# Supplementary Material: Toward Better Practice of Covariate Adjustment in Analyzing Randomized Clinical Trials

Ting Ye[1], Jun Shao[2], Yanyao Yi[3] and Qingyuan Zhao[4]

## 1 Additional Results on Simulation

The table below provides the mean vector and covariance matrix of $(Y^{(1)}, U, W)$ in the real clinical trial used in §4.1.

| | mean | SD | correlation | |
|---|---|---|---|---|
| $Y^{(1)}$ | −1.031 | 1.126 | $Y^{(1)}$ and $U$ | −0.216 |
| $U$ | 5.684 | 0.953 | $Y^{(1)}$ and $W$ | −0.168 |
| $W$ | 23.222 | 13.422 | $U$ and $W$ | 0.744 |

The following Table S1-S2 are simulation results with $n = 200$ based on 10,000 simulations for linear contrasts and ratios, respectively.

[1]Department of Biostatistics, University of Washington.
[2]School of Statistics, East China Normal University; Department of Statistics, University of Wisconsin.
[3]Global Statistical Sciences, Eli Lilly and Company.
[4]Department of Pure Mathematics and Mathematical Statistics, University of Cambridge.
Corresponding to Dr. Jun Shao. Email: shao@stat.wisc.edu.

Table S1: Bias, standard deviation (SD), average standard error (SE), and coverage probability (CP) of 95% asymptotic confidence interval under simple randomization (SR), stratified permuted block randomization (PB), and Pocock-Simon's minimization based on 10,000 simulations and setup in (21) with $n = 200$

| | | | | $\theta_2 - \theta_1$ | | | | $\theta_3 - \theta_1$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Allocation | Randomization | Method | $X$ | Bias | SD | SE | CP | Bias | SD | SE | CP |
| 1:1:1 | SR | ANOVA | | -0.002 | 0.467 | 0.463 | 0.944 | 0.000 | 0.284 | 0.285 | 0.950 |
| | | ANCOVA | $Z$ | 0.000 | 0.445 | 0.433 | 0.941 | -0.001 | 0.326 | 0.322 | 0.949 |
| | | ANCOVA | $Z, U, W$ | -0.023 | 0.428 | 0.407 | 0.933 | 0.009 | 0.351 | 0.339 | 0.942 |
| | | ANHECOVA | $Z$ | 0.000 | 0.384 | 0.372 | 0.939 | 0.003 | 0.238 | 0.235 | 0.943 |
| | | ANHECOVA | $Z, U, W$ | 0.000 | 0.325 | 0.315 | 0.943 | 0.001 | 0.220 | 0.213 | 0.941 |
| | PB | ANOVA | | -0.002 | 0.380 | 0.462 | 0.980 | -0.001 | 0.241 | 0.284 | 0.977 |
