On-Line Appendix

When Should You Adjust Standard Errors for Clustering?

Alberto Abadie, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge

Current version: September 19, 2022

A.1. Setting and notation

We have a sequence of populations indexed by k. The k-th population has n_k units, indexed by $i = 1, \ldots, n_k$. The population is partitioned into m_k strata or clusters. Let $m_{k,i} \in \{1, \ldots, m_k\}$ denote the stratum that unit i of population k belongs to. The number of units in cluster m of population k is $n_{k,m} \geqslant 1$. For each unit, i, there are two potential outcomes, $y_{k,i}(1)$ and $y_{k,i}(0)$, corresponding to treatment and no treatment. The parameter of interest is the population average treatment effect

$$\tau_k = \frac{1}{n_k} \sum_{i=1}^{n_k} (y_{k,i}(1) - y_{k,i}(0)).$$

The population treatment effect by cluster is

$$\tau_{k,m} = \frac{1}{n_{k,m}} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\}(y_{k,i}(1) - y_{k,i}(0)).$$

Therefore,

$$\tau_k = \sum_{m=1}^{m_k} \frac{n_{k,m}}{n_k} \tau_{k,m}.$$

We will assume that potential outcomes, $y_{k,i}(1)$ and $y_{k,i}(0)$, are bounded in absolute value, uniformly for all (k, i).

We next describe the two components of the stochastic nature of the sample. There is a stochastic binary treatment for each unit in each population, $W_{k,i} \in \{0, 1\}$. The realized outcome for unit i in population k is $Y_{k,i} = y_{k,i}(W_{k,i})$. For a random sample of the population, we observe the triple $(Y_{k,i}, W_{k,i}, m_{k,i})$. Inclusion in the sample is represented by the random variable $R_{k,i}$, which takes value one if unit i belongs to the sample, and value zero if not.

The sampling process that determines the values of $R_{k,i}$ is independent of the potential outcomes and the assignments. It consists of two stages. First, clusters are sampled with cluster sampling probability $q_k \in (0, 1]$. Second, units are sampled from the subpopulation consisting of all the sampled clusters, with unit sampling probability equal to $p_k \in (0, 1]$. Both q_k and p_k may be equal to one, or close to zero. If $q_k = 1$, we sample all clusters. If $p_k = 1$, we sample all units from the sampled clusters. If $q_k = p_k = 1$, all units in the population are sampled.

The assignment process that determines the values of $W_{k,i}$ also consists of two stages. In the first stage, for cluster m in population k, an assignment probability $A_{k,m} \in [0, 1]$ is drawn randomly from a distribution with mean μ_k, bounded away from zero and one uniformly in k, and variance σ_k^2, independently for each cluster. The variance σ_k^2 is key. If it is zero, we have random assignment across clusters. For positive values of σ_k^2 we have correlated assignment within the clusters. Because $A_{k,m}^2 \leqslant A_{k,m}$, it follows that σ_k^2 is bounded above by $\mu_k(1-\mu_k)$ and that the bound is attained when

$A_{k,m}$ can only take values zero or one (so all units within a cluster have the same values for the treatment). In the second stage, each unit in cluster m is assigned to the treatment independently, with cluster-specific probability $A_{k,m}$.

A.2. Base case: Difference in means

Let

$$N_{k,1} = \sum_{i=1}^{n_k} R_{k,i} W_{k,i} \quad \text{and} \quad N_{k,0} = \sum_{i=1}^{n_k} R_{k,i}(1 - W_{k,i})$$

be the number of treated and untreated units in the sample, respectively. The total sample size is $N_k = N_{k,1} + N_{k,0}$. We consider the simple difference of means between treated and non-treated, which is obtained as the coefficient on the treatment indicator in a regression of the outcome on a constant and the treatment,

$$\widehat{\tau}_k = \frac{1}{N_{k,1} \vee 1} \sum_{i=1}^{n_k} R_{k,i} W_{k,i} Y_{k,i} - \frac{1}{N_{k,0} \vee 1} \sum_{i=1}^{n_k} R_{k,i}(1 - W_{k,i}) Y_{k,i}.$$

We make the following assumptions about the sampling process and the cluster sizes: *(i)* $q_k m_k \to \infty$, *(ii)* $\liminf_{k\to\infty} p_k \min_m n_{k,m} > 0$, and *(iii)* $\limsup_{k\to\infty} \max_m n_{k,m} / \min_m n_{k,m} < \infty$. The first assumption implies that the expected number of sampled clusters goes to infinity as k increases. The second assumption implies that the average number of observations sampled per cluster, conditional on the cluster being sampled, does not go to zero. The third assumption restricts the imbalance between the number of units across clusters. Notice that assumptions *(i)* and *(ii)* imply $n_k p_k q_k \to \infty$, so the sample size becomes larger in expectation as k increases.

A.2.1. Large k distribution

Let $\alpha_k = (1/n_k) \sum_{i=1}^{n_k} y_{k,i}(0)$ and $\tau_k = (1/n_k) \sum_{i=1}^{n_k} (y_{k,i}(1) - y_{i,k}(0))$, $u_{k,i}(1) = y_{k,i}(1) - (\alpha_k + \tau_k)$, and $u_{k,i}(0) = y_{k,i}(0) - \alpha_k$. Notice that,

$$\sum_{i=1}^{n_k} u_{k,i}(1) = \sum_{i=1}^{n_k} u_{k,i}(0) = 0.$$

This implies

$$\sqrt{n_k p_k q_k}(\widehat{\tau}_k - \tau_k) = \frac{b_{k,1}}{\widehat{b}_{k,1}} \widehat{a}_{k,1} - \frac{b_{k,0}}{\widehat{b}_{k,0}} \widehat{a}_{k,0},$$

where

$$\widehat{a}_{k,1} = \frac{1}{\sqrt{n_k p_k q_k} \mu_k} \sum_{i=1}^{n_k} (R_{k,i} W_{k,i} - p_k q_k \mu_k) u_{k,i}(1),$$

$$\widehat{a}_{k,0} = \frac{1}{\sqrt{n_k p_k q_k}(1 - \mu_k)} \sum_{i=1}^{n_k} (R_{k,i}(1 - W_{k,i}) - p_k q_k (1 - \mu_k)) u_{k,i}(0),$$

$\widehat{b}_{k,1} = (N_{k,1} \vee 1)/n_k$, $\widehat{b}_{k,0} = (N_{k,0} \vee 1)/n_k$, $b_{k,1} = p_k q_k \mu_k$ and $b_{k,0} = p_k q_k (1-\mu_k)$. We will first derive the large sample distribution of

$$\begin{aligned}\widehat{a}_k &= \widehat{a}_{k,1} - \widehat{a}_{k,0} \\ &= \sum_{m=1}^{m_k} \big(\xi_{k,m,1} - \xi_{k,m,0}\big),\end{aligned}$$

where

$$\xi_{k,m,1} = \frac{1}{\sqrt{n_k p_k q_k}\mu_k} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\}\big(R_{k,i}W_{k,i} - p_k q_k \mu_k\big) u_{k,i}(1),$$

and

$$\xi_{k,m,0} = \frac{1}{\sqrt{n_k p_k q_k}(1-\mu_k)} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\}\big(R_{k,i}(1-W_{k,i}) - p_k q_k (1-\mu_k)\big) u_{k,i}(0).$$

Notice that $E[\xi_{k,m,1}] = E[\xi_{k,m,0}] = 0$. Moreover, notice that the terms $\xi_{k,m,1} - \xi_{k,m,0}$ are independent across clusters, m. In addition,

$$\begin{aligned}E[\xi^2_{k,m,1}] = \frac{1}{n_k}\sum_{i=1}^{n_k} & 1\{m_{k,i} = m\}\frac{1 - p_k q_k \mu_k}{\mu_k} u^2_{k,i}(1) \\ & + \frac{2}{n_k}\sum_{i=1}^{n_k-1}\sum_{j=i+1}^{n_k} 1\{m_{k,i} = m_{k,j} = m\}\frac{p_k(\sigma_k^2 + \mu_k^2(1-q_k))}{\mu_k^2} u_{k,i}(1)u_{k,j}(1).\end{aligned}$$

$$\begin{aligned}E[\xi^2_{k,m,0}] = \frac{1}{n_k}\sum_{i=1}^{n_k} & 1\{m_{k,i} = m\}\frac{1 - p_k q_k (1-\mu_k)}{(1-\mu_k)} u^2_{k,i}(0) \\ & + \frac{2}{n_k}\sum_{i=1}^{n_k-1}\sum_{j=i+1}^{n_k} 1\{m_{k,i} = m_{k,j} = m\}\frac{p_k(\sigma_k^2 + (1-\mu_k)^2(1-q_k))}{(1-\mu_k)^2} u_{k,i}(0)u_{k,j}(0),\end{aligned}$$

and

$$\begin{aligned}E[\xi_{k,m,1}\xi_{k,m,0}] &= -\frac{1}{n_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} p_k q_k u_{k,i}(1)u_{k,i}(0) \\ &+ \frac{1}{n_k}\sum_{i=1}^{n_k-1}\sum_{j=i+1}^{n_k} 1\{m_{k,i} = m_{k,j} = m\}\frac{p_k(\mu_k(1-\mu_k)(1-q_k) - \sigma_k^2)}{\mu_k(1-\mu_k)}\big(u_{k,i}(0)u_{k,j}(1) + u_{k,i}(1)u_{k,j}(0)\big).\end{aligned}$$

We obtain:

$$\begin{aligned}& n_k E[(\xi_{k,m,1} - \xi_{k,m,0})^2] \\ &= \frac{1}{\mu_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} u^2_{k,i}(1) + \frac{1}{1-\mu_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} u^2_{k,i}(0) \\ &+ 2p_k \sum_{i=1}^{n_k-1}\sum_{j=i+1}^{n_k} 1\{m_{k,i} = m_{k,j} = m\}\Big(u_{k,i}(1)u_{k,j}(1) + u_{k,i}(0)u_{k,j}(0) - u_{k,i}(0)u_{k,j}(1) - u_{k,i}(1)u_{k,j}(0)\Big)\end{aligned}$$

$$
\begin{aligned}
&- p_k q_k \Bigg(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} \big(u_{k,i}^2(1) + u_{k,i}^2(0) - 2u_{k,i}(1)u_{k,i}(0) \big) \\
&\qquad + 2 \sum_{i=1}^{n_k-1} \sum_{j=i+1}^{n_k} 1\{m_{k,i} = m_{k,j} = m\} \big(u_{k,i}(1)u_{k,j}(1) + u_{k,i}(0)u_{k,j}(0) - u_{k,i}(0)u_{k,j}(1) - u_{k,i}(1)u_{k,j}(0) \big) \Bigg) \\
&+ 2p_k\sigma_k^2 \Bigg(\sum_{i=1}^{n_k-1} \sum_{j=i+1}^{n_k} 1\{m_{k,i} = m_{k,j} = m\} \bigg(\frac{u_{k,i}(1)u_{k,j}(1)}{\mu_k^2} + \frac{u_{k,i}(0)u_{k,j}(0)}{(1-\mu_k)^2} + \frac{u_{k,i}(0)u_{k,j}(1)}{\mu_k(1-\mu_k)} + \frac{u_{k,i}(1)u_{k,j}(0)}{\mu_k(1-\mu_k)} \bigg) \Bigg).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&n_k E[(\xi_{k,m,1} - \xi_{k,m,0})^2] \\
&\quad = \frac{1}{\mu_k} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} u_{k,i}^2(1) + \frac{1}{1-\mu_k} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} u_{k,i}^2(0) \\
&\quad + p_k \Bigg[\bigg(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} \big(u_{k,i}(1) - u_{k,i}(0) \big) \bigg)^2 - \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} \big(u_{k,i}(1) - u_{k,i}(0) \big)^2 \Bigg] \\
&\quad - p_k q_k \bigg(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} \big(u_{k,i}(1) - u_{k,i}(0) \big) \bigg)^2 \\
&\quad + p_k \sigma_k^2 \Bigg[\bigg(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} \bigg(\frac{u_{k,i}(1)}{\mu_k} + \frac{u_{k,i}(0)}{1-\mu_k} \bigg) \bigg)^2 - \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} \bigg(\frac{u_{k,i}(1)}{\mu_k} + \frac{u_{k,i}(0)}{1-\mu_k} \bigg)^2 \Bigg].
\end{aligned}
$$

Let $v_k = \sum_{m=1}^{m_k} E[(\xi_{k,m,1} - \xi_{k,m,0})^2]$, then

$$
\begin{aligned}
n_k v_k &= \sum_{i=1}^{n_k} \bigg(\frac{u_{k,i}^2(1)}{\mu_k} + \frac{u_{k,i}^2(0)}{1-\mu_k} \bigg) \\
&\quad + p_k \sum_{m=1}^{m_k} \Bigg[\bigg(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} \big(u_{k,i}(1) - u_{k,i}(0) \big) \bigg)^2 - \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} \big(u_{k,i}(1) - u_{k,i}(0) \big)^2 \Bigg] \\
&\quad - p_k q_k \sum_{m=1}^{m_k} \bigg(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} \big(u_{k,i}(1) - u_{k,i}(0) \big) \bigg)^2 \\
&\quad + p_k \sigma_k^2 \sum_{m=1}^{m_k} \Bigg[\bigg(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} \bigg(\frac{u_{k,i}(1)}{\mu_k} + \frac{u_{k,i}(0)}{1-\mu_k} \bigg) \bigg)^2 - \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} \bigg(\frac{u_{k,i}(1)}{\mu_k} + \frac{u_{k,i}(0)}{1-\mu_k} \bigg)^2 \Bigg].
\end{aligned}
$$

Alternatively, we can write this expression as

$$
\begin{aligned}
n_k v_k &= \sum_{i=1}^{n_k} \bigg(\frac{u_{k,i}^2(1)}{\mu_k} + \frac{u_{k,i}^2(0)}{1-\mu_k} \bigg) \\
&\quad - p_k \sum_{i=1}^{n_k} \big(u_{k,i}(1) - u_{k,i}(0) \big)^2 - p_k \sigma_k^2 \sum_{i=1}^{n_k} \bigg(\frac{u_{k,i}(1)}{\mu_k} + \frac{u_{k,i}(0)}{1-\mu_k} \bigg)^2 \\
&\quad + p_k (1 - q_k) \sum_{m=1}^{m_k} \bigg(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} \big(u_{k,i}(1) - u_{k,i}(0) \big) \bigg)^2
\end{aligned}
$$

$$+ p_k\sigma_k^2 \sum_{m=1}^{m_k}\left(\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}\left(\frac{u_{k,i}(1)}{\mu_k}+\frac{u_{k,i}(0)}{1-\mu_k}\right)\right)^2.$$

The sum of the first three terms is minimized for $p_k = 1$ and $\sigma_k^2 = \mu_k(1-\mu_k)$, in which case this sum is equal to zero. Therefore,

$$v_k \geqslant (p_k \min_m n_{k,m})(1-q_k)\sum_{m=1}^{m_k}\frac{n_{k,m}}{n_k}\left(\frac{1}{n_{k,m}}\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}\big(u_{k,i}(1)-u_{k,i}(0)\big)\right)^2$$
$$+ (p_k \min_m n_{k,m})\,\sigma_k^2\sum_{m=1}^{m_k}\frac{n_{k,m}}{n_k}\left(\frac{1}{n_{k,m}}\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}\left(\frac{u_{k,i}(1)}{\mu_k}+\frac{u_{k,i}(0)}{1-\mu_k}\right)\right)^2. \qquad \text{(A.1)}$$

We will assume that $\liminf_{k\to\infty}((1-q_k)\vee\sigma_k^2) > 0$, so either sampling or assignment or both are correlated within cluster. (We study the case $q_k = 1$ and $\sigma_k^2 = 0$ separately below.) In addition, assume *(i)* $\liminf_{k\to\infty}(1-q_k) > 0$ and

$$\liminf_{k\to\infty}\sum_{m=1}^{m_k}\frac{n_{k,m}}{n_k}\left(\frac{1}{n_{k,m}}\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}\big(u_{k,i}(1)-u_{k,i}(0)\big)\right)^2 > 0, \qquad \text{(A.2)}$$

or *(ii)* $\liminf_{k\to\infty}\sigma_k^2 > 0$ and

$$\liminf_{k\to\infty}\sum_{m=1}^{m_k}\frac{n_{k,m}}{n_k}\left(\frac{1}{n_{k,m}}\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}\left(\frac{u_{k,i}(1)}{\mu_k}+\frac{u_{k,i}(0)}{1-\mu_k}\right)\right)^2 > 0. \qquad \text{(A.3)}$$

Equation (A.2) would be violated if, as k increases, there is no variation in average treatment effects across clusters. Equation (A.3) would be violated if as k increases there is no variation in average potential outcomes across clusters. If equations (A.2) and (A.3) hold, v_k is bounded below by a term of order at least $p_k \min_m n_{k,m}$. Recall our assumption, $\liminf_{k\to\infty} p_k \min_m n_{k,m} > 0$, so the average number of observations sampled per cluster, conditional on the cluster being sampled, does not go to zero. Then,

$$\liminf_{k\to\infty} v_k > 0.$$

To obtain a CLT, we will check Lyapunov's condition,

$$\lim_{k\to\infty}\sum_{m=1}^{m_k}\frac{1}{v_k^{1+\delta/2}}E[|\xi_{k,m,1}-\xi_{k,m,0}|^{2+\delta}] = 0,$$

for some $\delta > 0$. Because potential outcomes are uniformly bounded and μ_k is uniformly bounded away from zero, we obtain

$$|\xi_{k,m,1}|^{2+\delta} \leqslant c\frac{n_{k,m}^{2+\delta}}{(n_k p_k q_k)^{1+\delta/2}}\left|\frac{1}{n_{k,m}}\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}|R_{k,i}W_{k,i}-p_k q_k\mu_k|\right|^{2+\delta},$$

where c is some generic positive constant, whose value may change across equations. Consider $\delta = 1$, and let

$$S_{k,m,1}^3 = E\left[\left|\frac{1}{n_{k,m}}\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}|R_{k,i}W_{k,i}-p_k q_k\mu_k|\right|^3\right]$$

$$
\begin{aligned}
&\leqslant \frac{1}{n_{k,m}^3} n_{k,m} E[|R_{k,i}W_{k,i} - p_k q_k \mu_k|^3] \\
&+ \frac{3}{n_{k,m}^3} n_{k,m}(n_{k,m}-1) E\big[|R_{k,i}W_{k,i} - p_k q_k \mu_k|^2 |R_{k,j}W_{k,j} - p_k q_k \mu_k| \big| m_{k,i} = m_{k,j} = m\big] \\
&+ \frac{6}{n_{k,m}^3} \binom{n_{k,m}}{3} E\big[|R_{k,i}W_{k,i} - p_k q_k \mu_k||R_{k,j}W_{k,j} - p_k q_k \mu_k||R_{k,t}W_{k,t} - p_k q_k \mu_k| \big| m_{k,i} = m_{k,j} = m_{k,t} = m\big],
\end{aligned}
$$

for $i \neq j \neq t$. (The second and third terms on the left-hand side of last equation only appear when $n_{k,m} \geqslant 2$ and $n_{k,m} \geqslant 3$, respectively) As a result,

$$
\begin{aligned}
S_{k,m,1}^3 &\leqslant c \left(\frac{p_k q_k}{n_{k,m}^2} + \frac{p_k^2 q_k}{n_{k,m}} + p_k^3 q_k \right) \\
&\leqslant c\, p_k^3 q_k \left(\frac{1}{p_k^2 \min_m n_{k,m}^2} + \frac{1}{p_k \min_m n_{k,m}} + 1 \right).
\end{aligned}
$$

