and treatment modelling. Figure 6 presents the boxplots of $\hat{\tau} - \tau$ for the three versions of AIPW, RIPW, and unweighted TWFE estimator.

First, we can see that all estimators are unbiased when both models are correct and biased when both models are wrong. As expected, the RIPW estimator is also unbiased when one of the model is correct, and the unweighted estimator is unbiased when the outcome model is correct. However, none of AIPW estimators are doubly robust: the AIPW estimator with estimated fixed effects is biased when the treatment model is correct, and the AIPW estimator that zeros out fixed effects with or without cross-fitting are biased when the outcome model is correct.

The bias of AIPW estimator that zeros out the fixed effects can be attributed to biased estimates of the outcome model despite including the covariates. The bias of the in-sample AIPW estimator can be attributed to the dependence between the outcome model estimates and the treatment assignment. In fact, when T is small, this dependence is nonvanishing no matter how fixed effects are estimated. On the other hand, the AIPW estimator is valid under

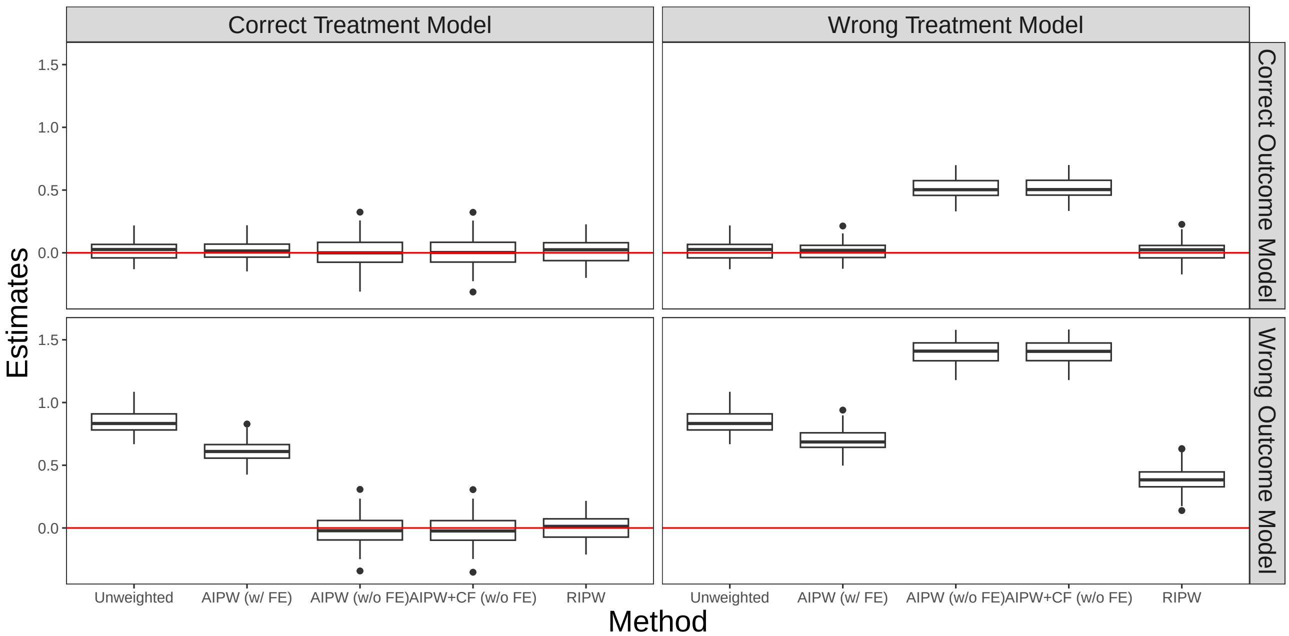

Figure 6: Boxplots of $\hat{\tau} - \tau$ for the RIPW estimator, unweighted TWFE estimator, and the three versions of AIPW: "AIPW (w/ FE)" for the one fit on the entire data with estimated fixed effects, "AIPW (w/o FE)" for the one fit on the entire data with fixed effects zeroed out, and "AIPW+CF (w/o FE)" for the cross-fitted one with fixed effects zeroed out.

a correct treatment model but a wrong outcome model only when the outcome model estimate is asymptotically independent of the assignments. In sum, there is no simple way to estimate fixed effects to make the resulting aggregated AIPW estimator doubly robust.

E More details of the OpenTable dataset

We collect the variables from different sources.

- Daily state-level year-over-year percentage change in seated diners provided by OpenTable [OpenTable] (outcome variable): https://www.opentable.com/state-of-industry.
- Indicator of whether the state of emergency has been declared [Perper et al.] (treatment variable): https://www.businessinsider.com/california-washington-state-of-emergency-coronavirus-what-it-means-2020-3.
- Daily state-level accumulated confirmed cases [Dong et al., 2020] (covariate): https://coronavirus.jhu.edu/.

Variable	Mean	SD	1st quartile	3rd quartile
Reservation Diff. (in %)	-8.89	14.94	-17.00	-1.00
Confirmed Cases	14.55	50.10	0.00	7.00
Vote Share	49.40	9.72	40.11	57.18
log(#Hospital Beds)	9.83	0.81	9.37	10.24

Table 4: Summary statistics.

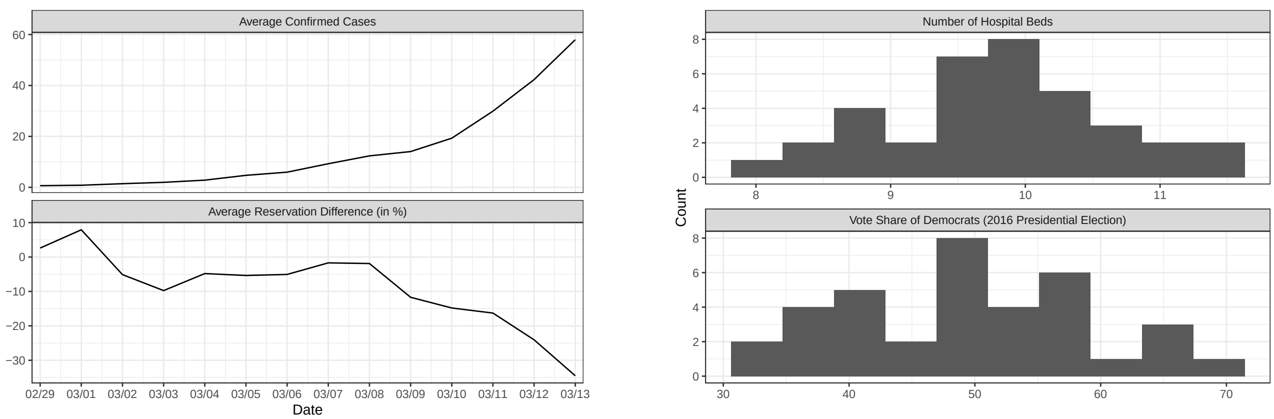

Figure 7: (Left) daily average of the reservation difference and confirmed cases. (Right) histograms of number of hospital beds and vote share.

- Vote share of Democrats based on the 2016 presidential election data [MIT Election Data and Science Lab, 2018](covariate): `https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/VOQCHQ`.

- Number of hospital beds [Zemel et al.](covariate): `https://github.com/rbracco/covidcompare`.

The summary statistics are reported in Table 4. We also plot the daily average of the reservation difference and confirmed cases as well as the histograms of the other two variables in Figure 7.