| | | ANCOVA | $Z$ | -0.002 | 0.379 | 0.432 | 0.972 | -0.001 | 0.242 | 0.321 | 0.991 |
| | | ANCOVA | $Z, U, W$ | -0.026 | 0.356 | 0.406 | 0.970 | 0.009 | 0.275 | 0.338 | 0.983 |
| | | ANHECOVA | $Z$ | -0.002 | 0.377 | 0.371 | 0.940 | -0.001 | 0.240 | 0.234 | 0.940 |
| | | ANHECOVA | $Z, U, W$ | -0.002 | 0.317 | 0.314 | 0.948 | -0.001 | 0.220 | 0.213 | 0.941 |
| | Minimization | ANOVA | | 0.003 | 0.378 | 0.463 | 0.980 | 0.002 | 0.236 | 0.284 | 0.981 |
| | | ANCOVA | $Z$ | 0.003 | 0.378 | 0.432 | 0.972 | 0.002 | 0.237 | 0.321 | 0.991 |
| | | ANCOVA | $Z, U, W$ | -0.021 | 0.356 | 0.406 | 0.968 | 0.012 | 0.270 | 0.338 | 0.985 |
| | | ANHECOVA | $Z$ | 0.002 | 0.376 | 0.372 | 0.946 | 0.002 | 0.236 | 0.234 | 0.947 |
| | | ANHECOVA | $Z, U, W$ | 0.002 | 0.319 | 0.314 | 0.945 | 0.003 | 0.217 | 0.213 | 0.943 |
| 1:2:2 | SR | ANOVA | | 0.001 | 0.446 | 0.441 | 0.946 | 0.003 | 0.289 | 0.289 | 0.949 |
| | | ANCOVA | $Z$ | 0.004 | 0.430 | 0.417 | 0.942 | 0.002 | 0.347 | 0.339 | 0.946 |
| | | ANCOVA | $Z, U, W$ | -0.019 | 0.420 | 0.398 | 0.934 | 0.012 | 0.380 | 0.365 | 0.943 |
| | | ANHECOVA | $Z$ | 0.003 | 0.386 | 0.382 | 0.945 | 0.004 | 0.257 | 0.256 | 0.947 |
| | | ANHECOVA | $Z, U, W$ | 0.004 | 0.345 | 0.342 | 0.949 | 0.006 | 0.247 | 0.241 | 0.942 |
| | PB | ANOVA | | 0.002 | 0.379 | 0.441 | 0.977 | 0.000 | 0.254 | 0.289 | 0.971 |
| | | ANCOVA | $Z$ | 0.002 | 0.378 | 0.414 | 0.968 | 0.000 | 0.257 | 0.337 | 0.988 |
| | | ANCOVA | $Z, U, W$ | -0.024 | 0.365 | 0.395 | 0.961 | 0.008 | 0.296 | 0.362 | 0.982 |
| | | ANHECOVA | $Z$ | 0.001 | 0.377 | 0.381 | 0.951 | 0.000 | 0.253 | 0.255 | 0.948 |
| | | ANHECOVA | $Z, U, W$ | 0.002 | 0.336 | 0.341 | 0.948 | 0.001 | 0.243 | 0.240 | 0.944 |
| | Minimization | ANOVA | | 0.003 | 0.384 | 0.441 | 0.971 | 0.000 | 0.251 | 0.288 | 0.974 |
| | | ANCOVA | $Z$ | 0.001 | 0.383 | 0.414 | 0.961 | -0.002 | 0.252 | 0.336 | 0.991 |
| | | ANCOVA | $Z, U, W$ | -0.023 | 0.371 | 0.395 | 0.959 | 0.008 | 0.294 | 0.361 | 0.985 |
| | | ANHECOVA | $Z$ | 0.002 | 0.382 | 0.381 | 0.944 | -0.001 | 0.250 | 0.254 | 0.950 |
| | | ANHECOVA | $Z, U, W$ | 0.000 | 0.338 | 0.341 | 0.952 | -0.001 | 0.239 | 0.239 | 0.948 |