Because $\liminf_{k\to\infty} p_k \min_m n_{k,m} > 0$, for large enough k we obtain,

$$
E[|\xi_{k,m,1}|^3] \leqslant c \frac{p_k^3 q_k n_{k,m}^3}{(n_k p_k q_k)^{3/2}},
$$

and the same bound applies for $E[|\xi_{k,m,0}|^3]$. Notice that

$$
\begin{aligned}
\sum_{m=1}^{m_k} E[|\xi_{k,m,1} - \xi_{k,m,0}|^3] &\leqslant \sum_{m=1}^{m_k} E[(|\xi_{k,m,1}| + |\xi_{k,m,0}|)^3] \\
&= \sum_{m=1}^{m_k} E[|\xi_{k,m,1}|^3] + \sum_{m=1}^{m_k} E[|\xi_{k,m,0}|^3] \\
&+ 3 \sum_{m=1}^{m_k} E[|\xi_{k,m,1}|^2 |\xi_{k,m,0}|] + 3 \sum_{m=1}^{m_k} E[|\xi_{k,m,1}||\xi_{k,m,0}|^2].
\end{aligned}
$$

Now, Hölder's inequality implies that

$$
\frac{p_k^3 q_k \sum_{m=1}^{m_k} n_{k,m}^3}{v_k^{3/2} (n_k p_k q_k)^{3/2}} \longrightarrow 0, \tag{A.4}
$$

is sufficient for the Lyapunov condition to hold. Because $\max_m n_{k,m} / \min_m n_{k,m}$ is bounded asymptotically, we obtain,

$$
\begin{aligned}
\limsup_{k\to\infty} \frac{p_k^3 q_k \sum_{m=1}^{m_k} n_{k,m}^3}{v_k^{3/2} (n_k p_k q_k)^{3/2}} &\leqslant \limsup_{k\to\infty} c \frac{p_k^3 q_k m_k \max_m n_{k,m}^3}{(p_k^2 q_k m_k \min_m n_{k,m}^2)^{3/2}} \\
&\leqslant \limsup_{k\to\infty} \left(\frac{\max_m n_{k,m}}{\min_m n_{k,m}} \right)^3 \frac{c}{\sqrt{q_k m_k}} = 0,
\end{aligned}
$$

and so the Lyapunov condition holds. As a result, we obtain

$$
\widehat{a}_k / \sqrt{v_k} \xrightarrow{d} N(0,1).
$$

We will next prove that both $\widehat{a}_{k,1}/\sqrt{v_k}$ and $\widehat{a}_{k,0}/\sqrt{v_k}$ are $\mathcal{O}_p(1)$.

$$
\begin{aligned}
E[\widehat{a}_{k,1}^2] &= \frac{1}{n_k p_k q_k}\frac{1}{\mu_k^2}\sum_{m=1}^{m_k} E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}(R_{k,i}W_{k,i}-p_kq_k\mu_k)u_{k,i}(1)\right)^2\right] \\
&\leqslant c\,\frac{1}{n_kp_kq_k}\sum_{m=1}^{m_k}\big(n_{k,m}p_kq_k + n_{k,m}(n_{k,m}-1)p_k^2q_k\big) \\
&= c\Big(1+\sum_{m=1}^{m_k}\frac{n_{k,m}(n_{k,m}-1)p_k}{n_k}\Big).
\end{aligned}
$$

Therefore,

$$
E[(\widehat{a}_{k,1}/\sqrt{v_k})^2] \leqslant c\left(\frac{1}{p_k\min_m n_{k,m}} + \sum_{m=1}^{m_k}\frac{(\max_m n_{k,m})(n_{k,m}-1)p_k}{n_kp_k\min_m n_{k,m}}\right).
$$

Because $\limsup \max_m n_{k,m}/\min_m n_{k,m}<\infty$, we obtain $\limsup_{k\to\infty} E[(\widehat{a}_{k,1}/v_k)^2]<\infty$. As a result, $\widehat{a}_{k,1}/\sqrt{v_k}$ is $\mathcal{O}_p(1)$.

Let $\widetilde{b}_{k,1}=N_{k,1}/n_k$. Consider k large enough, so $p_k\min_m n_{k,m}$ is bounded away from zero, making $\widetilde{b}_{k,1}/b_{k,1}$ well-defined. Notice that $E[\widetilde{b}_{k,1}/b_{k,1}]=1$ and

$$
\begin{aligned}
\mathrm{var}(\widetilde{b}_{k,1}/b_{k,1}) &= \frac{1}{(n_kp_kq_k\mu_k)^2}\sum_{m=1}^{m_k} E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}(R_{k,i}W_{k,i}-n_kp_kq_k\mu_k)\right)^2\right] \\
&= \frac{n_kp_kq_k\mu_k(1-p_kq_k\mu_k)}{(n_kp_kq_k\mu_k)^2} + \sum_{m=1}^{m_k}\frac{n_{k,m}(n_{k,m}-1)p_k^2q_k(\sigma_k^2+(1-q_k)\mu_k^2)}{(n_kp_kq_k\mu_k)^2} \\
&\leqslant \frac{1-p_kq_k\mu_k}{n_kp_kq_k\mu_k} + c\,\frac{n_k(\max_m n_{k,m}-1)p_k^2q_k}{(n_kp_kq_k)^2} \\
&\leqslant \frac{1-p_kq_k\mu_k}{n_kp_kq_k\mu_k} + c\,\frac{(\max_m n_{k,m}-1)}{\min n_{k,m}}\frac{1}{q_km_k} \longrightarrow 0.
\end{aligned}
$$

This implies $\widetilde{b}_{k,1}/b_{k,1}\xrightarrow{p}1$. Analogous calculations yield $\widetilde{b}_{0,k}/b_{0,k}\xrightarrow{p}1$. For large enough k, $\widetilde{b}_{k,1}/b_{k,1}=0$ if and only if $N_{k,1}=0$, which implies $\Pr(N_{k,1}=0)\to 0$. It follows that, for large enough k,

$$
\Pr(|\widetilde{b}_{k,1}/b_{k,1}-\widehat{b}_{k,1}/b_{k,1}|=0)=\Pr(N_{k,1}>0)\longrightarrow 1
$$

and $\widehat{b}_{k,1}/b_{k,1}\xrightarrow{p}1$. Using analogous calculations, we obtain $\widehat{b}_{k,0}/b_{k,0}\xrightarrow{p}1$. As a result,

$$
\begin{aligned}
\sqrt{n_kp_kq_k}(\widehat{\tau}_k-\tau_k)/v_k^{1/2} &= \frac{b_{k,1}}{\widehat{b}_{k,1}}\frac{\widehat{a}_{k,1}}{v_k^{1/2}} - \frac{b_{k,0}}{\widehat{b}_{k,0}}\frac{\widehat{a}_{k,0}}{v_k^{1/2}} \\
&= \frac{\widehat{a}_k}{v_k^{1/2}} + \left(\frac{b_{k,1}}{\widehat{b}_{k,1}}-1\right)\frac{\widehat{a}_{k,1}}{v_k^{1/2}} - \left(\frac{b_{k,0}}{\widehat{b}_{k,0}}-1\right)\frac{\widehat{a}_{k,0}}{v_k^{1/2}} \\
&= \widehat{a}_k/\sqrt{v_k} + o_p(1).
\end{aligned}
$$

Therefore,

$$\sqrt{n_k p_k q_k}(\widehat{\tau}_k - \tau_k)/v_k^{1/2} \xrightarrow{d} N(0,1).$$

Using $\widetilde{b}_{1,k}/b_{1,k} \xrightarrow{p} 1$ and $\widetilde{b}_{0,k}/b_{0,k} \xrightarrow{p} 1$, it is easy to show $N_k/(n_k p_k q_k) \xrightarrow{p} 1$, which implies

$$\sqrt{N_k}(\widehat{\tau}_k - \tau_k)/v_k^{1/2} \xrightarrow{d} N(0,1).$$

We will next consider the case of $q_k = 1$ and $\sigma_k^2 = 0$, where no clustering is required. Consider

$$\vartheta_{k,i,1} = \frac{1}{\sqrt{n_k p_k}\mu_k}\big(R_{k,i}W_{k,i} - p_k\mu_k\big)u_{k,i}(1)$$

and

$$\vartheta_{k,i,0} = \frac{1}{\sqrt{n_k p_k}(1-\mu_k)}\big(R_{k,i}(1-W_{k,i}) - p_k(1-\mu_k)\big)u_{k,i}(0).$$

Redefine now $v_k = \sum_{i=1}^{n_k} E\big[(\vartheta_{k,i,1} - \vartheta_{k,i,0})^2\big]$. Then,

$$v_k = \frac{1}{n_k}\sum_{i=1}^{n_k}\left(\frac{u_{k,i}^2(1)}{\mu_k} + \frac{u_{k,i}^2(0)}{1-\mu_k}\right) - p_k\frac{1}{n_k}\sum_{i=1}^{n_k}\big(u_{k,i}(1) - u_{k,i}(0)\big)^2.$$

Notice that v_k is minimized for $p_k = 1$, in which case

$$\begin{aligned}
v_k &= \frac{1}{n_k}\sum_{i=1}^{n_k}\left(\frac{u_{k,i}^2(1)}{\mu_k} + \frac{u_{k,i}^2(0)}{1-\mu_k}\right) - \frac{1}{n_k}\sum_{i=1}^{n_k}\big(u_{k,i}(1) - u_{k,i}(0)\big)^2\\
&= \frac{1}{n_k}\sum_{i=1}^{n_k}\left(\frac{1-\mu_k}{\mu_k}u_{k,i}^2(1) + \frac{\mu_k}{1-\mu_k}u_{k,i}^2(0) + 2u_{k,i}(1)u_{k,i}(0)\right)\\
&= \mu_k(1-\mu_k)\frac{1}{n_k}\sum_{i=1}^{n_k}\left(\frac{u_{k,i}^2(1)}{\mu_k^2} + \frac{u_{k,i}^2(0)}{(1-\mu_k)^2} + 2\,\frac{u_{k,i}(1)u_{k,i}(0)}{\mu_k(1-\mu_k)}\right)\\
&= \mu_k(1-\mu_k)\frac{1}{n_k}\sum_{i=1}^{n_k}\left(\frac{u_{k,i}(1)}{\mu_k} + \frac{u_{k,i}(0)}{1-\mu_k}\right)^2.
\end{aligned}$$

Therefore, the assumption

$$\liminf_{k\to\infty}\frac{1}{n_k}\sum_{i=1}^{n_k}\left(\frac{u_{k,i}(1)}{\mu_k} + \frac{u_{k,i}(0)}{1-\mu_k}\right)^2 > 0$$

is enough for $\liminf_{k\to\infty} v_k > 0$. Notice now that

$$\begin{aligned}
E[|\vartheta_{k,i,1}|^3] &\leqslant \frac{1}{(n_k p_k)^{3/2}}E[|R_{k,i}W_{k,i} - p_k\mu_k|^3]\\
&= \frac{1}{(n_k p_k)^{3/2}}(1-p_k\mu_k)^3 p_k\mu_k + (p_k\mu_k)^3(1-p_k\mu_k)\\
&\leqslant c\frac{p_k}{(n_k p_k)^{3/2}},
\end{aligned}$$

and the same bound holds for $E[|\vartheta_{k,i,0}|^3]$. Therefore, for the Lyapunov condition to hold, it is enough that

$$\frac{n_k p_k}{(n_k p_k)^{3/2}} = \frac{1}{\sqrt{n_k p_k}} \longrightarrow 0,$$

or $n_k p_k \to \infty$. That is, assumptions *(i)*-*(iii)*, which we used for the clustered case, are replaced by $n_k p_k \to \infty$.

A.2.2. Estimation of the variance

Let $\widehat{U}_{k,i} = Y_{k,i} - \widehat{\alpha}_k - \widehat{\tau}_k W_{k,i}$ be the residuals from the regression of $Y_{k,i}$ or a constant and $W_{k,i}$. Here, $\widehat{\alpha}_k$ is the coefficient on the constant regressor equal to one, and $\widehat{\tau}_k$ is the coefficient on $W_{k,i}$. We have already shown $v_k^{-1/2}(\widehat{\tau}_k - \tau_k) = \mathcal{O}_p(1/\sqrt{n_k p_k q_k})$. The same is true about $\widehat{\alpha}_k$ (e.g., apply the proof for $\widehat{\tau}_k$ after replacing each $y_{k,i}(1)$ with a zero). Define $\widehat{\Sigma}_k = \sum_{m=1}^{m_k} \widehat{\Sigma}_{k,m}$, where

$$\widehat{\Sigma}_{k,m} = \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} \begin{pmatrix} \widehat{U}_{k,i} \\ W_{k,i}\widehat{U}_{k,i} \end{pmatrix}\right) \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} \begin{pmatrix} \widehat{U}_{k,i} \\ W_{k,i}\widehat{U}_{k,i} \end{pmatrix}\right)'.$$

Also, let

$$\widehat{Q}_k = \sum_{i=1}^{n_k} R_{k,i} \begin{pmatrix} 1 \\ W_{k,i} \end{pmatrix} \begin{pmatrix} 1 \\ W_{k,i} \end{pmatrix}',$$

and $z = (0,1)'$. Then, the cluster estimator of the variance of $\sqrt{N_k}(\widehat{\tau}_k - \tau_k)$ is

$$\widehat{V}_k^{\text{cluster}} = N_k z' \widehat{Q}_k^{-1} \widehat{\Sigma}_k \widehat{Q}_k^{-1} z.$$

Notice that

$$(n_k p_k q_k)^{-1} E[\widehat{Q}_k] = \begin{pmatrix} 1 & \mu_k \\ \mu_k & \mu_k \end{pmatrix}.$$

In addition,

$$\frac{1}{n_k p_k q_k} \widehat{Q}_k(2,2) = \frac{1}{n_k p_k q_k} \sum_{m=1}^{m_k} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}.$$

$$\begin{aligned} \operatorname{var}\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}\right) &= n_{k,m} p_k q_k \mu_k (1 - p_k q_k \mu_k) \\ &\quad + n_{k,m}(n_{k,m} - 1) p_k^2 q_k (\sigma_k^2 + \mu_k^2 (1 - q_k)). \end{aligned}$$

Therefore, under conditions *(i)*-*(iii)*, we obtain

$$\begin{aligned} \operatorname{var}\left(\frac{1}{n_k p_k q_k} \widehat{Q}_k(2,2)\right) &\leqslant \frac{c}{n_k p_k q_k} \Big(1 + p_k(\max_m n_{k,m} - 1)\Big) \\ &= c\,\frac{\max_m n_{k,m}}{n_k q_k} + o(1) \\ &\leqslant c\,\frac{\max_m n_{k,m}}{\min_m n_{k,m}} \frac{1}{q_k m_k} + o(1) \longrightarrow 0. \end{aligned}$$

Analogous calculations yield $\operatorname{var}((n_k p_k q_k)^{-1}\widehat{Q}_k(1,1)) \to 0$. Therefore,

$$\frac{1}{n_k p_k q_k}\widehat{Q}_k = \begin{pmatrix} 1 & \mu_k \\ \mu_k & \mu_k \end{pmatrix} + o_p(1)$$

and

$$n_k q_k p_k \widehat{Q}_k^{-1} = H_k + o_p(1), \quad \text{where} \quad H_k = \frac{1}{\mu_k(1-\mu_k)}\begin{pmatrix} \mu_k & -\mu_k \\ -\mu_k & 1 \end{pmatrix}.$$

Now, let $U_{k,i} = Y_{k,i} - \alpha_k - \tau_k W_{k,i} = W_{k,i}u_{k,i}(1) + (1-W_{k,i})u_{k,i}(0)$. Notice that

$$v_k^{-1/2} \max_{i=1,\ldots,n_k} |\widehat{U}_{k,i} - U_{k,i}| \leqslant v_k^{-1/2}|\widehat{\alpha}_k - \alpha_k| + v_k^{-1/2}|\widehat{\tau}_k - \tau_k| = \mathcal{O}_p(1/\sqrt{n_k p_k q_k}).$$

Define $\overline{\Sigma}_k = \sum_{m=1}^{m_k} \overline{\Sigma}_{k,m}$, where

$$\overline{\Sigma}_{k,m} = \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} \begin{pmatrix} U_{k,i} \\ W_{k,i}U_{k,i} \end{pmatrix}\right)\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} \begin{pmatrix} U_{k,i} \\ W_{k,i}U_{k,i} \end{pmatrix}\right)'.$$

We will show

$$\frac{1}{n_k p_k q_k v_k}(\widehat{\Sigma}_k - \overline{\Sigma}_k) \xrightarrow{p} 0.$$

Notice that

$$\begin{aligned}\widehat{\Sigma}_{k,m}(2,2) - \overline{\Sigma}_{k,m}(2,2) &= \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}(\widehat{U}_{k,i} - U_{k,i})\right)^2 \\ &\quad + 2\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i} U_{k,i}\right)\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}(\widehat{U}_{k,i} - U_{k,i})\right).\end{aligned}$$

Therefore,

$$\begin{aligned}\frac{1}{n_k p_k q_k v_k}|\widehat{\Sigma}_k(2,2) - \overline{\Sigma}_k(2,2)| &\leqslant c\frac{1}{n_k p_k q_k v_k}\sum_{m=1}^{m_k}\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}\right)^2 \\ &\quad \times \left(\max_{i=1,\ldots,n_k} |\widehat{U}_{k,i} - U_{k,i}|^2 + \max_{i=1,\ldots,n_k} |\widehat{U}_{k,i} - U_{k,i}|\right).\end{aligned}$$

The same expression holds for the off-diagonal elements of $\widehat{\Sigma}_{k,m} - \overline{\Sigma}_{k,m}$. For $\widehat{\Sigma}_{k,m}(1,1) - \overline{\Sigma}_{k,m}(1,1)$, the expression holds once we replace each $W_{k,i}$ with a one. Let $\|\cdot\|$ be the Frobenius norm of a matrix. Then,

$$\begin{aligned}\frac{1}{n_k p_k q_k v_k}\|\widehat{\Sigma}_k - \overline{\Sigma}_k\| &\leqslant c\frac{1}{n_k p_k q_k v_k}\sum_{m=1}^{m_k}\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i}\right)^2 \\ &\quad \times \left(\max_{i=1,\ldots,n_k} |\widehat{U}_{k,i} - U_{k,i}|^2 + \max_{i=1,\ldots,n_k} |\widehat{U}_{k,i} - U_{k,i}|\right).\end{aligned}$$

We will prove that the right-hand side of the previous equation converges to zero in probability. We will factorize each term into a expression that is bounded in probability and one that converges to zero in L_1.

$$E\left[\sum_{m=1}^{m_k}\left(\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}R_{k,i}\right)^2\right] \leqslant n_k p_k q_k + n_k(\max_m n_{k,m}-1)p_k^2 q_k.$$

For the first term, notice that

$$\begin{aligned}
&\max_{i=1,\ldots,n_k} |\widehat{U}_{k,i}-U_{k,i}|^2 \frac{n_k p_k q_k + n_k(\max_m n_{k,m}-1)p_k^2 q_k}{n_k p_k q_k v_k} \\
&\qquad = \frac{n_k p_k q_k}{v_k} \max_{i=1,\ldots,n_k} |\widehat{U}_{k,i}-U_{k,i}|^2 \left(\frac{n_k p_k q_k + n_k(\max_m n_{k,m}-1)p_k^2 q_k}{(n_k p_k q_k)^2}\right) \\
&\qquad \leqslant \frac{n_k p_k q_k}{v_k} \max_{i=1,\ldots,n_k} |\widehat{U}_{k,i}-U_{k,i}|^2 \left(\frac{1}{n_k p_k q_k} + \frac{\max_m n_{k,m}-1}{\min_m n_{k,m}}\frac{1}{q_k m_k}\right) \\
&\qquad = \mathcal{O}_p(1)\, o(1).
\end{aligned}$$

For the second term, using the fact that v_k is greater or equal to $p_k \min_m n_{k,m} > 0$ times a term with limit inferior that is bounded away from zero, we obtain

$$\begin{aligned}
&\max_{i=1,\ldots,n_k} |\widehat{U}_{k,i}-U_{k,i}| \frac{n_k p_k q_k + n_k(\max_m n_{k,m}-1)p_k^2 q_k}{n_k p_k q_k v_k} \\
&\qquad = \left(\frac{n_k p_k q_k}{v_k}\right)^{1/2} \max_{i=1,\ldots,n_k} |\widehat{U}_{k,i}-U_{k,i}| \left(\frac{n_k p_k q_k + n_k(\max_m n_{k,m}-1)p_k^2 q_k}{(n_k p_k q_k)^{3/2} v_k^{1/2}}\right) \\
&\qquad \leqslant \left(\frac{n_k p_k q_k}{v_k}\right)^{1/2} \max_{i=1,\ldots,n_k} |\widehat{U}_{k,i}-U_{k,i}| \left(\frac{1}{(n_k p_k q_k v_k)^{1/2}} + \frac{\max_m n_{k,m}-1}{\min_m n_{k,m}}\frac{1}{(q_k m_k)^{1/2}}\right) \\
&\qquad = \mathcal{O}_p(1)\, o(1).
\end{aligned}$$

As a result, we obtain

$$\frac{1}{n_k p_k q_k v_k}\|\widehat{\Sigma}_k - \overline{\Sigma}_k\| = o_p(1).$$

Notice that

$$\begin{aligned}
\frac{n_k p_k q_k}{v_k}\widehat{Q}_k^{-1}\widehat{\Sigma}_k\widehat{Q}_k^{-1} - H_k\frac{\overline{\Sigma}_k}{n_k p_k q_k v_k}H_k &= H_k\frac{\overline{\Sigma}_k}{n_k p_k q_k v_k}\left(n_k p_k q_k \widehat{Q}_k^{-1} - H_k\right) \\
&+ \left(n_k p_k q_k \widehat{Q}_k^{-1} - H_k\right)\frac{\overline{\Sigma}_k}{n_k p_k q_k v_k}\left(n_k p_k q_k \widehat{Q}_k^{-1}\right) + \left(n_k p_k q_k \widehat{Q}_k^{-1}\right)\frac{\widehat{\Sigma}_k - \overline{\Sigma}_k}{n_k p_k q_k v_k}\left(n_k p_k q_k \widehat{Q}_k^{-1}\right).
\end{aligned}$$

Therefore, to show that the left-hand side of the last equation is $o_p(1)$, it is only left to show that $\overline{\Sigma}_k/(n_k p_k q_k v_k)$ is $\mathcal{O}_p(1)$. We will prove this next. Notice that

$$\frac{1}{n_k p_k q_k v_k}\|\overline{\Sigma}_k\| \leqslant c\frac{1}{n_k p_k q_k v_k}\sum_{m=1}^{m_k}\left(\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}R_{k,i}\right)^2.$$

Therefore,

$$E\left[\frac{1}{n_k p_k q_k v_k}\|\overline{\Sigma}_k\|\right] \leqslant c\frac{1}{n_k p_k q_k v_k}\left(n_k p_k q_k + n_k(\max_m n_{k,m} - 1)p_k^2 q_k\right).$$

Then,

$$E\left[\frac{1}{n_k p_k q_k v_k}\|\overline{\Sigma}_k\|\right] \leqslant c\left(\frac{1}{v_k} + \frac{p_k(\max_m n_{k,m} - 1)}{p_k \min_m n_{k,m}}\right) < \infty.$$

We, therefore, obtain,

$$\frac{n_k p_k q_k}{v_k}\widehat{Q}_k^{-1}\widehat{\Sigma}_k\widehat{Q}_k^{-1} - H_k\frac{\overline{\Sigma}_k}{n_k p_k q_k v_k}H_k \xrightarrow{p} 0.$$

Because $N_k/(n_k p_k q_k) \xrightarrow{p} 1$, we obtain

$$\begin{aligned}
\widehat{V}_k^{\text{cluster}}/v_k &= z' H_k \frac{\overline{\Sigma}_k}{n_k p_k q_k v_k} H_k z + o_p(1) \\
&= \frac{1}{n_k p_k q_k v_k}\left(\frac{1}{\mu_k(1-\mu_k)}\right)^2 \sum_{m=1}^{m_k}\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}(W_{k,i} - \mu_k)U_{k,i}\right)^2 + o_p(1).
\end{aligned}$$

Recall that $U_{k,i}^2 = u_{k,i}^2(1)W_{k,i} + u_{k,i}^2(0)(1 - W_{k,i})$. Notice that

$$\begin{aligned}
&E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}(W_{k,i} - \mu_k)U_{k,i}\right)^2\right] \\
&\quad= \sum_{i=1}^{n_k} 1\{m_{k,i} = m\}p_k q_k \mu_k(1-\mu_k)\Big((1-\mu_k)u_{k,i}^2(1) + \mu_k u_{k,i}^2(0)\Big) \\
&\quad+ 2\sum_{i=1}^{n_k-1}\sum_{j=i+1}^{n_k} 1\{m_{k,i} = m_{k,j} = m\}p_k^2 q_k\Big[(\sigma_k^2 + \mu_k^2)(1-\mu_k)^2 u_{k,i}(1)u_{k,j}(1) \\
&\qquad+ \mu_k(1-\mu_k)(\sigma_k^2 - \mu_k(1-\mu_k))(u_{k,i}(0)u_{k,j}(1) + u_{k,i}(1)u_{k,j}(0)) \\
&\qquad+ (\sigma_k^2 + (1-\mu_k)^2)\mu_k^2 u_{k,i}(0)u_{k,j}(0)\Big].
\end{aligned}$$

Let

$$v_k^{\text{cluster}} = \frac{1}{n_k p_k q_k}\left(\frac{1}{\mu_k(1-\mu_k)}\right)^2 \sum_{m=1}^{m_k} E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}(W_{k,i} - \mu_k)U_{k,i}\right)^2\right].$$

Then,

$$\begin{aligned}
n_k v_k^{\text{cluster}} &= \sum_{i=1}^{n_k}\left(\frac{u_{k,i}^2(1)}{\mu_k} + \frac{u_{k,i}^2(0)}{1-\mu_k}\right) \\
&\quad+ p_k\sum_{m=1}^{m_k}\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}\big(u_{k,i}(1) - u_{k,i}(0)\big)\right)^2 - \sum_{i=1}^{n_k} 1\{m_{k,i} = m\}\big(u_{k,i}(1) - u_{k,i}(0)\big)^2\right] \\
&\quad+ p_k\sigma_k^2\sum_{m=1}^{m_k}\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}\left(\frac{u_{k,i}(1)}{\mu_k} + \frac{u_{k,i}(0)}{1-\mu_k}\right)\right)^2 - \sum_{i=1}^{n_k} 1\{m_{k,i} = m\}\left(\frac{u_{k,i}(1)}{\mu_k} + \frac{u_{k,i}(0)}{1-\mu_k}\right)^2\right].
\end{aligned}$$

Alternatively, we can write

$$\begin{aligned}
n_k v_k^{\text{cluster}} &= \sum_{i=1}^{n_k}\left(\frac{u_{k,i}^2(1)}{\mu_k} + \frac{u_{k,i}^2(0)}{1-\mu_k}\right) \\
&\quad - p_k \sum_{i=1}^{n_k}\big(u_{k,i}(1) - u_{k,i}(0)\big)^2 - p_k\sigma_k^2 \sum_{i=1}^{n_k}\left(\frac{u_{k,i}(1)}{\mu_k} + \frac{u_{k,i}(0)}{1-\mu_k}\right)^2 \\
&\quad + p_k \sum_{m=1}^{m_k}\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}\big(u_{k,i}(1) - u_{k,i}(0)\big)\right)^2 \\
&\quad + p_k\sigma_k^2 \sum_{m=1}^{m_k}\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}\left(\frac{u_{k,i}(1)}{\mu_k} + \frac{u_{k,i}(0)}{1-\mu_k}\right)\right)^2.
\end{aligned}$$

We will next show that

$$z' H_k \frac{\overline{\Sigma}_k}{n_k p_k q_k v_k} H_k z - \frac{v_k^{\text{cluster}}}{v_k} \xrightarrow{p} 0.$$

Given the $\mu_k(1-\mu_k)$ is bounded away from zero, by the weak law of large numbers for arrays, it is enough to show

$$\frac{1}{(n_k p_k q_k v_k)^2} \sum_{m=1}^{m_k} E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i}(W_{k,i} - \mu_k) U_{k,i}\right)^4\right] \longrightarrow 0.$$

Applying the multinomial theorem and the fact that all moments of $W_{k,i}$ as well as all potential outcomes are bounded, we obtain:

$$\begin{aligned}
&\frac{1}{(n_k p_k q_k v_k)^2} \sum_{m=1}^{m_k} E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i}(W_{k,i} - \mu_k) U_{k,i}\right)^4\right] \\
&\qquad \leqslant \frac{c}{(n_k p_k q_k v_k)^2}\Big(n_k p_k q_k + n_k p_k^2 q_k \max_m n_{k,m} + n_k p_k^3 q_k \max_m n_{k,m}^2 + n_k p_k^4 q_k \max_m n_{k,m}^3\Big).
\end{aligned}$$

Now, using $\limsup_{k\to\infty} \max_m n_{k,m}/\min_m n_{k,m} < \infty$, $\limsup_{k\to\infty} p_k \min_m n_{k,m}/v_k < \infty$, and $q_k m_k \to \infty$ we obtain

$$\begin{aligned}
&\frac{1}{(n_k p_k q_k v_k)^2} \sum_{m=1}^{m_k} E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i}(W_{k,i} - \mu_k) U_{k,i}\right)^4\right] \\
&\qquad \leqslant c\left(\frac{1}{n_k p_k q_k v_k^2} + \frac{\max_m n_{k,m}}{\min_m n_{k,m}}\frac{1}{q_k m_k v_k^2} + \frac{p_k \max_m n_{k,m}^2}{v_k \min_m n_{k,m}}\frac{1}{q_k m_k v_k} + \frac{p_k^2 \max_m n_{k,m}^3}{v_k^2 \min_m n_{k,m}}\frac{1}{q_k m_k}\right) \\
&\qquad \longrightarrow 0.
\end{aligned}$$

As a result,

$$\frac{\widehat{V}_k^{\text{cluster}}}{v_k} = \frac{v_k^{\text{cluster}}}{v_k} + o_p(1).$$

The robust (sandwich) estimator of the variance of $\sqrt{N_k}(\widehat{\tau}_k - \tau_k)$ is given by

$$\widehat{V}_k^{\text{robust}} = N_k z' \widehat{Q}_k^{-1} \widehat{\Omega}_k \widehat{Q}_k^{-1} z.$$

where

$$\widehat{\Omega}_k = \sum_{i=1}^{n_k} R_{k,i} \begin{pmatrix} \widehat{U}_{k,i} \\ W_{k,i}\widehat{U}_{k,i} \end{pmatrix} \begin{pmatrix} \widehat{U}_{k,i} \\ W_{k,i}\widehat{U}_{k,i} \end{pmatrix}'.$$

We will derive the limit of $\widehat{V}_k^{\text{robust}}/v_k$. Let

$$\overline{\Omega}_k = \sum_{i=1}^{n_k} R_{k,i} \begin{pmatrix} U_{k,i} \\ W_{k,i}U_{k,i} \end{pmatrix} \begin{pmatrix} U_{k,i} \\ W_{k,i}U_{k,i} \end{pmatrix}'.$$

Because potential outcomes (and $W_{k,i}$) are bounded, we obtain

$$\frac{1}{n_k p_k q_k v_k}\|\widehat{\Omega}_k - \overline{\Omega}_k\| \leqslant c\left(\frac{1}{n_k p_k q_k v_k}\sum_{i=1}^{n_k} R_{k,i}\right)\max_{i=1,\dots,n_k}|\widehat{U}_{k,i}^2 - U_{k,i}^2|.$$

Because the limsup of the expectation of the first factor (which is non-negative) is bounded and the second factor converges to zero in probability as proved above, we obtain

$$\frac{1}{n_k p_k q_k v_k}\|\widehat{\Omega}_k - \overline{\Omega}_k\| = o_p(1).$$

Notice that

$$\frac{1}{n_k p_k q_k v_k}\|\overline{\Omega}_k\| \leqslant c\left(\frac{1}{n_k p_k q_k v_k}\sum_{i=1}^{n_k} R_{k,i}\right)$$

Again, the limsup of the expectation of the right-hand side of this equation is non-negative and bounded. As a result, we obtain $\|\overline{\Omega}_k\|/(n_k p_k q_k v_k) = \mathcal{O}_p(1)$.

$$\begin{aligned}
\widehat{V}_k^{\text{robust}}/v_k &= z' H_k \frac{\overline{\Omega}_k}{n_k p_k q_k v_k} H_k z + o_p(1) \\
&= \frac{1}{n_k p_k q_k v_k}\left(\frac{1}{\mu_k(1-\mu_k)}\right)^2 \sum_{i=1}^{n_k} R_{k,i}(W_{k,i}-\mu_k)^2 U_{k,i}^2 + o_p(1).
\end{aligned}$$

Notice that

$$E\left[\sum_{i=1}^{n_k} R_{k,i}(W_{k,i}-\mu_k)^2 U_{k,i}^2\right] = \sum_{i=1}^{n_k} p_k q_k \mu_k(1-\mu_k)\Big((1-\mu_k)u_{k,i}^2(1) + \mu_k u_{k,i}^2(0)\Big).$$

Finally, notice that

$$\begin{aligned}
\frac{1}{(n_k p_k q_k v_k)^2}\sum_{m=1}^{m_k} E\left[\left(\sum_{i=1}^{n_k} R_{k,i}(W_{k,i}-\mu_k)^2 U_{k,i}^2\right)^2\right] &\leqslant c\frac{n_k p_k q_k + n_k p_k^2 q_k \max_m n_{k,m}}{(n_k p_k q_k v_k)^2} \\
&\leqslant c\left(\frac{1}{n_k p_k q_k v_k^2} + \frac{\max_m n_{k,m}}{\min_m n_{k,m}}\frac{1}{q_k m_k v_k^2}\right) \\
&\longrightarrow 0.
\end{aligned}$$

Therefore, by the weak law of large numbers for arrays, we obtain

$$\frac{\widehat{V}_k^{\text{robust}}}{v_k} = \frac{v_k^{\text{robust}}}{v_k} + o_p(1),$$

where

$$v_k^{\text{robust}} = \frac{1}{n_k}\sum_{i=1}^{n_k}\left(\frac{u_{k,i}^2(1)}{\mu_k} + \frac{u_{k,i}^2(0)}{1-\mu_k}\right).$$

A.3. Fixed effects

A.3.1. Large k distribution

Let

$$\overline{N}_{k,m} = \sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}$$

and

$$\widehat{\tau}_k^{\text{fixed}} = \frac{\sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}Y_{k,i}(W_{k,i} - \overline{W}_{k,m})}{\sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}W_{k,i}(W_{k,i} - \overline{W}_{k,m})}, \tag{A.5}$$

where

$$\overline{W}_{k,m} = \frac{1}{\overline{N}_{k,m}\vee 1}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}W_{k,i}.$$

Notice that we need $\liminf_{k\to\infty}\mu_k(1-\mu_k) - \sigma_k^2 = \liminf_{k\to\infty} E[A_{k,m}(1-A_{k,m})] > 0$ for this estimator to be well-defined in large samples (otherwise, the denominator in the formula for $\widehat{\tau}_k^{\text{fixed}}$ could be equal to zero). Although it is not strictly necessary, and because it entails little loss of generality and simplifies the exposition, we will assume that the supports of the cluster probabilities, $A_{k,m}$, are bounded away from zero and one (uniformly in k and m). In finite samples we assign $\widehat{\tau}_k^{\text{fixed}} = 0$ to the cases when the denominator of $\widehat{\tau}_k^{\text{fixed}}$ in equation (A.5) is equal to zero. Notice that

$$\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}(W_{k,i} - \overline{W}_{k,m}) = 0.$$

Let

$$\alpha_{k,m} = \frac{1}{n_{k,m}}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}y_{k,i}(0), \quad \tau_{k,m} = \frac{1}{n_{k,m}}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}(y_{k,i}(1) - y_{k,i}(0)),$$

$e_{k,i}(0) = y_{k,i}(0) - \alpha_{k,m_{k,i}}$, and $e_{k,i}(1) = y_{k,i}(1) - \alpha_{k,m_{k,i}} - \tau_{k,m_{k,i}}$. It follows that

$$\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}e_{k,i}(1) = \sum_{i=1}^{n_k} 1\{m_{k,i} = m\}e_{k,i}(0) = 0.$$

Now, $Y_{k,i} = e_{k,i}(1)W_{k,i} + e_{k,i}(0)(1 - W_{k,i}) + \alpha_{k,m_{k,i}} + \tau_{k,m_{k,i}} W_{k,i}$. Then,

$$\widehat{\tau}_k^{\text{fixed}} = \frac{\displaystyle\sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i}((e_{k,i}(1) + \tau_{k,m})W_{k,i} + e_{k,i}(0)(1 - W_{k,i}))(W_{k,i} - \overline{W}_{k,m})}{\displaystyle\sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}(W_{k,i} - \overline{W}_{k,m})}.$$

Let

$$\overline{\tau}_k = \frac{\displaystyle\sum_{m=1}^{m_k} \tau_{k,m} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}(W_{k,i} - \overline{W}_{k,m})}{\displaystyle\sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}(W_{k,i} - \overline{W}_{k,m})}, \tag{A.6}$$

where, as before, we make $\overline{\tau}_k = 0$ if the denominator on the right-hand side of (A.6) is equal to zero. Now, $\widehat{\tau}_k^{\text{fixed}} - \tau_k = (\widehat{\tau}_k^{\text{fixed}} - \overline{\tau}_k) + (\overline{\tau}_k - \tau_k)$, where

$$\widehat{\tau}_k^{\text{fixed}} - \overline{\tau}_k = \frac{\displaystyle\sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i}(e_{k,i}(1)W_{k,i} + e_{k,i}(0)(1 - W_{k,i}))(W_{k,i} - \overline{W}_{k,m})}{\displaystyle\sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}(W_{k,i} - \overline{W}_{k,m})}$$

and

$$\overline{\tau}_k - \tau_k = \frac{\displaystyle\sum_{m=1}^{m_k} (\tau_{k,m} - \tau_k) \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}(W_{k,i} - \overline{W}_{k,m})}{\displaystyle\sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}(W_{k,i} - \overline{W}_{k,m})}.$$

Notice that outcomes enter the term $\widehat{\tau}_k^{\text{fixed}} - \overline{\tau}_k$ only through the intra-cluster errors, $e_{k,i}(1)$ and $e_{k,i}(0)$. In contrast, the term $\overline{\tau}_k - \tau_k$ depends on outcomes only through inter-cluster variability in treatment effects, $\tau_{k,m} - \tau_k$. The numerator in the expression for $\overline{\tau}_k - \tau_k$ in the last displayed equation does not have mean zero in general, and this will be reflected in a bias term, B_k, which we define next. Let,

$$D_k = \frac{1}{n_k p_k q_k} \sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}(W_{k,i} - \overline{W}_{k,m}),$$

and

$$B_k = -\frac{\displaystyle\frac{1}{n_k p_k} E[A_{k,m}(1 - A_{k,m})] \sum_{m=1}^{m_k} (\tau_{k,m} - \tau_k)(1 - (1 - p_k)^{n_{k,m}})}{\displaystyle\frac{1}{n_k p_k q_k} \sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}(W_{k,i} - \overline{W}_{k,m})}.$$

Then, $\sqrt{n_k p_k q_k}(\widehat{\tau}_k^{\text{fixed}} - \tau_k - B_k) = F_k / D_k$, where

$$F_k = \sum_{m=1}^{m_k} (\psi_{k,m} - \overline{\psi}_{k,m}) + (\varphi_{k,m} - \overline{\varphi}_{k,m}),$$

$$\psi_{k,m} = \frac{1}{\sqrt{n_k p_k q_k}} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i}(e_{k,i}(1)W_{k,i} + e_{k,i}(0)(1 - W_{k,i}))(W_{k,i} - A_{k,m}),$$

$$\overline{\psi}_{k,m} = \frac{1}{\sqrt{n_k p_k q_k}} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i}(e_{k,i}(1)W_{k,i} + e_{k,i}(0)(1 - W_{k,i}))(\overline{W}_{k,m} - A_{k,m}),$$

$$\varphi_{k,m} = \frac{1}{\sqrt{n_k p_k q_k}} (\tau_{k,m} - \tau_k) \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} \big(R_{k,i} W_{k,i}(W_{k,i} - A_{k,m}) - p_k q_k E[A_{k,m}(1 - A_{k,m})] \big),$$

and

$$\overline{\varphi}_{k,m} = \frac{1}{\sqrt{n_k p_k q_k}} (\tau_{k,m} - \tau_k) \Bigg(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}(\overline{W}_{k,m} - A_{k,m}) - q_k E[A_{k,m}(1 - A_{k,m})](1 - (1 - p_k)^{n_{k,m}}) \Bigg).$$

The terms $\psi_{k,m}$ and $\overline{\psi}_{k,m}$ depend on the within-cluster errors $e_{k,i}(1)$ and $e_{k,i}(0)$. The terms $\varphi_{k,m}$ and $\overline{\varphi}_{k,m}$ depend on the inter-clusters errors $\tau_{k,m} - \tau_k$. $\psi_{k,m}$ and $\varphi_{k,m}$ replace $\overline{W}_{k,m}$ with $A_{k,m}$, while $\overline{\psi}_{k,m}$ and $\overline{\varphi}_{k,m}$ correct for the difference, $\overline{W}_{k,m} - A_{k,m}$.