Table S2: Bias, standard deviation (SD), average standard error (SE), and coverage probability (CP) of 95% asymptotic confidence interval under simple randomization (SR), stratified permuted block randomization (PB), and Pocock-Simon's minimization based on 10,000 simulations and setup in (21) with $n = 200$

| | | | | $\theta_2/\theta_1$ | | | | $\theta_3/\theta_1$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Allocation | Randomization | Method | $X$ | Bias | SD | SE | CP | Bias | SD | SE | CP |
| 1:1:1 | SR | ANOVA | | 0.046 | 0.556 | 0.547 | 0.958 | 0.037 | 0.377 | 0.373 | 0.949 |
| | | ANCOVA | $Z$ | 0.048 | 0.546 | 0.526 | 0.952 | 0.049 | 0.440 | 0.427 | 0.948 |
| | | ANCOVA | $Z, U, W$ | 0.084 | 0.555 | 0.518 | 0.954 | 0.054 | 0.490 | 0.460 | 0.942 |
| | | ANHECOVA | $Z$ | 0.047 | 0.494 | 0.473 | 0.951 | 0.034 | 0.340 | 0.328 | 0.944 |
| | | ANHECOVA | $Z, U, W$ | 0.047 | 0.455 | 0.427 | 0.945 | 0.035 | 0.328 | 0.308 | 0.941 |
| | PB | ANOVA | | 0.048 | 0.483 | 0.544 | 0.980 | 0.038 | 0.339 | 0.371 | 0.969 |
| | | ANCOVA | $Z$ | 0.048 | 0.483 | 0.518 | 0.972 | 0.038 | 0.340 | 0.418 | 0.983 |
| | | ANCOVA | $Z, U, W$ | 0.083 | 0.480 | 0.509 | 0.972 | 0.043 | 0.393 | 0.451 | 0.973 |
| | | ANHECOVA | $Z$ | 0.048 | 0.481 | 0.471 | 0.955 | 0.038 | 0.338 | 0.327 | 0.942 |
| | | ANHECOVA | $Z, U, W$ | 0.048 | 0.438 | 0.426 | 0.950 | 0.037 | 0.324 | 0.307 | 0.942 |
| | Minimization | ANOVA | | 0.046 | 0.480 | 0.543 | 0.979 | 0.036 | 0.333 | 0.370 | 0.969 |
| | | ANCOVA | $Z$ | 0.046 | 0.479 | 0.517 | 0.971 | 0.036 | 0.335 | 0.417 | 0.981 |
| | | ANCOVA | $Z, U, W$ | 0.080 | 0.480 | 0.507 | 0.972 | 0.041 | 0.384 | 0.449 | 0.972 |
| | | ANHECOVA | $Z$ | 0.046 | 0.478 | 0.470 | 0.952 | 0.036 | 0.332 | 0.326 | 0.944 |
| | | ANHECOVA | $Z, U, W$ | 0.044 | 0.440 | 0.424 | 0.948 | 0.034 | 0.322 | 0.306 | 0.940 |
| 1:2:2 | SR | ANOVA | | 0.074 | 0.617 | 0.596 | 0.952 | 0.062 | 0.457 | 0.439 | 0.941 |
| | | ANCOVA | $Z$ | 0.095 | 0.678 | 0.621 | 0.940 | 0.094 | 0.605 | 0.540 | 0.933 |
| | | ANCOVA | $Z, U, W$ | 0.150 | 0.754 | 0.656 | 0.938 | 0.117 | 0.717 | 0.611 | 0.927 |
| | | ANHECOVA | $Z$ | 0.079 | 0.594 | 0.556 | 0.944 | 0.062 | 0.446 | 0.412 | 0.938 |
| | | ANHECOVA | $Z, U, W$ | 0.079 | 0.578 | 0.526 | 0.937 | 0.062 | 0.447 | 0.397 | 0.930 |
| | PB | ANOVA | | 0.073 | 0.562 | 0.593 | 0.968 | 0.062 | 0.423 | 0.437 | 0.956 |
| | | ANCOVA | $Z$ | 0.074 | 0.562 | 0.592 | 0.963 | 0.063 | 0.426 | 0.509 | 0.973 |
| | | ANCOVA | $Z, U, W$ | 0.132 | 0.626 | 0.624 | 0.962 | 0.088 | 0.531 | 0.576 | 0.963 |
| | | ANHECOVA | $Z$ | 0.074 | 0.561 | 0.550 | 0.955 | 0.063 | 0.423 | 0.408 | 0.943 |
| | | ANHECOVA | $Z, U, W$ | 0.075 | 0.543 | 0.521 | 0.948 | 0.062 | 0.421 | 0.394 | 0.938 |
| | Minimization | ANOVA | | 0.068 | 0.562 | 0.586 | 0.965 | 0.057 | 0.415 | 0.429 | 0.956 |
| | | ANCOVA | $Z$ | 0.083 | 0.572 | 0.589 | 0.962 | 0.071 | 0.430 | 0.505 | 0.974 |
| | | ANCOVA | $Z, U, W$ | 0.137 | 0.635 | 0.618 | 0.958 | 0.093 | 0.532 | 0.569 | 0.966 |
| | | ANHECOVA | $Z$ | 0.077 | 0.565 | 0.546 | 0.949 | 0.064 | 0.416 | 0.403 | 0.945 |
| | | ANHECOVA | $Z, U, W$ | 0.077 | 0.549 | 0.516 | 0.945 | 0.065 | 0.418 | 0.389 | 0.937 |

## 2 Two Lemmas

**Lemma 2.** *Assume (C1), (C2), and that* $P(A_i = a_t \mid Z_1, \ldots, Z_n) = \pi_t$ *for all* $t = 1, \ldots, k$ *and* $i = 1, \ldots, n$. *We have the following conclusions.*
*(i) For any integrable function* $f$,

$$E\{f(Y_i^{(t)}, X_i)\} = E(f(Y_i, X_i) \mid A_i = a_t)$$

*and*

$$E\{f(Y_i^{(t)}, X_i) \mid X_i\} = E(f(Y_i, X_i) \mid X_i, A_i = a_t).$$

*(ii) Let* $\theta = (E(Y^{(1)}, ..., E(Y^{(k)}))^\top$ *be the potential response mean vector,* $\beta = \sum_{t=1}^k \pi_t \beta_t$, *and* $\beta_t = \Sigma_X^{-1} \mathrm{cov}(X_i, Y_i^{(t)})$, $t = 1, ..., k$. *Then*

$$(\theta, \beta) = \arg\min_{(\vartheta, \not\beta)} E\left[\left\{Y_i - \vartheta^\top A_i - \not\beta^\top (X_i - \mu_X)\right\}^2\right]$$