It can be seen (in intermediate calculations below) that

$$E\left[\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}(W_{k,i} - A_{k,m}) \right] = n_{k,m} p_k q_k E[A_{k,m}(1 - A_{k,m})]$$

and

$$E\left[\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}(\overline{W}_{k,m} - A_{k,m}) \right] = q_k E[A_{k,m}(1 - A_{k,m})](1 - (1 - p_k)^{n_{k,m}}).$$

These two expectations are substracted in $\varphi_{k,m}$ and $\overline{\varphi}_{k,m}$, so $\varphi_{k,m}$ and $\overline{\varphi}_{k,m}$ have mean zero. Doing so for $\varphi_{k,m}$ does not require adjustments elsewhere. Because

$$\sum_{m=1}^{m_k} (\tau_{k,m} - \tau_k) n_{k,m} = 0,$$

the $n_{k,m} p_k q_k E[A_{k,m}(1 - A_{k,m})]$ terms do not change the sum F_k. In contrast, demeaning $\overline{\varphi}_{k,m}$ creates the bias term B_k. If the size of the clusters $n_{k,m}$ does not vary across clusters, then B_k is equal to zero. More generally, $\sqrt{n_k p_k q_k} D_k B_k = \mathcal{O}(m_k \sqrt{q_k/(n_k p_k)})$. Therefore, if

$$\frac{m_k q_k}{p_k(n_k/m_k)} \longrightarrow 0, \tag{A.7}$$

(that is, if the expected number of sampled clusters is small relative to the expected number of sampled observations per sampled cluster) then $\sqrt{n_k p_k q_k} D_k B_k$ converges to zero. As a result, $\sqrt{n_k p_k q_k} B_k$ converges in probability to zero, because, as we will show later, D_k converges in probability to $\mu_k(1 - \mu_k) - \sigma_k^2$, which is bounded away from zero. In our large sample analysis, we will assume that the expected number of sampled clusters grows to infinity, $m_k q_k \to \infty$. Then, equation (A.7) implies that the expected number of observations per sampled cluster goes to infinity, $p_k(n_k/m_k) \to \infty$. Notice also that $n_k p_k q_k = (n_k p_k/m_k)(m_k q_k) \to \infty$.

We summarize now the assumptions we made thus far. We first assumed that the supports of the cluster probabilities, $A_{k,m}$, are bounded away from zero and one (uniformly in k and m), and that potential outcomes are bounded. Moreover, we assumed $m_k q_k \to \infty$ and $(m_k q_k)/((p_k n_k)/m_k) \to 0$. These imply $(p_k n_k)/m_k \to \infty$ and $n_k p_k q_k \to \infty$. We will add the assumption that the ratio between maximum and minimum cluster size is bounded, $\limsup_{k\to\infty} \max_m n_{k,m}/\min_m n_{k,m} < \infty$. This assumption implies $p_k \min_m n_{k,m} \to \infty$ and $(m_k q_k)/(p_k \min_m n_{k,m}) \to 0$.

We will now study the behavior of D_k. Notice that

$$\begin{aligned} E&\left[\sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}R_{k,i}W_{k,i}(W_{k,i}-\overline{W}_{k,m})\right] \\ &= E\left[\sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}R_{k,i}W_{k,i}(W_{k,i}-A_{k,m})\right] \\ &\quad - E\left[\sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}R_{k,i}W_{k,i}(\overline{W}_{k,m}-A_{k,m})\right] \\ &= n_k p_k q_k E[A_{k,m}(1-A_{k,m})] - q_k E[A_{k,m}(1-A_{k,m})]\sum_{m=1}^{m_k}(1-(1-p_k)^{n_{k,m}}). \end{aligned}$$

In addition,

$$\begin{aligned} \frac{1}{(n_k p_k q_k)^2}\sum_{m=1}^{m_k} E&\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}R_{k,i}W_{k,i}(W_{k,i}-\overline{W}_{k,m})\right)^2\right] \\ &\leqslant c\,\frac{n_k p_k q_k + n_k p_k^2 q_k \max_m n_{k,m}}{(n_k p_k q_k)^2} \\ &= c\left(\frac{1}{n_k p_k q_k} + \frac{\max_m n_{k,m}}{\min_m n_{k,m}}\frac{1}{m_k q_k}\right) \longrightarrow 0. \end{aligned}$$

The weak law of large numbers for arrays implies

$$D_k - E[A_{k,m}(1-A_{k,m})] + \frac{1}{n_k p_k}E[A_{k,m}(1-A_{k,m})]\sum_{m=1}^{m_k}(1-(1-p_k)^{n_{k,m}}) \xrightarrow{p} 0.$$

Because $m_k/(n_k p_k) \to 0$ and $E[A_{k,m}(1-A_{k,m})] = \mu_k(1-\mu_k) - \sigma_k^2$, we obtain

$$D_k - (\mu_k(1-\mu_k) - \sigma_k^2) \xrightarrow{p} 0.$$

We now turn our attention to F_k. We will first calculate the variance of $\psi_{k,m}$. Let $Q_{k,m}$ be a binary variable that takes value one if cluster m in population k is sampled, and zero otherwise. Notice that

$$E[R_{k,i}W_{k,i}(W_{k,i}-A_{k,m})|A_{k,m}, Q_{k,m}=1, m_{k,i}=m] = p_k A_{k,m}(1-A_{k,m}),$$

and

$$E[R_{k,i}(1-W_{k,i})(W_{k,i}-A_{k,m})|A_{k,m}, Q_{k,m}=1, m_{k,i}=m] = -p_k A_{k,m}(1-A_{k,m}).$$

Consider now

$$\begin{aligned}\psi_{k,m,1} &= \frac{1}{\sqrt{n_k p_k q_k}} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}(W_{k,i} - A_{k,m}) e_{k,i}(1) \\ &= \frac{Q_{k,m}}{\sqrt{n_k p_k q_k}} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} \Big(R_{k,i} W_{k,i}(W_{k,i} - A_{k,m}) - p_k A_{k,m}(1 - A_{k,m}) \Big) e_{k,i}(1),\end{aligned}$$

and

$$\begin{aligned}\psi_{k,m,0} &= \frac{1}{\sqrt{n_k p_k q_k}} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i}(1 - W_{k,i})(W_{k,i} - A_{k,m}) e_{k,i}(0) \\ &= \frac{Q_{k,m}}{\sqrt{n_k p_k q_k}} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} \Big(R_{k,i}(1 - W_{k,i})(W_{k,i} - A_{k,m}) + p_k A_{k,m}(1 - A_{k,m}) \Big) e_{k,i}(0).\end{aligned}$$

It holds that $\psi_{k,m} = \psi_{k,m,1} + \psi_{k,m,0}$ and $E[\psi_{k,m}] = 0$. Now, notice that

$$\begin{aligned}E[\psi^2_{k,m,1}] &= \frac{1}{n_k} E[A_{k,m}(1 - A_{k,m})^2 - p_k A^2_{k,m}(1 - A_{k,m})^2] \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} e^2_{k,i}(1), \\ E[\psi^2_{k,m,0}] &= \frac{1}{n_k} E[A^2_{k,m}(1 - A_{k,m}) - p_k A^2_{k,m}(1 - A_{k,m})^2] \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} e^2_{k,i}(0),\end{aligned}$$

and

$$E[\psi_{k,m,1}\psi_{k,m,0}] = \frac{1}{n_k} p_k E[A^2_{k,m}(1 - A_{k,m})^2] \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} e_{k,i}(1) e_{k,i}(0).$$

Therefore,

$$\begin{aligned}E[(\psi_{k,m,1} + \psi_{k,m,0})^2] = & \frac{1}{n_k} E[A_{k,m}(1 - A_{k,m})^2] \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} e^2_{k,i}(1) \\ & + \frac{1}{n_k} E[A^2_{k,m}(1 - A_{k,m})] \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} e^2_{k,i}(0) \\ & - \frac{1}{n_k} p_k E[A^2_{k,m}(1 - A_{k,m})^2] \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} (e_{k,i}(1) - e_{k,i}(0))^2,\end{aligned}$$

and

$$\begin{aligned}\sum_{m=1}^{m_k} E[(\psi_{k,m,1} + \psi_{k,m,0})^2] = & E[A_{k,m}(1 - A_{k,m})^2] \frac{1}{n_k} \sum_{i=1}^{n_k} e^2_{k,i}(1) \\ & + E[A^2_{k,m}(1 - A_{k,m})] \frac{1}{n_k} \sum_{i=1}^{n_k} e^2_{k,i}(0) \\ & - p_k E[A^2_{k,m}(1 - A_{k,m})^2] \frac{1}{n_k} \sum_{i=1}^{n_k} (e_{k,i}(1) - e_{k,i}(0))^2. \qquad \text{(A.8)}\end{aligned}$$

We will next show that the terms $\overline{\psi}_{k,m}$ do not matter for the asymptotic distribution of $\sqrt{n_k p_k q_k}(\widehat{\tau}_k - \tau_k)$. Notice that, because the cluster sum of $e_{k,i}(1)$ is equal to zero, we obtain $E[\overline{\psi}_{k,m}] = 0$ and, therefore,

$$\sum_{m=1}^{m_k} E\Big[\overline{\psi}_{k,m}\Big] = 0.$$

Moreover

$$2\sum_{i=1}^{n_k-1}\sum_{j=i+1}^{n_k} 1\{m_{k,i} = m_{k,j} = m\}e_{k,i}(1)e_{k,j}(1) = -\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}e_{k,i}^2(1) \leqslant 0.$$

In addition, $E[R_{k,i}W_{k,i}(\overline{W}_{k,m} - A_{k,m})^2|m_{k,i} = m] \leqslant q_k E[A_{k,m}(1 - A_{k,m})]/n_{k,m}$ (see intermediate calculations). Therefore,

$$\begin{aligned}
&E\Bigg[\bigg(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}W_{k,i}(\overline{W}_{k,m} - A_{k,m})e_{k,i}(1)\bigg)^2\Bigg] \\
&\quad = \sum_{i=1}^{n_k} 1\{m_{k,i} = m\}E\Big[R_{k,i}W_{k,i}(\overline{W}_{k,m} - A_{k,m})^2|m_{k,i} = m\Big]e_{k,i}^2(1) \\
&\quad + 2\sum_{i=1}^{n_k-1}\sum_{j=i+1}^{n_k} E\Big[1\{m_{k,i} = m_{k,j} = m\}R_{k,i}R_{k,j}W_{k,i}W_{k,j}(\overline{W}_{k,m} - A_{k,m})^2\Big]e_{k,i}(1)e_{k,j}(1) \\
&\quad \leqslant q_k E[A_{k,m}(1 - A_{k,m})]\frac{1}{n_{k,m}}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}e_{k,i}^2(1).
\end{aligned}$$

Now, because errors are bounded, we obtain

$$\sum_{m=1}^{m_k} E\Bigg[\bigg(\frac{1}{\sqrt{n_k p_k q_k}}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}W_{k,i}(\overline{W}_{k,m} - A_{k,m})e_{k,i}(1)\bigg)^2\Bigg] \leqslant c\,\frac{m_k}{n_k p_k}. \tag{A.9}$$

Because $m_k/(n_k p_k) \to 0$, the weak law of large numbers for arrays, implies,

$$\frac{1}{\sqrt{n_k p_k q_k}}\sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}W_{k,i}(\overline{W}_{k,m} - A_{k,m})e_{k,i}(1) \xrightarrow{p} 0.$$

with the analogous result involving the errors $e_{k,i}(0)$. If follows that

$$\sum_{m=1}^{m_k} \overline{\psi}_{k,m} \xrightarrow{p} 0.$$

Consider now $\varphi_{k,m}$. Notice that

$$\begin{aligned}
&E\Big[\Big(R_{k,i}W_{k,i}(W_{k,i} - A_{k,m}) - p_k q_k E[A_{k,m}(1 - A_{k,m})]\Big)\Big)^2\Big] \\
&\qquad = p_k q_k E[A_{k,m}(1 - A_{k,m})^2] - p_k^2 q_k^2\big(E[A_{k,m}(1 - A_{k,m})]\big)^2,
\end{aligned}$$

and

$$
\begin{aligned}
E\Big[&\Big(R_{k,i}W_{k,i}(W_{k,i}-A_{k,m})-p_kq_kE[A_{k,m}(1-A_{k,m})]\Big)\Big)\\
&\qquad\qquad\times\Big(R_{k,j}W_{k,j}(W_{k,j}-A_{k,m})-p_kq_kE[A_{k,m}(1-A_{k,m})]\Big)\Big)\big|m_{k,i}=m_{k,j}=m\Big]\\
&\qquad\qquad\quad=p_k^2q_kE[A_{k,m}^2(1-A_{k,m})^2]-p_k^2q_k^2\big(E[A_{k,m}(1-A_{k,m})]\big)^2.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
E[\varphi_{k,m}^2]=&\Big(E[A_{k,m}(1-A_{k,m})^2]-p_kq_k(E[A_{k,m}(1-A_{k,m})])^2\Big)\frac{n_{k,m}}{n_k}(\tau_{k,m}-\tau_k)^2\\
&+\Big(p_kE[A_{k,m}^2(1-A_{k,m})^2]-p_kq_k(E[A_{k,m}(1-A_{k,m})])^2\Big)\frac{n_{k,m}(n_{k,m}-1)}{n_k}(\tau_{k,m}-\tau_k)^2,
\end{aligned}
$$

and

$$
\begin{aligned}
\sum_{m=1}^{m_k}E[\varphi_{k,m}^2]=&\Big(E[A_{k,m}(1-A_{k,m})^2]-p_kq_k(E[A_{k,m}(1-A_{k,m})])^2\Big)\sum_{m=1}^{m_k}\frac{n_{k,m}}{n_k}(\tau_{k,m}-\tau_k)^2\\
&+\Big(p_kE[A_{k,m}^2(1-A_{k,m})^2]-p_kq_k(E[A_{k,m}(1-A_{k,m})])^2\Big)\sum_{m=1}^{m_k}\frac{n_{k,m}(n_{k,m}-1)}{n_k}(\tau_{k,m}-\tau_k)^2.
\end{aligned}
$$

Next, we calculate the variance of $\overline{\varphi}_{k,m}$. Using results on the moments of a Binomial distribution, we obtain, for $n\geqslant 1$,

$$
\begin{aligned}
E&\left[\left(\sum_{i=1}^{n_k}1\{m_{k,i}=m\}R_{k,i}W_{k,i}(\overline{W}_{k,m}-A_{k,m})\right)^2\Big|\ \begin{matrix}Q_{k,m}=1,\\ \overline{N}_{k,m}=n\end{matrix}\right]\\
&=\frac{1}{n^2}E\left[\left(\sum_{i=1}^{n_k}1\{m_{k,i}=m\}R_{k,i}W_{k,i}\Big(\sum_{i=1}^{n_k}1\{m_{k,i}=m\}R_{k,i}W_{k,i}-nA_{k,m}\Big)\right)^2\Big|\ \begin{matrix}Q_{k,m}=1,\\ \overline{N}_{k,m}=n\end{matrix}\right]\\
&=nE[A_{k,m}^3(1-A_{k,m})]+E[A_{k,m}^2(1-A_{k,m})(5-7A_{k,m})]\\
&+\frac{1}{n}E[A_{k,m}(1-A_{k,m})(6A_{k,m}^2-6A_{k,m}+1)].
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
E&\left[\left(\sum_{i=1}^{n_k}1\{m_{k,i}=m\}R_{k,i}W_{k,i}(\overline{W}_{k,m}-A_{k,m})\right)^2\right]\\
&=n_{k,m}p_kq_kE[A_{k,m}^3(1-A_{k,m})]+q_kE[A_{k,m}^2(1-A_{k,m})(5-7A_{k,m})](1-(1-p_k)^{n_{k,m}})\\
&+q_kE[A_{k,m}(1-A_{k,m})(6A_{k,m}^2-6A_{k,m}+1)]r_{k,m},
\end{aligned}
$$

where

$$
r_{k,m}=\sum_{n=1}^{n_{k,m}}\frac{1}{n}\Pr(\overline{N}_{k,m}=n|Q_{k,m}=1)\leqslant\sum_{n=1}^{n_{k,m}}\Pr(\overline{N}_{k,m}=n|Q_{k,m}=1)\leqslant 1.
$$

It follows that,

$$
E[\overline{\varphi}_{k,m}^2]=(\tau_{k,m}-\tau_k)^2\Big(\frac{n_{k,m}}{n_k}E[A_{k,m}^3(1-A_{k,m})]+\frac{1}{n_kp_k}E[A_{k,m}^2(1-A_{k,m})(5-7A_{k,m})](1-(1-p_k)^{n_{k,m}})
$$

$$+ \frac{1}{n_k p_k} E[A_{k,m}(1-A_{k,m})(6A_{k,m}^2 - 6A_{k,m}+1)] r_{k,m}$$
$$- \frac{q_k}{n_k p_k} (E[A_{k,m}(1-A_{k,m})])^2 (1-(1-p_k)^{n_{k,m}})^2 \Big).$$

Therefore,

$$\sum_{m=1}^{m_k} E[\overline{\varphi}_{k,m}^2] = \sum_{m=1}^{m_k} (\tau_{k,m} - \tau_k)^2 \Big(\frac{n_{k,m}}{n_k}\Big) E[A_{k,m}^3(1-A_{k,m})] + o(1).$$

We will now study the covariance between $\varphi_{k,m}$ and $\overline{\varphi}_{k,m}$. Using results on the moments of a Binomial distribution, we obtain, for $n \geqslant 1$,

$$E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i}=m\} R_{k,i} W_{k,i}(W_{k,i} - A_{k,m})\right)\left(\sum_{i=1}^{n_k} 1\{m_{k,i}=m\} R_{k,i} W_{k,i}(\overline{W}_{k,m} - A_{k,m})\right) \Big| \begin{array}{l} Q_{k,m}=1, \\ \overline{N}_{k,m}=n \end{array}\right]$$
$$= E\left[\frac{1-A_{k,m}}{n}\left(\sum_{i=1}^{n_k} 1\{m_{k,i}=m\} R_{k,i} W_{k,i}\right)^2 \left(\sum_{i=1}^{n_k} 1\{m_{k,i}=m\} R_{k,i} W_{k,i} - nA_{k,m}\right) \Big| \begin{array}{l} Q_{k,m}=1, \\ \overline{N}_{k,m}=n \end{array}\right]$$
$$= 2nE[A_{k,m}^2(1-A_{k,m})^2] + E[A_{k,m}(1-A_{k,m})^2(1-2A_{k,m})].$$

Therefore,

$$E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i}=m\} R_{k,i} W_{k,i}(W_{k,i} - A_{k,m})\right)\left(\sum_{i=1}^{n_k} 1\{m_{k,i}=m\} R_{k,i} W_{k,i}(\overline{W}_{k,m} - A_{k,m})\right)\right]$$
$$= 2n_{k,m} p_k q_k E[A_{k,m}^2(1-A_{k,m})^2] + q_k E[A_{k,m}(1-A_{k,m})^2(1-2A_{k,m})] \Pr(\overline{N}_{k,m} \geqslant 1 | Q_{k,m}=1).$$

In addition,

$$E\left[\sum_{i=1}^{n_k} 1\{m_{k,i}=m\} R_{k,i} W_{k,i}(W_{k,i} - A_{k,m})\right] E\left[\sum_{i=1}^{n_k} 1\{m_{k,i}=m\} R_{k,i} W_{k,i}(\overline{W}_{k,m} - A_{k,m})\right]$$
$$= n_{k,m} p_k q_k^2 (E[A_{k,m}(1-A_{k,m})])^2 \Pr(\overline{N}_{k,m} \geqslant 1 | Q_{k,m}=1).$$

As a result,

$$E[\varphi_{k,m}\overline{\varphi}_{k,m}] = \Big(2E[A_{k,m}^2(1-A_{k,m})^2] - q_k (E[A_{k,m}(1-A_{k,m})])^2\Big)(\tau_{k,m}-\tau_k)^2 \left(\frac{n_{k,m}}{n_k}\right) + \mathcal{O}\left(\frac{1}{n_k p_k}\right)$$
$$+ \mathcal{O}\left(\frac{q_k}{p_k \min_m n_{k,m}} \big(p_k \min_m n_{k,m}(1-p_k)^{\min_m n_{k,m}}\big)\right).$$

Notice that $m_k/(n_k p_k) \to 0$. In addition, $m_k q_k/(p_k \min_m n_{k,m}) \to 0$ and

$$p_k \min_m n_{k,m}(1-p_k)^{\min_m n_{k,m}} = p_k \min_m n_{k,m} \left(1 - \frac{p_k \min_m n_{k,m}}{\min_m n_{k,m}}\right)^{\min_m n_{k,m}}$$
$$< p_k \min_m n_{k,m} e^{-p_k \min_m n_{k,m}} \longrightarrow 0.$$

Therefore,

$$\sum_{m=1}^{m_k} E[\varphi_{k,m}\overline{\varphi}_{k,m}] = \Big(2E[A_{k,m}^2(1-A_{k,m})^2] - q_k(E[A_{k,m}(1-A_{k,m})])^2\Big) \sum_{m=1}^{m_k} (\tau_{k,m}-\tau_k)^2 \left(\frac{n_{k,m}}{n_k}\right) + o(1).$$

Next, we will study the remaining covariances between $\psi_{k,m}$, $\varphi_{k,m}$, $\overline{\psi}_{k,m}$, and $\overline{\varphi}_{k,m}$. Because the intra-cluster errors, $e_{k,i}(1)$ and $e_{k,i}(0)$ sum to zero, it can be easily seen that $E[\psi_{k,m}\varphi_{k,m}] = E[\psi_{k,m}\overline{\varphi}_{k,m}] = 0$. It can also be seen that the inter-clusters sums of covariances between $\overline{\psi}_{k,m}$ and any of the other terms go to zero. To prove this for the covariance with $\psi_{k,m}$, we have

$$\begin{aligned}\left(\sum_{m=1}^{m_k} E[|\psi_{k,m}\overline{\psi}_{k,m}|]\right)^2 &\leqslant \left(\sum_{m=1}^{m_k} (E[\psi_{k,m}^2]E[\overline{\psi}_{k,m}^2])^{1/2}\right)^2 \\ &\leqslant \sum_{m=1}^{m_k} E[\psi_{k,m}^2] \sum_{m=1}^{m_k} E[\overline{\psi}_{k,m}^2] \\ &= \mathcal{O}(1)\mathcal{o}(1) = \mathcal{o}(1).\end{aligned}$$

The same argument and result applies to $E[\overline{\psi}_{k,m}\varphi_{k,m}]$ and $E[\overline{\psi}_{k,m}\overline{\varphi}_{k,m}]$. Putting all the pieces together, we obtain

$$n_k p_k q_k E[D_k^2(\widehat{\tau}_k^{\text{fixed}} - \tau_k)^2] = f_k + \mathcal{o}(1),$$

where

$$\begin{aligned}f_k &= E[A_{k,m}(1-A_{k,m})^2]\frac{1}{n_k}\sum_{i=1}^{n_k} e_{k,i}^2(1) + E[A_{k,m}^2(1-A_{k,m})]\frac{1}{n_k}\sum_{i=1}^{n_k} e_{k,i}^2(0) \\ &- p_k E[A_{k,m}^2(1-A_{k,m})^2]\frac{1}{n_k}\sum_{i=1}^{n_k}(e_{k,i}(1)-e_{k,i}(0))^2 \\ &+ \Big(E[A_{k,m}(1-A_{k,m})^2] - p_k q_k (E[A_{k,m}(1-A_{k,m})])^2\Big)\sum_{m=1}^{m_k}\frac{n_{k,m}}{n_k}(\tau_{k,m}-\tau_k)^2 \\ &+ \Big(p_k E[A_{k,m}^2(1-A_{k,m})^2] - p_k q_k (E[A_{k,m}(1-A_{k,m})])^2\Big)\sum_{m=1}^{m_k}\frac{n_{k,m}(n_{k,m}-1)}{n_k}(\tau_{k,m}-\tau_k)^2 \\ &+ E[A_{k,m}^3(1-A_{k,m})]\sum_{m=1}^{m_k}(\tau_{k,m}-\tau_k)^2\Big(\frac{n_{k,m}}{n_k}\Big) \\ &- 2\Big(2E[A_{k,m}^2(1-A_{k,m})^2] - q_k (E[A_{k,m}(1-A_{k,m})])^2\Big)\sum_{m=1}^{m_k}(\tau_{k,m}-\tau_k)^2\left(\frac{n_{k,m}}{n_k}\right).\end{aligned}$$

Collecting terms with identical factors, we obtain

$$\begin{aligned}f_k &= E[A_{k,m}(1-A_{k,m})^2]\frac{1}{n_k}\sum_{i=1}^{n_k} e_{k,i}^2(1) + E[A_{k,m}^2(1-A_{k,m})]\frac{1}{n_k}\sum_{i=1}^{n_k} e_{k,i}^2(0) \\ &- p_k E[A_{k,m}^2(1-A_{k,m})^2]\frac{1}{n_k}\sum_{i=1}^{n_k}(e_{k,i}(1)-e_{k,i}(0))^2 \\ &+ \Big(E[A_{k,m}(1-A_{k,m})^2] - (4+p_k)E[A_{k,m}^2(1-A_{k,m})^2] \\ &\qquad + E[A_{k,m}^3(1-A_{k,m})] + 2q_k (E[A_{k,m}(1-A_{k,m})])^2\Big)\sum_{m=1}^{m_k}\frac{n_{k,m}}{n_k}(\tau_{k,m}-\tau_k)^2\end{aligned}$$

$$+\Big(p_k E[A_{k,m}^2(1-A_{k,m})^2] - p_k q_k (E[A_{k,m}(1-A_{k,m})])^2\Big) \sum_{m=1}^{m_k} \frac{n_{k,m}^2}{n_k}(\tau_{k,m}-\tau_k)^2.$$

The first three terms in the expression above depend on intra-cluster heterogeneity in potential outcomes and treatment effects. The last two terms depend on inter-cluster variation in average treatment effects.

A more compact expression for f_k is

$$\begin{aligned}
f_k &= E[A_{k,m}(1-A_{k,m})^2]\frac{1}{n_k}\sum_{i=1}^{n_k} e_{k,i}^2(1) + E[A_{k,m}^2(1-A_{k,m})]\frac{1}{n_k}\sum_{i=1}^{n_k} e_{k,i}^2(0) \\
&\quad - p_k E[A_{k,m}^2(1-A_{k,m})^2]\frac{1}{n_k}\sum_{i=1}^{n_k}(e_{k,i}(1)-e_{k,i}(0))^2 \\
&\quad + \Big(E[A_{k,m}(1-A_{k,m})] - (5+p_k)E[A_{k,m}^2(1-A_{k,m})^2] \\
&\qquad\qquad + 2q_k(E[A_{k,m}(1-A_{k,m})])^2\Big)\sum_{m=1}^{m_k}\frac{n_{k,m}}{n_k}(\tau_{k,m}-\tau_k)^2 \\
&\quad + \Big(p_k E[A_{k,m}^2(1-A_{k,m})^2] - p_k q_k(E[A_{k,m}(1-A_{k,m})])^2\Big)\sum_{m=1}^{m_k}\frac{n_{k,m}^2}{n_k}(\tau_{k,m}-\tau_k)^2. \qquad \text{(A.10)}
\end{aligned}$$

Notice that the first four terms in (A.10) are bounded, and that

$$E[A_{k,m}^2(1-A_{k,m})^2] - q_k(E[A_{k,m}(1-A_{k,m})])^2 = \mathrm{var}(A_{k,m}(1-A_{k,m})) + (1-q_k)(E[A_{k,m}(1-A_{k,m})])^2.$$

Assume that

$$\liminf_{k\to\infty} \sum_{m=1}^{m_k}\frac{n_{k,m}}{n_k}(\tau_{k,m}-\tau_k)^2 > 0, \qquad \text{(A.11)}$$

and

$$\liminf_{k\to\infty} \mathrm{var}(A_{k,m}(1-A_{k,m})) \vee (1-q_k) > 0. \qquad \text{(A.12)}$$

The last term in equation (A.10) is greater than

$$p_k \min_m n_{k,m}\big(E[A_{k,m}^2(1-A_{k,m})^2] - (E[A_{k,m}(1-A_{k,m})])^2\big)\sum_{m=1}^{m_k}\frac{n_{k,m}}{n_k}(\tau_{k,m}-\tau_k)^2,$$

which converges to infinity because $p_k \min_m n_{k,m} \to \infty$. That is, the last term dominates the variance in large samples provided that (A.11) and (A.12) hold.

We will now derive the large sample distribution of $\widehat{\tau}_k^{\text{fixed}}$. To show that Lyapunov's condition holds for F_k, notice that

$$\begin{aligned}
&|(\psi_{k,m}-\overline{\psi}_{k,m}) + (\varphi_{k,m}-\overline{\varphi}_{k,m})|^3 \\
&\quad = \frac{1}{(n_k p_k q_k)^{3/2}}\Bigg|\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}R_{k,i}((e_{k,i}(1)+\tau_{k,m}-\tau_k)W_{k,i} + e_{k,i}(0)(1-W_{k,i}))(W_{k,i}-\overline{W}_{k,m}) \\
&\qquad\qquad - (\tau_{k,m}-\tau)q_k E[A_{k,m}(1-A_{k,m})](1-(1-p_k)^{n_{k,m}})\Bigg|^3,
\end{aligned}$$

where the last term inside the absolute value comes from the bias correction. Notice that,

$$\begin{aligned}
&\left|\sum_{i=1}^{n_k} 1\{m_{k,i}=m\} R_{k,i}(e_{k,i}(1)+\tau_{k,m}-\tau_k) W_{k,i}(W_{k,i}-\overline{W}_{k,m})\right|^3 \\
&\qquad\qquad = \left|(1-\overline{W}_{k,m}) \sum_{i=1}^{n_k} 1\{m_{k,i}=m\} R_{k,i}(e_{k,i}(1)+\tau_{k,m}-\tau_k) W_{k,i}\right|^3 \\
&\qquad\qquad \leqslant c \left|\sum_{i=1}^{n_k} 1\{m_{k,i}=m\} R_{k,i} W_{k,i}\right|^3 \\
&\qquad\qquad \leqslant c\,\overline{N}_{k,m}^3.
\end{aligned}$$

From the formula of the third moment of a binomial random variable, we obtain

$$\begin{aligned}
E[\overline{N}_{k,m}^3] &= q_k E[\overline{N}_{k,m}^3 | Q_{k,m}=1] \\
&= n_{k,m}^3 p_k^3 q_k + o(n_{k,m}^3 p_k^3 q_k),
\end{aligned}$$

as $p_k n_{k,m} \to \infty$. Now,

$$\begin{aligned}
&\frac{1}{f_k^{3/2}} \sum_{k=1}^{m_k} E\left[\left|\frac{1}{\sqrt{n_k p_k q_k}} \sum_{i=1}^{n_k} 1\{m_{k,i}=m\} R_{k,i}(e_{k,i}(1)+\tau_{k,m}-\tau_k) W_{k,i}(W_{k,i}-\overline{W}_{k,m})\right|^3\right] \\
&\qquad\leqslant c\,\frac{n_k \max_n n_{k,m}^2 p_k^3 q_k}{(n_k p_k q_k)^{3/2} (p_k \min_m n_{k,m})^{3/2}} = c\left(\frac{\max_m n_{k,m}}{\min_m n_{k,m}}\right)^2 \frac{1}{(m_k q_k)^{1/2}} \longrightarrow 0.
\end{aligned}$$

Similar calculations deliver the analogous result for the term involving $e_{k,i}(0)$, and proving the result for the bias term is straightforward. Therefore, we obtain

$$\frac{1}{f_k^{3/2}} \sum_{m=1}^{m_k} |(\psi_{k,m}-\overline{\psi}_{k,m}) + (\varphi_{k,m}-\overline{\varphi}_{k,m})|^3 \longrightarrow 0.$$

By the Central Limit Theorem for arrays, this implies

$$\sqrt{n_k p_k q_k} F_k / f_k^{1/2} \xrightarrow{d} N(0,1).$$

Let $\tilde{v}_k = f_k/(\mu_k(1-\mu_k)-\sigma_k^2)^2$. Then,

$$\sqrt{n_k p_k q_k}(\widehat{\tau}_k^{\text{fixed}} - \tau_k)/\tilde{v}_k^{1/2} \xrightarrow{d} N(0,1).$$

As a result,

$$\sqrt{N_k}(\widehat{\tau}_k^{\text{fixed}} - \tau_k)/\tilde{v}_k^{1/2} \xrightarrow{d} N(0,1).$$

A.3.2. Estimation of the variance

Let

$$N_{k,m,0} = \sum_{i=1}^{n_k} 1\{m_{k,i}=m\} R_{k,i}(1-W_{k,i})$$

and

$$N_{k,m,1} = \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}.$$

Let

$$\overline{Y}_{k,m} = \frac{1}{\overline{N}_{k,m} \vee 1} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} Y_{k,i}.$$

Then,

$$\overline{Y}_{k,m} = \widehat{\alpha}_{k,m} + \widehat{\tau}_{k,m} \overline{W}_{k,m},$$

where

$$\widehat{\alpha}_{k,m} = \frac{1}{N_{k,m,0} \vee 1} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i}(1 - W_{k,i}) Y_{k,i},$$

$$\widehat{\tau}_{k,m} = \frac{1}{N_{k,m,1} \vee 1} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i} Y_{k,i} - \frac{1}{N_{k,m,0} \vee 1} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i}(1 - W_{k,i}) Y_{k,i},$$

and, as before,

$$\overline{W}_{k,m} = \frac{1}{\overline{N}_{k,m} \vee 1} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}.$$

Let $\widetilde{U}_{k,i} = \widetilde{Y}_{k,i} - \widehat{\tau}_k^{\text{fixed}} \widetilde{W}_{k,i}$, where $\widetilde{Y}_{k,i} = Y_{k,i} - \overline{Y}_{k,m_{k,i}}$, $\widetilde{W}_{k,i} = (W_{k,i} - \overline{W}_{k,m_{k,i}})$, and $\widehat{\tau}_k^{\text{fixed}}$ is the within estimator of τ_k. Let $\widetilde{\Sigma}_k = \sum_{m=1}^{m_k} \widetilde{\Sigma}_{k,m}$, where

$$\widetilde{\Sigma}_{k,m} = \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} \widetilde{W}_{k,i} \widetilde{U}_{k,i} \right)^2.$$

Also, let

$$\widetilde{Q}_k = \sum_{i=1}^{n_k} R_{k,i} \widetilde{W}_{k,i}^2.$$

Then, the cluster estimator of the variance of $\sqrt{N_k}(\widehat{\tau}_k^{\text{fixed}} - \tau_k)$ is

$$\widetilde{V}_k^{\text{cluster}} = N_k \widetilde{Q}_k^{-1} \widetilde{\Sigma}_k \widetilde{Q}_k^{-1}.$$

We know already that

$$\frac{1}{n_k p_k q_k} \widetilde{Q}_k - (\mu_k(1 - \mu_k) - \sigma_k^2) \xrightarrow{p} 0,$$

with $\mu_k(1 - \mu_k) - \sigma_k^2$ bounded away from zero. To establish convergence of $\widetilde{\Sigma}_k / (n_k p_k q_k f_k)$, first notice that, for $m_{k,i} = m$, we have

$$\begin{aligned}
\widetilde{U}_{k,i} &= Y_{k,i} - (\widehat{\alpha}_{k,m} + \widehat{\tau}_{k,m} \overline{W}_{k,m}) - \widehat{\tau}_k^{\text{fixed}}(W_{k,i} - \overline{W}_{k,m}) \\
&= y_{k,i}(1) W_{k,i} + y_{k,i}(0)(1 - W_{k,i}) - (\alpha_{k,m} + \tau_{k,m} \overline{W}_{k,m}) - \widehat{\tau}_k^{\text{fixed}}(W_{k,i} - \overline{W}_{k,m}) \\
&\quad - (\widehat{\alpha}_{k,m} - \alpha_{k,m}) - (\widehat{\tau}_{k,m} - \tau_{k,m}) \overline{W}_{k,m} \\
&= e_{k,i}(1) W_{k,i} + e_{k,i}(0)(1 - W_{k,i}) + (\tau_{k,m} - \widehat{\tau}_k^{\text{fixed}})(W_{k,i} - \overline{W}_{k,m})
\end{aligned}$$

$$
\begin{aligned}
&- (\widehat{\alpha}_{k,m} - \alpha_{k,m}) - (\widehat{\tau}_{k,m} - \tau_{k,m})\overline{W}_{k,m} \\
&= e_{k,i}(1)W_{k,i} + e_{k,i}(0)(1 - W_{k,i}) + (\tau_{k,m} - \tau_k)(W_{k,i} - \overline{W}_{k,m}) \\
&- (\widehat{\tau}_k^{\text{fixed}} - \tau_k)(W_{k,i} - \overline{W}_{k,m}) - (\widehat{\alpha}_{k,m} - \alpha_{k,m}) - (\widehat{\tau}_{k,m} - \tau_{k,m})\overline{W}_{k,m}.
\end{aligned}
$$

For $m_{k,i} = m$ and $N_{k,m,0}, N_{k,m,1} \geqslant 1$, let

$$
\overline{U}_{k,i} = e_{k,i}(1)W_{k,i} + e_{k,i}(0)(1 - W_{k,i}) + (\tau_{k,m} - \tau_k)(W_{k,i} - \overline{W}_{k,m}),
$$

and let $\overline{U}_{k,i} = 0$ for $m_{k,i} = m$ and $N_{k,m,0}N_{k,m,1} = 0$. Then, for $m_{k,i} = m$ and $N_{k,m,0}N_{k,m,1} \geqslant 1$, we have

$$
\widetilde{U}_{k,i} - \overline{U}_{k,i} = -(\widehat{\tau}_k^{\text{fixed}} - \tau_k)(W_{k,i} - \overline{W}_{k,m}) - (\widehat{\alpha}_{k,m} - \alpha_{k,m}) - (\widehat{\tau}_{k,m} - \tau_{k,m})\overline{W}_{k,m}.
$$

Then,

$$
\begin{aligned}
&\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}\widetilde{W}_{k,i}\widetilde{U}_{k,i}\right)^2 \\
&= \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}\widetilde{W}_{k,i}\Big(\overline{U}_{k,i} + \big(\widetilde{U}_{k,i} - \overline{U}_{k,i}\big)\Big)\right)^2 \\
&= \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}\widetilde{W}_{k,i}\Big(\overline{U}_{k,i} - (\widehat{\tau}_k^{\text{fixed}} - \tau_k)(W_{k,i} - \overline{W}_{k,m})\Big)\right)^2 \\
&= \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}\widetilde{W}_{k,i}\overline{U}_{k,i} - (\widehat{\tau}_k^{\text{fixed}} - \tau_k)\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}W_{k,i}(W_{k,i} - \overline{W}_{k,m})\right)^2.
\end{aligned}
$$

Using the formula for the second moment of a binomial distribution and $n \geqslant 1$, we obtain,

$$
\begin{aligned}
&E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}W_{k,i}(W_{k,i} - \overline{W}_{k,m})\right)^2 \middle| \overline{N}_{k,m} = n\right] \\
&\qquad = E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}(1 - \overline{W}_{k,m})R_{k,i}W_{k,i}\right)^2 \middle| \overline{N}_{k,m} = n\right] \\
&\qquad \leqslant E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}W_{k,i}\right)^2 \middle| \overline{N}_{k,m} = n\right] \\
&\qquad \leqslant n^2 + n.
\end{aligned}
$$

From the formula of the sum of the first two moments of a binomial distribution, we obtain

$$
\sum_{m=1}^{m_k} E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}W_{k,i}(W_{k,i} - \overline{W}_{k,m})\right)^2\right] \leqslant \sum_{m=1}^{m_k} (n_{k,m}^2 p_k^2 q_k + n_{k,m}p_k q_k).
$$

Therefore,

$$
\frac{1}{n_k p_k q_k f_k}(\widehat{\tau}_k^{\text{fixed}} - \tau_k)^2 \sum_{m=1}^{m_k} E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}W_{k,i}(W_{k,i} - \overline{W}_{k,m})\right)^2\right]
$$

$$\leqslant \frac{n_k p_k q_k}{f_k}(\widehat{\tau}_k^{\text{fixed}} - \tau_k)^2 \frac{1}{(n_k p_k q_k)^2} \sum_{m=1}^{m_k} (n_{k,m}^2 p_k^2 q_k + n_{k,m} p_k q_k)$$
$$= \mathcal{O}_p(1) \left(\frac{\max_m n_{k,m}}{\min_m n_{k,m}} \frac{1}{m_k q_k} + \frac{1}{n_k p_k q_k} \right) \xrightarrow{p} 0.$$

Now, notice that

$$\frac{1}{n_k p_k q_k f_k} \sum_{m=1}^{m_k} \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} \widetilde{W}_{k,i} \overline{U}_{k,i} \right)^2$$
$$= \frac{1}{n_k p_k q_k f_k} \sum_{m=1}^{m_k} \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} \big(e_{k,i}(1) W_{k,i} + e_{k,i}(0)(1 - W_{k,i})\big)(W_{k,i} - \overline{W}_{k,m}) \right.$$
$$\left. + (\tau_{k,m} - \tau_k) \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} (W_{k,i} - \overline{W}_{k,m})^2 \right)^2.$$

Equation (A.9) (and the analogous result for the sum involving terms with $e_{k,i}(0)$), implies

$$\frac{1}{n_k p_k q_k f_k} \sum_{m=1}^{m_k} \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} \big(e_{k,i}(1) W_{k,i} + e_{k,i}(0)(1 - W_{k,i})\big)(\overline{W}_{k,m} - A_{k,m}) \right)^2 \xrightarrow{p} 0.$$

As a result, it is enough to establish convergence of $\overline{\Sigma}_k/(n_k p_k q_k f_k)$, where

$$\overline{\Sigma}_k = \sum_{m=1}^{m_k} \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} \big(e_{k,i}(1) W_{k,i} + e_{k,i}(0)(1 - W_{k,i})\big)(W_{k,i} - A_{k,m}) \right.$$
$$\left. + (\tau_{k,m} - \tau_k) \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} (W_{k,i} - \overline{W}_{k,m})^2 \right)^2$$
$$= \sum_{m=1}^{m_k} \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} \Big(R_{k,i} W_{k,i} (W_{k,i} - A_{k,m}) - p_k q_k A_{k,m} (1 - A_{k,m}) \Big) e_{k,i}(1) \right.$$
$$+ \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} \Big(R_{k,i} (1 - W_{k,i})(W_{k,i} - A_{k,m}) + p_k q_k A_{k,m} (1 - A_{k,m}) \Big) e_{k,i}(0)$$
$$\left. + (\tau_{k,m} - \tau_k) \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} (W_{k,i} - \overline{W}_{k,m})^2 \right)^2.$$

We will next show that

$$\frac{1}{n_k p_k q_k f_k} \overline{\Sigma}_k - \frac{f_k^{\text{cluster}}}{f_k} \xrightarrow{p} 0, \tag{A.13}$$

where

$$f_k^{\text{cluster}} = \frac{1}{n_k} E[A_{k,m}(1 - A_{k,m})^2] \sum_{i=1}^{n_k} e_{k,i}^2(1)$$
$$+ \frac{1}{n_k} E[A_{k,m}^2 (1 - A_{k,m})] \sum_{i=1}^{n_k} e_{k,i}^2(0)$$

$$
\begin{aligned}
&- \frac{1}{n_k} p_k E[A_{k,m}^2(1-A_{k,m})^2] \sum_{i=1}^{n_k} (e_{k,i}(1) - e_{k,i}(0))^2 \\
&+ (E[A_{k,m}(1-A_{k,m})] - (5+p_k)E[A_{k,m}^2(1-A_{k,m})^2]) \sum_{m=1}^{m_k} \frac{n_{k,m}}{n_k}(\tau_{k,m} - \tau_k)^2 \\
&+ p_k E[A_{k,m}^2(1-A_{k,m})^2] \sum_{m=1}^{m_k} \frac{n_{k,m}^2}{n_k}(\tau_{k,m} - \tau_k)^2.
\end{aligned}
$$

Let

$$
\begin{aligned}
X_{k,m} = \frac{1}{n_k p_k q_k} \Bigg(& \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} \big(e_{k,i}(0)(1 - W_{k,i}) \\
& + e_{k,i}(1) W_{k,i}\big)(W_{k,i} - A_{k,m}) + (\tau_{k,m} - \tau_k) \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} (W_{k,i} - \overline{W}_{k,m})^2 \Bigg)^2
\end{aligned}
$$

Using the result in equation (A.8) and results on the moments of the binomial distribution (see intermediate calculations in section A.7), we obtain

$$
\begin{aligned}
\frac{1}{n_k p_k q_k} E[\overline{\Sigma}_k] &= \sum_{m=1}^{m_k} E[X_{k,m}] \\
&= f_k^{\text{cluster}} + o(1).
\end{aligned}
$$

Therefore, to show that equation (A.13) holds, we will show

$$
\frac{1}{f_k^2} \sum_{m=1}^{m_k} E[X_{k,m}^2] \longrightarrow 0. \tag{A.14}
$$

Let

$$
\begin{aligned}
\theta_k &= E[(R_{k,i} W_{k,i}(W_{k,i} - A_{k,m}) - p_k A_{k,m}(1 - A_{k,m}))^2 | m_{k,i} = m, Q_{k,m} = 1] \\
&= p_k \left(E[A_{k,m}(1-A_{k,m})^2] - p_k E[A_{k,m}^2(1-A_{k,m})^2] \right),
\end{aligned}
$$

and

$$
\begin{aligned}
\pi_k &= E[(R_{k,i} W_{k,i}(W_{k,i} - A_{k,m}) - p_k A_{k,m}(1 - A_{k,m}))^4 | m_{k,i} = m, Q_{k,m} = 1] \\
&= p_k E[(W_{k,i}(W_{k,i} - A_{k,m}) - p_k A_{k,m}(1 - A_{k,m}))^4 | m_{k,i} = m] + p_k^4 (1-p_k) E[A_{k,m}^4 (1-A_{k,m})^4].
\end{aligned}
$$

Let

$$
\begin{aligned}
X_{k,m,1} &= \frac{1}{n_k p_k q_k} \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i}(W_{k,i} - A_{k,m}) e_{k,i}(1) \right)^2 \\
&= \frac{Q_{k,m}}{n_k p_k q_k} \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} (R_{k,i} W_{k,i}(W_{k,i} - A_{k,m}) - p_k A_{k,m}(1 - A_{k,m})) e_{k,i}(1) \right)^2.
\end{aligned}
$$

Then,

$$
\begin{aligned}
E[X_{k,m,1}^2] &= q_k E[X_{k,m,1}^2 | Q_{k,m} = 1] \\
&= \frac{\pi_k}{n_k^2 p_k^2 q_k} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} e_{k,i}^4(1) \\
&+ \frac{6\theta_k^2}{n_k^2 p_k^2 q_k} \sum_{i=1}^{n_k-1} \sum_{j=i+1}^{n_k} 1\{m_{k,i} = m_{k,j} = m\} e_{k,i}^2(1) e_{k,j}^2(1).
\end{aligned}
$$

Therefore, because $n_k p_k q_k \to \infty$ and $m_k q_k \to \infty$, we obtain

$$
\begin{aligned}
\sum_{m=1}^{m_k} E[X_{k,m,1}^2] \leqslant{}& \frac{c}{n_k p_k q_k} \left(\frac{1}{n_k} \sum_{i=1}^{n_k} e_{k,i}^4(1) \right) + \frac{c}{m_k q_k} \frac{\max_m n_{k,m}^2}{\min_m n_{k,m}^2} \\
& \times \left(\frac{1}{m_k} \sum_{m=1}^{m_k} \frac{1}{\max_m n_{k,m}^2} \sum_{i=1}^{n_k-1} \sum_{j=i+1}^{n_k} 1\{m_{k,i} = m_{k,j} = m\} e_{k,i}^2(1) e_{k,j}^2(1) \right) \\
& \longrightarrow 0.
\end{aligned} \tag{A.15}
$$

Using the same argument, we obtain

$$
\sum_{m=1}^{m_k} E[X_{k,m,2}^2] \longrightarrow 0, \tag{A.16}
$$

where

$$
X_{k,m,2} = \frac{1}{n_k p_k q_k} \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} (1 - W_{k,i})(W_{k,i} - A_{k,m}) e_{k,i}(0) \right)^2.
$$

Notice that equations (A.15) and (A.16) imply

$$
\frac{1}{f_k^2} \sum_{m=1}^{m_k} E[X_{k,m,1}^2] \longrightarrow 0
$$

and

$$
\frac{1}{f_k^2} \sum_{m=1}^{m_k} E[X_{k,m,2}^2] \longrightarrow 0.
$$

Notice that the last two equations hold even if f_k is bounded (e.g., when $\tau_{k,m} - \tau_k = 0$ for all k and m), as long as f_k is bounded away from zero in large samples. In section A.3.3 we derive conditions so that f_k is bounded away from zero in large samples even if $\tau_{k,m} - \tau_k = 0$ for all k and m. Now, let

$$
X_{k,m,3} = \frac{1}{n_k p_k q_k} \left((\tau_{k,m} - \tau_k) \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i} (W_{k,i} - \overline{W}_{k,m}) \right)^2.
$$

Recall that, under the conditions in (A.11) and (A.12), $f_k \to \infty$ and $p_k \min n_{k,m}/f_k$ is bounded for large k and, therefore, $p_k \max n_{k,m}/f_k$ is bounded for large k. Then (see intermediate calculations at the end of this document), for large k,

$$
\frac{1}{f_k^2} \sum_{m=1}^{m_k} E[X_{k,m,3}^2] = \frac{1}{(n_k p_k q_k f_k)^2} \sum_{m=1}^{m_k} n_{k,m}^4 p_k^4 q_k (\tau_{k,m} - \tau_k)^4 \left(1 + \mathcal{O}\left(\frac{1}{p_k \min_m n_{k,m}} \right) \right)
$$

$$= \frac{p_k \max_m n_{k,m}^2}{m_k q_k f_k \min_m n_{k,m}} \left(\frac{p_k}{f_k} \sum_{m=1}^{m_k} \frac{n_{k,m}^2}{n_k} (\tau_{k,m} - \tau_k)^4 \right) \left(1 + \mathcal{O}\left(\frac{1}{p_k \min_m n_{k,m}} \right) \right)$$
$$= \mathcal{O}\left(\frac{1}{m_k q_k} \right) \left(1 + \mathcal{O}\left(\frac{1}{p_k \min_m n_{k,m}} \right) \right) \to 0.$$

Now, Hölder's inequality implies that equation (A.14) holds (see intermediate calculations).

Now let,

$$\tilde{v}_k^{\text{cluster}} = f_k^{\text{cluster}} / (\mu_k(1-\mu_k) - \sigma_k^2)^2.$$

We obtain,

$$\frac{\widetilde{V}_k^{\text{cluster}}}{\tilde{v}_k} = \frac{\tilde{v}_k^{\text{cluster}}}{\tilde{v}_k} + o_p(1).$$

We will next establish the analogous result for the heteroskedaticity-robust variance estimator. Let

$$\widetilde{\Sigma}_k^{\text{robust}} = \sum_{i=1}^{n_k} R_{k,i} \widetilde{W}_{k,i}^2 \widetilde{U}_{k,i}^2.$$

Then, the heteroskedasticity-robust estimator of the variance of $\sqrt{N_k}(\hat{\tau}_k^{\text{fixed}} - \tau_k)$ is

$$\widetilde{V}_k^{\text{robust}} = N_k \widetilde{Q}_k^{-1} \widetilde{\Sigma}_k^{\text{robust}} \widetilde{Q}_k^{-1}.$$

As we have established before,

$$\begin{aligned} \widetilde{U}_{k,i} &= e_{k,i}(1) W_{k,i} + e_{k,i}(0)(1 - W_{k,i}) + (\tau_{k,m} - \tau_k)(W_{k,i} - \overline{W}_{k,m}) \\ &\quad - (\hat{\tau}_k^{\text{fixed}} - \tau_k)(W_{k,i} - \overline{W}_{k,m}) - (\hat{\alpha}_{k,m} - \alpha_{k,m}) - (\hat{\tau}_{k,m} - \tau_{k,m}) \overline{W}_{k,m}. \end{aligned}$$

For $m_{k,i} = m$ and $N_{k,m,0} N_{k,m,1} \geqslant 1$, let

$$\overline{U}_{k,i} = e_{k,i}(1) W_{k,i} + e_{k,i}(0)(1 - W_{k,i}) + (\tau_{k,m} - \tau_k)(W_{k,i} - \overline{W}_{k,m}),$$

and let $\overline{U}_{k,i} = 0$ for $m_{k,i} = m$ and $N_{k,m,0} N_{k,m,1} = 0$. Then, for $m_{k,i} = m$ and $N_{k,m,0} N_{k,m,1} \geqslant 1$, we have

$$\widetilde{U}_{k,i} - \overline{U}_{k,i} = -(\hat{\tau}_k^{\text{fixed}} - \tau_k)(W_{k,i} - \overline{W}_{k,m}) - (\hat{\alpha}_{k,m} - \alpha_{k,m}) - (\hat{\tau}_{k,m} - \tau_{k,m}) \overline{W}_{k,m},$$

and

$$\frac{1}{n_k p_k q_k} \sum_{i=1}^{n_k} R_{k,i} \widetilde{W}_{k,i}^2 \widetilde{U}_{k,i}^2 = \frac{1}{n_k p_k q_k} \sum_{i=1}^{n_k} R_{k,i} \widetilde{W}_{k,i}^2 \Big(\overline{U}_{k,i} + \big(\widetilde{U}_{k,i} - \overline{U}_{k,i} \big) \Big)^2. \tag{A.17}$$

Focusing on the part of the right hand side of last equation that depends on the first term of $\widetilde{U}_{k,i} - \overline{U}_{k,i}$, we obtain

$$\frac{1}{n_k p_k q_k} \sum_{i=1}^{n_k} R_{k,i} \widetilde{W}_{k,i}^4 (\hat{\tau}_k^{\text{fixed}} - \tau_k)^2 \leqslant (\hat{\tau}_k^{\text{fixed}} - \tau_k)^2 \frac{1}{n_k p_k q_k} \sum_{i=1}^{n_k} R_{k,i} \widetilde{W}_{k,i}^2 \xrightarrow{p} 0.$$

We will focus now on the part of the right-hand side of equation (A.17) that that depends on the second term of $\widetilde{U}_{k,i} - \overline{U}_{k,i}$,

$$\frac{1}{n_k p_k q_k} \sum_{m=1}^{m_k} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} \widetilde{W}_{k,i}^2 (\hat{\alpha}_{k,m} - \alpha_{k,m})^2.$$

Using the formula for the variance of a sample mean under sampling without replacement (e.g., in the supplement of **?**), we obtain for $1 \leqslant n \leqslant n_{k,m} - 1$,

$$
\begin{aligned}
E\Bigg[(\widehat{\alpha}_{k,m} - \alpha_{k,m})^2 \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} \widetilde{W}_{k,i}^2 \Big| N_{k,m,0} = n\Bigg] \\
&= E\Bigg[(\widehat{\alpha}_{k,m} - \alpha_{k,m})^2 \overline{N}_{k,m} \overline{W}_{k,m}(1 - \overline{W}_{k,m}) \Big| N_{k,m,0} = n\Bigg] \\
&\leqslant E\Big[n(\widehat{\alpha}_{k,m} - \alpha_{k,m})^2 \Big| N_{k,m,0} = n\Big] \\
&= n \operatorname{var}(\widehat{\alpha}_{k,m} | N_{k,m,0} = n) \\
&= s_{k,m,0}^2 \Big(1 - \frac{n}{n_{k,m}}\Big), \qquad \text{(A.18)}
\end{aligned}
$$

where

$$
s_{k,m,0}^2 = \frac{1}{n_{k,m} - 1} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\}(y_{k,i}(0) - \alpha_{k,m})^2.
$$

Because $s_{k,m,0}^2$ is bounded, so is the right-hand side of equation (A.18). As a result

$$
E\Bigg[\frac{1}{n_k p_k q_k} \sum_{m=1}^{m_k} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} \widetilde{W}_{k,i}^2 (\widehat{\alpha}_{k,m} - \alpha_{k,m})^2\Bigg] \leqslant c \, \frac{m_k}{n_k p_k} \longrightarrow 0.
$$

An analogous derivation applies to the part of the right-hand side of equation (A.17) that depends on the third term of $\widetilde{U}_{k,i} - \overline{U}_{k,i}$. (Notice that $\overline{W}_{k,m} \leqslant 1$ and that $\widehat{\tau}_{k,m} - \tau_{k,m}$ is equal to minus the difference between $\widehat{\alpha}_{k,m} - \alpha_{k,m}$ and the analogous difference for the treated.

Therefore, we will study the behavior of

$$
\frac{1}{n_k p_k q_k} \sum_{i=1}^{n_k} R_{k,i} \widetilde{W}_{k,i}^2 \overline{U}_{k,i}^2. \qquad \text{(A.19)}
$$

First, notice that

$$
\begin{aligned}
\frac{1}{n_k p_k q_k} \sum_{m=1}^{m_k} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} \Big| (W_{k,i} - \overline{W}_{k,m})^2 - (W_{k,i} - A_{k,m})^2 \Big| W_{k,i} e_{k,i}^2(1) \\
&\leqslant c \frac{1}{n_k p_k q_k} \sum_{m=1}^{m_k} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} \Big| (W_{k,i} - \overline{W}_{k,m}) + (W_{k,i} - A_{k,m}) \Big| \big| \overline{W}_{k,m} - A_{k,m} \big| W_{k,i} \\
&\leqslant c \Bigg(\frac{1}{n_k p_k q_k} \sum_{m=1}^{m_k} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} \Big((W_{k,i} - \overline{W}_{k,m}) + (W_{k,i} - A_{k,m}) \Big)^2 W_{k,i} \Bigg)^{1/2} \\
&\quad \times \Bigg(\frac{1}{n_k p_k q_k} \sum_{m=1}^{m_k} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} \big(\overline{W}_{k,m} - A_{k,m} \big)^2 \Bigg)^{1/2}. \qquad \text{(A.20)}
\end{aligned}
$$

The inside of the first square root in equation (A.20) is bounded by a constant times

$$
\frac{1}{n_k p_k q_k} \sum_{m=1}^{m_k} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i},
$$

which converges in probability to one. The expectation of the inside of the second square root in equation (A.20) is

$$\frac{1}{n_k p_k q_k}\sum_{m=1}^{m_k} E\big[\overline{N}_{k,m}\big(\overline{W}_{k,m} - A_{k,m}\big)^2\big] \leqslant c\frac{m_k}{n_k p_k} \longrightarrow 0.$$

As a result, the right-hand side of equation (A.20) converges to zero in probability. The derivation with $(1 - W_{k,i})e_{k,i}^2(0)$ replacing $W_{k,i}e_{k,i}^2(1)$ in equation (A.20) is analogous. Now, notice that

$$\begin{aligned}&(W_{k,i} - \overline{W}_{k,m})^4 - (W_{k,i} - \overline{A}_{k,m})^4\\ &\quad = -\big((W_{k,i} - \overline{W}_{k,m})^2 + (W_{k,i} - A_{k,m})^2\big)\big((W_{k,i} - \overline{W}_{k,m}) + (W_{k,i} - A_{k,m})\big)\big(\overline{W}_{k,m} - A_{k,m}\big).\end{aligned}$$

Because the first factor of the expression above is bounded, we obtain

$$\begin{aligned}&\frac{1}{n_k p_k q_k}\sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}\Big|(W_{k,i} - \overline{W}_{k,m})^4 - (W_{k,i} - A_{k,m})^4\Big|(\tau_{k,m} - \tau_k)^2\\ &\quad\leqslant c\left(\frac{1}{n_k p_k q_k}\sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}\right)^{1/2}\\ &\quad\quad\times\left(\frac{1}{n_k p_k q_k}\sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}\big(\overline{W}_{k,m} - A_{k,m}\big)^2\right)^{1/2}. \qquad (A.21)\end{aligned}$$

Now, the right-hand side of equation (A.21) converges to zero in probability by the same argument as for equation (A.20). Cauchy-Schwarz inequality implies,

$$\frac{1}{n_k p_k q_k}\sum_{i=1}^{n_k} R_{k,i}\widetilde{W}_{k,i}^2\overline{U}_{k,i}^2 = \frac{1}{n_k p_k q_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}(W_{k,i} - A_{k,m})^2\breve{U}_{k,i}^2 + o_p(1),$$

where

$$\breve{U}_{k,i} = e_{k,i}(1)W_{k,i} + e_{k,i}(0)(1 - W_{k,i}) + (\tau_{k,m} - \tau_k)(W_{k,i} - A_{k,m}), \qquad (A.22)$$

for $m_{k,i} = m$ and $N_{k,m,0}N_{k,m,1} \geqslant 1$, and $\breve{U}_{k,i} = 0$ for $N_{k,m,0}N_{k,m,1} \geqslant 0$. Therefore, we will study the behavior of

$$\frac{1}{n_k p_k q_k}\sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}(W_{k,i} - A_{k,m})^2\breve{U}_{k,i}^2.$$

We know,

$$\begin{aligned}&\frac{1}{n_k p_k q_k}\sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}(W_{k,i} - A_{k,m})^2 W_{k,i}e_{k,i}^2(1)\\ &\qquad\qquad - E[A_{k,m}(1 - A_{k,m})^2]\frac{1}{n_k}\sum_{i=1}^{n_k} e_{k,i}^2(1) \xrightarrow{p} 0,\end{aligned}$$

and

$$\frac{1}{n_k p_k q_k}\sum_{m=1}^{m_k}\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}(W_{k,i} - A_{k,m})^2(1 - W_{k,i})e_{k,i}^2(0)$$

$$- E[A_{k,m}^2(1 - A_{k,m})] \frac{1}{n_k} \sum_{i=1}^{n_k} e_{k,i}^2(0) \xrightarrow{p} 0.$$

Now, notice that

$$\begin{aligned} E[(W_{k,i} - A_{k,m})^4 | m_{k,i} = m, R_{k,i} = 1, A_{k,m} = a] &= (1-a)^4 a + a^4(1-a) \\ &= a(1-a)[(1-a)^3 + a^3] \\ &= a(1-a)[1 - 3a(1-a)], \end{aligned}$$

which implies

$$\begin{aligned} &E\left[\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} (W_{k,i} - A_{k,m})^4 (\tau_{k,m} - \tau_k)^2 \right] \\ &\qquad = n_{k,m} p_k q_k E[A_{k,m}(1 - A_{k,m})(1 - 3A_{k,m}(1 - A_{k,m}))](\tau_{k,m} - \tau_k)^2, \end{aligned}$$

and

$$\begin{aligned} &E\left[\frac{1}{n_k p_k q_k} \sum_{m=1}^{m_k} \sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} (W_{k,i} - A_{k,m})^4 (\tau_{k,m} - \tau_k)^2 \right] \\ &\qquad = E[A_{k,m}(1 - A_{k,m})(1 - 3A_{k,m}(1 - A_{k,m}))] \sum_{m=1}^{m_k} \frac{n_{k,m}}{n_k} (\tau_{k,m} - \tau_k)^2. \end{aligned}$$

Notice now that

$$\begin{aligned} \frac{1}{(n_k p_k q_k)^2} \sum_{m=1}^{m_k} E&\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} (W_{k,i} - A_{k,m})^4 (\tau_{k,m} - \tau_k)^2 \right)^2 \right] \\ &\leqslant c \, \frac{1}{(n_k p_k q_k)^2} \sum_{m=1}^{m_k} E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} \right)^2 \right] \\ &\leqslant c \, \frac{q_k}{(n_k p_k q_k)^2} \sum_{m=1}^{m_k} (n_{k,m} p_k + n_{k,m}^2 p_k^2) \\ &= c \left(\frac{1}{n_k p_k q_k} + \frac{\max_m n_{k,m}}{\min_m n_{k,m}} \frac{1}{m_k q_k} \right) \xrightarrow{p} 0. \end{aligned}$$

Notice also that expectations of the sums of products of the terms on the right-hand side of equation (A.22) are equal to zero. Then,

$$\frac{1}{n_k p_k q_k} \widetilde{\Sigma}_k^{\text{robust}} - f_k^{\text{robust}} \xrightarrow{p} 0,$$

where

$$f_k^{\text{robust}} = E[A_{k,m}(1 - A_{k,m})^2] \frac{1}{n_k} \sum_{i=1}^{n_k} e_{k,i}^2(1) + E[A_{k,m}^2(1 - A_{k,m})] \frac{1}{n_k} \sum_{i=1}^{n_k} e_{k,i}^2(0)$$

$$+ E[A_{k,m}(1-A_{k,m})(1-3A_{k,m}(1-A_{k,m}))] \sum_{m=1}^{m_k} \frac{n_{k,m}}{n_k}(\tau_{k,m}-\tau_k)^2.$$

Now let,

$$\tilde{v}_k^{\text{robust}} = f_k^{\text{robust}}/(\mu_k(1-\mu_k)-\sigma_k^2)^2.$$

We obtain,

$$\widetilde{V}_k^{\text{robust}} = \tilde{v}_k^{\text{robust}} + o_p(1).$$

A.3.3. Large k results the fixed effects case under homogeneous average treatment effects across clusters

We will now study the Lyapounov's condition for the case $\tau_{k,m} = \tau_k$ for all k and $m = 1, \ldots, m_k$, so

$$f_k = \sum_{m=1}^{m_k} E[\psi_{k,m}^2].$$

Notice that

$$\begin{aligned}
\sum_{m=1}^{m_k} E[\psi_{k,m}^2] \geqslant{} & \frac{1}{n_k} E[A_{k,m}(1-A_{k,m})^2] \sum_{i=1}^{n_k} 1\{m_{k,i}=m\} e_{k,i}^2(1) \\
& + \frac{1}{n_k} E[A_{k,m}^2(1-A_{k,m})] \sum_{i=1}^{n_k} 1\{m_{k,i}=m\} e_{k,i}^2(0) \\
& - \frac{1}{n_k} E[A_{k,m}^2(1-A_{k,m})^2] \sum_{i=1}^{n_k} 1\{m_{k,i}=m\} (e_{k,i}(1)-e_{k,i}(0))^2 \\
={} & \frac{1}{n_k} E[A_{k,m}(1-A_{k,m})^3] \sum_{i=1}^{n_k} 1\{m_{k,i}=m\} e_{k,i}^2(1) \\
& + \frac{1}{n_k} E[A_{k,m}^3(1-A_{k,m})] \sum_{i=1}^{n_k} 1\{m_{k,i}=m\} e_{k,i}^2(0) \\
& + \frac{2}{n_k} E[A_{k,m}^2(1-A_{k,m})^2] \sum_{i=1}^{n_k} 1\{m_{k,i}=m\} e_{k,i}(1) e_{k,i}(0) \\
={} & E\left[\frac{1}{n_k} \sum_{m=1}^{m_k} A_{k,m}^3(1-A_{k,m})^3 \sum_{i=1}^{n_k} 1\{m_{k,i}=m\} \left(\frac{e_{k,i}(1)}{A_{k,m}} + \frac{e_{k,i}(0)}{1-A_{k,m}}\right)^2\right].
\end{aligned}$$

Therefore,

$$\liminf_{k\to\infty} E\left[\frac{1}{n_k} \sum_{m=1}^{m_k} A_{k,m}^3(1-A_{k,m})^3 \sum_{i=1}^{n_k} 1\{m_{k,i}=m\} \left(\frac{e_{k,i}(1)}{A_{k,m}} + \frac{e_{k,i}(0)}{1-A_{k,m}}\right)^2\right] > 0$$

is sufficient for $\liminf_{k\to\infty} f_k > 0$ (even if condition (A.11) does not hold). Given our assumption that the supports of the cluster probabilities, $A_{k,m}$, are bounded away from zero and one (uniformly in k and m), then

$$\liminf_{k\to\infty} E\left[\frac{1}{n_k} \sum_{m=1}^{m_k} \sum_{i=1}^{n_k} 1\{m_{k,i}=m\} \left(\frac{e_{k,i}(1)}{A_{k,m}} + \frac{e_{k,i}(0)}{1-A_{k,m}}\right)^2\right] > 0 \tag{A.23}$$

is sufficient for $\liminf_{k\to\infty} f_k > 0$. Assume that (A.23) holds, so $\liminf_{k\to\infty} f_k > 0$. We now obtain,

$$
\begin{aligned}
E\left[\left|\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}R_{k,i}W_{k,i}(W_{k,i}-\overline{W}_{k,m})e_{k,i}(1)\right|^4 \middle| Q_{k,m}=1, A_{k,m}\right]& \\
&= E\left[(1-\overline{W}_{k,m})^4\left|\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}R_{k,i}W_{k,i}e_{k,i}(1)\right|^4 \middle| Q_{k,m}=1, A_{k,m}\right]\\
&\leqslant E\left[\left|\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}R_{k,i}W_{k,i}e_{k,i}(1)\right|^4 \middle| Q_{k,m}=1, A_{k,m}\right],
\end{aligned}
$$

and

$$
\begin{aligned}
E\left[\left|\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}R_{k,i}W_{k,i}e_{k,i}(1)\right|^4 \middle| Q_{k,m}=1, A_{k,m}\right]& \\
&= E\left[\left|\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}(R_{k,i}W_{k,i}-p_kA_{k,m})e_{k,i}(1)\right|^4 \middle| Q_{k,m}=1, A_{k,m}\right]\\
&= n_{k,m}E[(R_{k,i}W_{k,i}-p_kA_{k,m})^4|Q_{k,m}=1, A_{k,m}]\\
&+ 3n_{k,m}(n_{k,m}-1)(E[(R_{k,i}W_{k,i}-p_kA_{k,m})^2|Q_{k,m}=1, A_{k,m}])^2.
\end{aligned}
$$

The first equality holds because the terms $e_{k,i}(1)$ sum to zero within clusters. The second equality holds because, if $m_{k,i} = m_{k,j} = m$, with $i \neq j$, then $R_{k,i}W_{k,i}$ and $R_{k,i}W_{k,i}$ are independent conditional on $Q_{k,m} = 1, A_{k,m}$, and $E[R_{k,i}W_{k,i} - p_kA_{k,m}|Q_{k,m} = 1, A_{k,m}] = 0$. Notice that

$$E[(R_{k,i}W_{k,i}-p_kA_{k,m})^2|Q_{k,m}=1, A_{k,m}] = p_kA_{k,m}(1-p_kA_{k,m}) \leqslant p_k,$$

which also implies $E[(R_{k,i}W_{k,i} - p_kA_{k,m})^4|Q_{k,m} = 1, A_{k,m}] \leqslant p_k$. As a result,

$$
\begin{aligned}
\sum_{m=1}^{m_k} E\left[\left|\frac{1}{\sqrt{n_kp_kq_k}}\sum_{i=1}^{n_k} 1\{m_{k,i}=m\}R_{k,i}W_{k,i}e_{k,i}(1)\right|^4\right]& \\
&\leqslant \frac{1}{n_kp_kq_k} + 3\frac{\max_m n_{k,m}}{\min_m n_{k,m}}\frac{1}{m_kq_k} \to 0.
\end{aligned}
$$

A.4. Derivations of the variance estimators

In this section, we derive the adjustments in the CCV variance. (We do this under the assumption that the Z_i are independent. In our simulations we actually use a slightly different sampling scheme for the Z_i where the average $\overline{Z}_{k,m}$ is identical and fixed in each cluster.) To derive the CCV variance of the least squares estimator, consider first a variance estimator of the form

$$\left(\sum_{i=1}^{n} V_i\right)^2.$$

We aim, however, to design an estimator based on a subsample consisting of units with $Z_i = 1$, where $Z_i \in \{0,1\}$ is i.i.d. binary with $\Pr(Z_i = 1) = p_Z$ and independent of V_i. First, notice that

$$E\left[\left(\sum_{i=1}^{n} V_i\right)^2\right] = \sum_{i=1}^{n} E[V_i^2] + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} E[V_iV_j],$$

and

$$E\left[\left(\sum_{i=1}^{n} Z_i V_i\right)^2\right] = p_Z \sum_{i=1}^{n} E[V_i^2] + 2p_Z^2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E[V_i V_j].$$

Therefore,

$$E\left[\frac{1}{p_Z}\left(\sum_{i=1}^{n} Z_i V_i\right)^2\right] = \sum_{i=1}^{n} E[V_i^2] + 2p_Z \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E[V_i V_j],$$

and

$$\frac{(1-p_Z)}{p_Z^2}\left(E\left[\left(\sum_{i=1}^{n} Z_i V_i\right)^2\right] - p_Z \sum_{i=1}^{n} E[V_i^2]\right) = 2(1-p_Z) \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E[V_i V_j].$$

Adding the last two equations,

$$\begin{aligned} E\left[\left(\sum_{i=1}^{n} V_i\right)^2\right] &= \frac{1}{p_Z^2} E\left[\left(\sum_{i=1}^{n} Z_i V_i\right)^2\right] - \frac{(1-p_Z)}{p_Z} \sum_{i=1}^{n} E[V_i^2] \\ &= \frac{1}{p_Z^2} E\left[\left(\sum_{i=1}^{n} Z_i V_i\right)^2\right] - \frac{(1-p_Z)}{p_Z^2} \sum_{i=1}^{n} E[Z_i V_i^2]. \end{aligned} \tag{A.24}$$

The first term of the CCV variance estimator for least squares is based on the sample counterpart of the right-hand side of equation (A.24), with $1\{m_{k,i} = m\}R_{k,i}((W_{k,i} - \overline{W}_k)\widehat{U}_{k,i} - (\widehat{\tau}_{k,m} - \widehat{\tau}_k)\overline{W}_k(1 - \overline{W}_k))$ in the role of V_i.

To derive the CCV variance estimator for the fixed effect case, consider

$$\lambda_k = 1 - q_k \frac{(E[A_{k,m}(1-A_{k,m})])^2}{E[A_{k,m}^2(1-A_{k,m})^2]},$$

and let $f_k^{\text{CCV}} = \lambda_k f_k^{\text{cluster}} + (1-\lambda_k) f_k^{\text{robust}}$. This transformation is designed to reproduce the terms in f_k with factor

$$\sum_{m=1}^{m_k} \frac{n_{k,m}^2}{n_k} (\tau_{k,m} - \tau_k)^2.$$

These terms dominate f_k as k increases. It also reproduces several lower order terms.

Notice that

$$\begin{aligned} f_k^{\text{robust}} &= E[A_{k,m}(1-A_{k,m})^2] \frac{1}{n_k} \sum_{i=1}^{n_k} e_{k,i}^2(1) + E[A_{k,m}^2(1-A_{k,m})] \frac{1}{n_k} \sum_{i=1}^{n_k} e_{k,i}^2(0) \\ &\quad + \Big(E[A_{k,m}(1-A_{k,m})] - (5+p_k) E[A_{k,m}^2(1-A_{k,m})^2]\Big) \sum_{m=1}^{m_k} \frac{n_{k,m}}{n_k} (\tau_{k,m} - \tau_k)^2 \\ &\quad + (2+p_k) E[A_{k,m}^2(1-A_{k,m})^2] \sum_{m=1}^{m_k} \frac{n_{k,m}}{n_k} (\tau_{k,m} - \tau_k)^2. \end{aligned}$$

Then,

$$f_k^{\mathrm{CCV}} - f_k = (1-\lambda_k) p_k E[A_{k,m}^2 (1-A_{k,m})^2] \left(\sum_{m=1}^{m_k} \frac{n_{k,m}}{n_k} (\tau_{k,m} - \tau_k)^2 + \frac{1}{n_k} \sum_{i=1}^{n_k} (e_{k,i}(1) - e_{k,i}(0))^2 \right)$$
$$= p_k q_k (E[A_{k,m}(1-A_{k,m})])^2 \left(\sum_{m=1}^{m_k} \frac{n_{k,m}}{n_k} (\tau_{k,m} - \tau_k)^2 + \frac{1}{n_k} \sum_{i=1}^{n_k} (e_{k,i}(1) - e_{k,i}(0))^2 \right).$$

For $\tilde{v}_k^{\mathrm{CCV}} = f_k^{\mathrm{CCV}} / (\mu_k(1-\mu_k) - \sigma_k^2)^2$, we obtain,

$$\tilde{v}_k^{\mathrm{CCV}} - \tilde{v}_k = p_k q_k \sum_{m=1}^{m_k} \frac{n_{k,m}}{n_k} (\tau_{k,m} - \tau_k)^2 + p_k q_k \frac{1}{n_k} \sum_{i=1}^{n_k} (e_{k,i}(1) - e_{k,i}(0))^2. \tag{A.25}$$

The difference $\tilde{v}_k^{\mathrm{CCV}} - \tilde{v}_k$ is non-negative and of smaller order than $\tilde{v}_k$. Therefore, $\tilde{v}_k^{\mathrm{CCV}} / \tilde{v}_k \to 1$ (even if $\tilde{v}_k^{\mathrm{CCV}} - \tilde{v}_k$ is bounded away from zero). The first term on the right-hand side of (A.25) could be estimated to further correct the difference between the CCV estimator and the variance of $\widehat{\tau}_k^{\mathrm{fixed}}$.

A.5. Limit results

Let $X_{k,m}$ be an infinite array of random variables, with rows indexed by $k = 1, 2, \ldots$, and the columns of the k-th row indexed by $m = 1, \ldots, m_k$. Let

$$S_k = \sum_{m=1}^{m_k} X_{k,m},$$

and $a_k = E[S_k]$.

A Weak Law of Large Numbers for Arrays: For each $k = 1, 2, \ldots$, suppose that $X_{k,1}, \ldots, X_{k,m_k}$ are independent and have finite second moments. In addition, let b_k be a sequence of positive constants such that

$$\frac{1}{b_k^2} \sum_{m=1}^{m_k} E[X_{k,m}^2] \longrightarrow 0.$$

Then,

$$\frac{S_k - a_k}{b_k} \xrightarrow{p} 0.$$

Proof: By Chebyshev's inequality, for any $\varepsilon > 0$

$$\begin{aligned}
\Pr\left(\left| \frac{S_k - a_k}{b_k} \right| > \varepsilon \right) &\leqslant \frac{1}{b_k^2 \varepsilon^2} \mathrm{var}(S_k) \\
&= \frac{1}{b_k^2 \varepsilon^2} \sum_{m=1}^{m_k} \mathrm{var}(X_{k,m}) \\
&\leqslant \frac{1}{b_k^2 \varepsilon^2} \sum_{m=1}^{m_k} E[X_{k,m}^2] \longrightarrow 0.
\end{aligned}$$

□

A Central Limit Theorem for Arrays: For each $k = 1, 2, \ldots$, suppose that $X_{k,1}, \ldots, X_{k,m_k}$ are independent, with zero means, $E[X_{k,m}] = 0$, and finite variances, $\sigma^2_{k,m} = E[X^2_{k,m}]$, for $m = 1, \ldots, m_k$. Let

$$s_k^2 = \sum_{m=1}^{m_k} \sigma^2_{k,m}.$$

Assume also that Lyapounov's condition holds,

$$\lim_{k\to\infty} \frac{1}{s_k^{2+\delta}} \sum_{m=1}^{m_k} E[|X_{k,m}|^{2+\delta}] = 0,$$

for some $\delta > 0$. Then,

$$\frac{S_k}{s_k} \xrightarrow{d} N(0,1).$$

Proof: **?**, Chapter 27.

A.6. Intermediate calculations for Section A.2

The calculation of v_k uses the following results.

$$E[(R_{k,i}W_{k,i} - p_k q_k \mu_k)^2] = p_k q_k \mu_k (1 - p_k q_k \mu_k),$$

$$E[(R_{k,i}(1 - W_{k,i}) - p_k q_k (1 - \mu_k))^2] = p_k q_k (1 - \mu_k)(1 - p_k q_k (1 - \mu_k)),$$

$$E[(R_{k,i}W_{k,i} - p_k q_k \mu_k)(R_{k,i}(1 - W_{k,i}) - p_k q_k (1 - \mu_k))] = -p_k^2 q_k^2 \mu_k (1 - \mu_k),$$

$$E[R_{k,i}W_{k,i}R_{k,j}W_{k,j} | m_{k,i} = m_{k,j}] = E[p_k^2 q_k A^2_{k,m}] = p_k^2 q_k (\sigma_k^2 + \mu_k^2),$$

and

$$\begin{aligned} E[(R_{k,i}W_{k,i} - p_k q_k \mu_k)(R_{k,j}W_{k,j} - p_k q_k \mu_k) | m_{k,i} = m_{k,j}] &= p_k^2 q_k (\sigma_k^2 + \mu_k^2) - (p_k q_k \mu_k)^2 \\ &= p_k^2 q_k (\sigma_k^2 + (1 - q_k)\mu_k^2). \end{aligned}$$

Similarly,

$$\begin{aligned} &E[(R_{k,i}(1 - W_{k,i}) - p_k q_k (1 - \mu_k))(R_{k,j}(1 - W_{k,j}) - p_k q_k (1 - \mu_k)) | m_{k,i} = m_{k,j}] \\ &\quad = p_k^2 q_k (\sigma_k^2 + (1 - q_k)(1 - \mu_k)^2). \end{aligned}$$

Notice also that

$$\begin{aligned} E[R_{k,i}W_{k,i}R_{k,j}(1 - W_{k,j}) | m_{k,i} = m_{k,j}] &= E[p_k^2 q_k A_{k,m}(1 - A_{k,m})] \\ &= p_k^2 q_k (\mu_k (1 - \mu_k) - \sigma_k^2), \end{aligned}$$

and

$$\begin{aligned} &E[(R_{k,i}W_{k,i} - p_k q_k \mu_k)(R_{k,j}(1 - W_{k,j}) - p_k q_k (1 - \mu_k)) | m_{k,i} = m_{k,j}] \\ &\quad = p_k^2 q_k (\mu_k (1 - \mu_k) - \sigma_k^2) - p_k^2 q_k^2 \mu_k (1 - \mu_k) \\ &\quad = p_k^2 q_k (\mu_k (1 - \mu_k)(1 - q_k) - \sigma_k^2). \end{aligned}$$

The following bounds are useful to prove Lyapunov's condition.

$$\begin{aligned}E[|R_{k,i}W_{k,i} - p_kq_k\mu_k|^3] &= (1-p_kq_k\mu_k)^3p_kq_k\mu_k + (p_kq_k\mu_k)^3(1-p_kq_k\mu_k)\\ &\leqslant c\,p_kq_k.\end{aligned}$$

Let $Q_{k,m}$ be a binary indicator that takes value one if cluster m of population k is sampled.

$$\begin{aligned}&E\big[|R_{k,i}W_{k,i} - p_kq_k\mu_k|^2|R_{k,j}W_{k,j} - p_kq_k\mu_k|\big|m_{k,i} = m_{k,j} = m\big]\\ &\quad= E\big[\big((1-p_kq_k\mu_k)^2p_kA_{k,m} + (p_kq_k\mu_k)^2(1-p_kA_{k,m})\big)\\ &\qquad\qquad\times\big((1-p_kq_k\mu_k)p_kA_{k,m} + (p_kq_k\mu_k)(1-p_kA_{k,m})\big)\big|m_{k,i} = m_{k,j} = m, Q_{k,m} = 1\big]q_k\\ &\quad+ E\big[\big(p_kq_k\mu_k\big)^3\big|m_{k,i} = m_{k,j} = m, Q_{k,m} = 0\big](1-q_k)\\ &\quad\leqslant cp_k^2q_k.\end{aligned}$$

$$\begin{aligned}&E\big[|R_{k,i}W_{k,i} - p_kq_k\mu_k||R_{k,j}W_{k,j} - p_kq_k\mu_k||R_{k,t}W_{k,t} - p_kq_k\mu_k|\big|m_{k,i} = m_{k,j} = m_{k,t} = m\big]\\ &\quad= E\Big[\big((1-p_kq_k\mu_k)p_kA_{k,m} + (p_kq_k\mu_k)(1-p_kA_{k,m})\big)^3\big|m_{k,i} = m_{k,j} = m_{k,t} = m, Q_{k,m} = 1\Big]q_k\\ &\quad+ E\Big[\big(p_kq_k\mu_k\big)^3\big|m_{k,i} = m_{k,j} = m_{k,t} = m, Q_{k,m} = 1\Big](1-q_k)\\ &\quad\leqslant cp_k^3q_k.\end{aligned}$$

Other useful intermediate calculations.

For the moments of treatment indicators, notice that $E[(W_{k,i} - \mu_k)^2W_{k,i}] = \mu_k(1-\mu_k)^2$, and $E[(W_{k,i} - \mu_k)^2(1-W_{k,i})] = (1-\mu_k)\mu_k^2$. In addition,

$$\begin{aligned}E[W_{k,i}W_{k,j}|m_{k,i} = m_{k,j}] &= E[A_{k,m}^2] \quad (\text{for } m \in \{1,\dots,m_k\})\\ &= \sigma_k^2 + \mu_k^2.\end{aligned}$$

Similarly, $E[(1-W_{k,i})(1-W_{k,j})|m_{k,i} = m_{k,j}] = \sigma_k^2 + (1-\mu_k)^2$. Therefore, $E[(W_{k,i}-\mu_k)W_{k,j}|m_{k,i} = m_{k,j}] = \sigma_k^2$ and $E[(W_{k,i} - \mu_k)(1-W_{k,j})|m_{k,i} = m_{k,j}] = -\sigma_k^2$. In addition,

$$\begin{aligned}&E[(W_{k,i} - \mu_k)(W_{k,j} - \mu_k)W_{k,i}W_{k,j}|m_{k,i} = m_{k,j}]\\ &\qquad= E[A_{k,m}^2](1-\mu_k)^2 \quad (\text{for } m \in \{1,\dots,m_k\})\\ &\qquad= (\sigma_k^2 + \mu_k^2)(1-\mu_k)^2.\end{aligned}$$

Similarly,

$$E[(W_{k,i} - \mu_k)(W_{k,j} - \mu_k)(1-W_{k,i})(1-W_{k,j})|m_{k,i} = m_{k,j}] = (\sigma_k^2 + (1-\mu_k)^2)\mu_k^2,$$

and

$$E[(W_{k,i} - \mu_k)(W_{k,j} - \mu_k)W_{k,i}(1-W_{k,j})|m_{k,i} = m_{k,j}] = \mu_k(1-\mu_k)(\sigma_k^2 - \mu_k(1-\mu_k)).$$

$\text{var}(R_{k,i}W_{k,i}) = p_kq_k\mu_k(1-p_kq_k\mu_k)$, $\text{var}(R_{k,i}(1-W_{k,i})) = p_kq_k(1-\mu_k)(1-p_kq_k(1-\mu_k))$. Moreover,

$$\begin{aligned}\text{cov}(R_{k,i}W_{k,i}, R_{k,i}(1-W_{k,i})) &= E[R_{k,i}W_{k,i}R_{k,i}(1-W_{k,i})] - E[R_{k,i}W_{k,i}]E[R_{k,i}(1-W_{k,i})] \\ &= -p_k^2q_k^2\mu_k(1-\mu_k).\end{aligned}$$

Recall that $E[W_{k,i}W_{k,j}|m_{k,i} = m_{k,j}] = \sigma_k^2 + \mu_k^2$. Therefore, $\text{cov}(W_{k,i}, W_{k,j}|m_{k,i} = m_{k,j}) = \sigma_k^2$. Also,

$$E[W_{k,i}(1-W_{k,j})|m_{k,i} = m_{k,j}] = \mu_k(1-\mu_k) - \sigma_k^2.$$

$$\begin{aligned}E[R_{k,i}W_{k,i}R_{k,j}W_{k,j}|m_{k,i} = m_{k,j}] &= E[R_{k,i}R_{k,j}|m_{k,i} = m_{k,j}]E[W_{k,i}W_{k,j}|m_{k,i} = m_{k,j}] \\ &= p_k^2q_k(\sigma_k^2 + \mu_k^2).\end{aligned}$$

Similarly,

$$E[R_{k,i}(1-W_{k,i})R_{k,j}(1-W_{k,j})|m_{k,i} = m_{k,j}] = p_k^2q_k(\sigma_k^2 + (1-\mu_k)^2).$$

Therefore,

$$\begin{aligned}\text{cov}(R_{k,i}W_{k,i}, R_{k,j}W_{k,j}|m_{k,i} = m_{k,j}) &= p_k^2q_k(\sigma_k^2 + \mu_k^2) - p_k^2q_k^2\mu_k^2 \\ &= p_k^2q_k(\sigma_k^2 + \mu_k^2(1-q_k)),\end{aligned}$$

and

$$\begin{aligned}\text{cov}(R_{k,i}(1-W_{k,i}), R_{k,j}(1-W_{k,j})|m_{k,i} = m_{k,j}) &= p_k^2q_k(\sigma_k^2 + (1-\mu_k)^2) - p_k^2q_k^2(1-\mu_k)^2 \\ &= p_k^2q_k(\sigma_k^2 + (1-\mu_k)^2(1-q_k)).\end{aligned}$$

In addition,

$$\begin{aligned}\text{cov}(R_{k,i}W_{k,i}, R_{k,j}(1-W_{k,j})|m_{k,i} = m_{k,j}) &= E[R_{k,i}W_{k,i}R_{k,j}(1-W_{k,j})|m_{k,i} = m_{k,j}] \\ &\quad - E[R_{k,i}W_{k,i}|m_{k,i} = m_{k,j}]E[R_{k,j}(1-W_{k,j})|m_{k,i} = m_{k,j}] \\ &= E[R_{k,i}R_{k,j}|m_{k,i} = m_{k,j}]E[W_{k,i}(1-W_{k,j})|m_{k,i} = m_{k,j}] \\ &\quad - E[R_{k,i}W_{k,i}|m_{k,i} = m_{k,j}]E[R_{k,j}(1-W_{k,j})|m_{k,i} = m_{k,j}] \\ &= p_k^2q_k(\mu_k(1-\mu_k) - \sigma_k^2) - p_k^2q_k^2\mu_k(1-\mu_k) \\ &= p_k^2q_k(\mu_k(1-\mu_k)(1-q_k) - \sigma_k^2).\end{aligned}$$

A.7. Intermediate calculations for Section A.3

$$E[R_{k,i}W_{k,i}(W_{k,i} - A_{k,m})|A_{k,m}, Q_{k,m} = 1, m_{k,i} = m] = p_kA_{k,m}(1-A_{k,m}).$$

This implies

$$E[R_{k,i}W_{k,i}(W_{k,i} - A_{k,m})|m_{k,i} = m] = p_kq_kE[A_{k,m}(1-A_{k,m})].$$

Therefore,

$$E\left[\sum_{i=1}^{n_k} 1\{m_{k,i} = m\}R_{k,i}W_{k,i}(W_{k,i} - A_{k,m})\right] = n_{k,m}p_kq_kE[A_{k,m}(1-A_{k,m})].$$

For $n \geqslant 1$,

$$
\begin{aligned}
&E\left[\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i} (\overline{W}_{k,m} - A_{k,m}) \middle| \overline{N}_{k,m} = n\right] \\
&\qquad = \frac{1}{n} E\left[\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i} \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i} - n A_{k,m}\right) \middle| \overline{N}_{k,m} = n\right] \\
&\qquad = E[A_{k,m}(1 - A_{k,m})].
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
E\left[\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i} (\overline{W}_{k,m} - A_{k,m})\right] &= E[A_{k,m}(1 - A_{k,m})] \Pr(\overline{N}_{k,m} \geqslant 1) \\
&= q_k E[A_{k,m}(1 - A_{k,m})](1 - (1 - p_k)^{n_{k,m}}).
\end{aligned}
$$

For $n \geqslant 1$

$$
\begin{aligned}
E[R_{k,i} W_{k,i} (\overline{W}_{k,m} &- A_{k,m})^2 | m_{k,i} = m, \overline{N}_{k,m} = n, R_{k,i} = 1] \\
&\leqslant E[(\overline{W}_{k,m} - A_{k,m})^2 | m_{k,i} = m, \overline{N}_{k,m} = n, R_{k,i} = 1] \\
&\leqslant \frac{E[A_{k,m}(1 - A_{k,m})]}{n}.
\end{aligned}
$$

Because $\Pr(R_{k,i} = 1 | \overline{N}_{k,m} = n, m_{k,i} = m) = n/n_{k,m}$, we obtain

$$
E[R_{k,i} W_{k,i} (\overline{W}_{k,m} - A_{k,m})^2 | m_{k,i} = m, \overline{N}_{k,m} = n] \leqslant \frac{E[A_{k,m}(1 - A_{k,m})]}{n_{k,m}},
$$

which implies

$$
E[R_{k,i} W_{k,i} (\overline{W}_{k,m} - A_{k,m})^2 | m_{k,i} = m, \overline{N}_{k,m} \geqslant 1] \leqslant \frac{E[A_{k,m}(1 - A_{k,m})]}{n_{k,m}}.
$$

Therefore,

$$
\begin{aligned}
&E[R_{k,i} W_{k,i} (\overline{W}_{k,m} - A_{k,m})^2 | m_{k,i} = m] \\
&\qquad = E[R_{k,i} W_{k,i} (\overline{W}_{k,m} - A_{k,m})^2 | m_{k,i} = m, \overline{N}_{k,m} \geqslant 1] \Pr(\overline{N}_{k,m} \geqslant 1 | m_{k,i} = m) \\
&\qquad \leqslant q_k \frac{E[A_{k,m}(1 - A_{k,m})]}{n_{k,m}}.
\end{aligned}
$$

Conditional on $\overline{N}_{k,m} = n$ and $A_{k,m}$, the variable $N_{k,m,1}$ has a binomial distribution with parameters $(n, A_{k,m})$. Then, using the formulas for the moments of a binomial distribution, we find that for any integer n, such that $1 \leqslant n \leqslant n_{k,m}$,

$$
E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i} (W_{k,i} - \overline{W}_{k,m})\right)^2 \middle| A_{k,m} = a, \overline{N}_{k,m} = n\right]
$$

$$
\begin{aligned}
&= E[(N_{k,m,1} - N_{k,m,1}^2/n)^2 | A_{k,m} = a, \overline{N}_{k,m} = n] \\
&= n^2 a^2 (1-a)^2 + na(1-a)(1-6a+6a^2) + r_1(a) + r_2(a)/n,
\end{aligned}
$$

where $|r_1(a)|$ and $|r_2(a)|$ are uniformly bounded in $a \in [0,1]$. Therefore,

$$
\begin{aligned}
&E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i} (W_{k,i} - \overline{W}_{k,m})\right)^2 \middle| \overline{N}_{k,m} = n\right] \\
&\qquad = n^2 E[A_{k,m}^2 (1 - A_{k,m})^2] + nE[A_{k,m}(1 - A_{k,m})(1 - 6A_{k,m} + 6A_{k,m}^2)] \\
&\qquad + E[r_1(A_{k,m})] + E[r_2(A_{k,m})]/n.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&E\left[\sum_{m=1}^{m_k} (\tau_{k,m} - \tau_k)^2 \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i} (W_{k,i} - \overline{W}_{k,m})\right)^2\right] \\
&\quad = \left(\sum_{m=1}^{m_k} (\tau_{k,m} - \tau_k)^2 (n_{k,m}(n_{k,m} - 1) p_k^2 q_k + n_{k,m} p_k q_k)\right) E[A_{k,m}^2 (1 - A_{k,m})^2] \\
&\quad + \sum_{m=1}^{m_k} (\tau_{k,m} - \tau_k)^2 n_{k,m} p_k q_k E[A_{k,m}(1 - A_{k,m})(1 - 6A_{k,m}(1 - A_{k,m}))] + \mathcal{O}(m_k q_k).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\frac{1}{n_k p_k q_k} E\left[\sum_{m=1}^{m_k} (\tau_{k,m} - \tau_k)^2 \left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i} (W_{k,i} - \overline{W}_{k,m})\right)^2\right] \\
&\quad \longrightarrow (E[A_{k,m}(1 - A_{k,m})] - (5 + p_k) E[A_{k,m}^2 (1 - A_{k,m})^2]) \sum_{m=1}^{m_k} \frac{n_{k,m}}{n_k} (\tau_{k,m} - \tau_k)^2 \\
&\quad + p_k E[A_{k,m}^2 (1 - A_{k,m})^2] \sum_{m=1}^{m_k} \frac{n_{k,m}^2}{n_k} (\tau_{k,m} - \tau_k)^2.
\end{aligned}
$$

Notice that,

$$
\begin{aligned}
&E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i} (W_{k,i} - \overline{W}_{k,m})\right)^4 \middle| A_{k,m} = a, \overline{N}_{k,m} = n\right] \\
&\qquad = E[(N_{k,m,1}(1 - N_{k,m,1}/n))^4 | A_{k,m} = a, \overline{N}_{k,m} = n] \\
&\qquad \leqslant E[N_{k,m,1}^4 | A_{k,m} = a, \overline{N}_{k,m} = n] \\
&\qquad \leqslant n^4,
\end{aligned}
$$

Therefore,

$$
E\left[\left(\sum_{i=1}^{n_k} 1\{m_{k,i} = m\} R_{k,i} W_{k,i} (W_{k,i} - \overline{W}_{k,m})\right)^4\right] = n_{k,m}^4 p_k^4 q_k \left(1 + \mathcal{O}\left(\frac{1}{p_k \min_m n_{k,m}}\right)\right),
$$

uniformly in m.

Suppose $X_{k,m} = (Z_{k,m,1} + Z_{k,m,2})^2$. Let $X_{k,m,1} = Z_{k,m,1}^2$ and $X_{k,m,2} = Z_{k,m,2}^2$. Now suppose,

$$\sum_{m=1}^{m_k} E[X_{k,m,1}^2] \longrightarrow 0,$$

and

$$\sum_{m=1}^{m_k} E[X_{k,m,2}^2] \longrightarrow 0.$$

Using the binomial theorem and Hölder's inequality, we obtain

$$\begin{aligned}
\sum_{m=1}^{m_k} E[X_{k,m}^2] &= \sum_{m=1}^{m_k} \sum_{p=0}^{4} c_p E[Z_{k,m,1}^p Z_{k,m,2}^{(4-p)}] \\
&\leqslant c \sum_{m=1}^{m_k} \sum_{p=0}^{4} E[|Z_{k,m,1}|^p |Z_{k,m,2}|^{(4-p)}] \\
&\leqslant c \sum_{m=1}^{m_k} \sum_{p=0}^{4} (E[X_{k,m,1}^2])^{p/4} (E[X_{k,m,2}^2])^{(4-p)/4} \\
&\leqslant c \sum_{p=0}^{4} \left(\sum_{m=1}^{m_k} E[X_{k,m,1}^2] \right)^{p/4} \left(\sum_{m=1}^{m_k} E[X_{k,m,2}^2] \right)^{(4-p)/4} \longrightarrow 0.
\end{aligned}$$