*and*

$$(\theta, \beta_1, \ldots, \beta_k) = \arg\min_{(\vartheta, \not\beta_1, \ldots, \not\beta_k)} E\left[\left\{Y_i - \vartheta^\top A_i - \sum_{t=1}^k \not\beta_t^\top (X_i - \mu_X) I(A_i = a_t)\right\}^2\right].$$

The condition $P(A_i = a_t \mid Z_1, \ldots, Z_n) = \pi_t$ for all $t$ and $i$ holds for most covariate-adaptive randomization schemes. Note that it does not exclude the possibility that the set of random variables $\{A_i, i = 1, \ldots, n\}$ is dependent of $\{Z_i, i = 1, \ldots, n\}$, which is indeed the case for covariate-adaptive randomization schemes. We impose this condition only in Lemma 2 to facilitate understanding the working models. This additional assumption is not needed for our asymptotic theory in §3, as condition (C2) is sufficient.

*Proof.* (i) We focus on proving the second result; the first result can be shown similarly. For simple randomization, this result immediately follows (C2) (i) as $(Y_i^{(1)}, \ldots, Y_i^{(k)}, X_i, A_i)$ are independent and identically distributed. For covariate-adaptive randomization, we remark that the property of conditional independence (Dawid, 1979, Lemma 4.3), (C2) (i) and the third condition in Lemma 2 imply that $A_i$ is independent of $\{(Y_i^{(1)}, ..., Y_i^{(k)}, X_i, Z_i), i = 1, \ldots, n\}$. Then, it can be shown that

$$\begin{aligned}
E\{f(Y_i, X_i) \mid X_i, A_i = a_t\} &= E\{f(Y_i^{(t)}, X_i) \mid X_i, A_i = a_t\} \\
&= \sum_{z_1,\ldots,z_n \in \mathcal{Z}} E\{f(Y_i^{(t)}, X_i) \mid X_i, A_i = a_t, G_n\} P(G_n \mid X_i, A_i = a_t) \\
&= \sum_{z_1,\ldots,z_n \in \mathcal{Z}} E\{f(Y_i^{(t)}, X_i) \mid X_i, G_n\} P(G_n \mid X_i, A_i = a_t) \\
&= \sum_{z_1,\ldots,z_n \in \mathcal{Z}} E\{f(Y_i^{(t)}, X_i) \mid X_i, G_n\} P(G_n \mid X_i) \\
&= E\{f(Y_i^{(t)}, X_i) \mid X_i\},
\end{aligned}$$

where $G_n$ is the event that $\{Z_i = z_i, i = 1, ..., n\}$, and the equalities follow from the consistency of potential responses, the law of iterated expectation, (C2) (i), and the remark above.

(ii) We only prove the first result. The second result can be proved similarly. Let $(\theta, \beta)$ be the optimality points satisfying the following estimation equations:

$$E\big[I(A_i = a_t)\{Y_i - \theta^\top A_i - \beta^\top (X_i - \mu_X)\}\big] = 0, \qquad \text{for any } t \tag{S1}$$

$$E\big[(X_i - \mu_X)\{Y_i - \theta^\top A_i - \beta^\top (X_i - \mu_X)\}\big] = 0. \tag{S2}$$

From Lemma 2(i), (S1) implies that for any $t$,

$$E\big[Y_i - \theta^\top A_i - \beta^\top (X_i - \mu_X) \mid A_i = a_t\big] = E\big[Y_i^{(t)} - \theta_t - \beta^\top (X_i - \mu_X)\big] = E[Y_i^{(t)} - \theta_t] = 0$$

and, thus, $\theta_t = E(Y_i^{(t)})$, $t = 1, ..., k$. Then (S2) implies that

$$\begin{aligned}
0 =& E\big[(X_i - \mu_X)\{Y_i - \theta^\top A_i - \beta^\top (X_i - \mu_X)\}\big] \\
=& \sum_{t=1}^{k} E\big[I(A_i = a_t)(X_i - \mu_X)\{Y_i - \theta^\top A_i - \beta^\top (X_i - \mu_X)\}\big] \\
=& \sum_{t=1}^{k} E\big[(X_i - \mu_X)\{Y_i - \theta^\top A_i - \beta^\top (X_i - \mu_X)\} \mid A_i = a_t\big] \pi_t
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{t=1}^{k} E\big[(X_i - \mu_X)\{Y_i^{(t)} - \theta_t - \beta^\top (X_i - \mu_X)\}\big]\pi_t \\
&= \sum_{t=1}^{k} \left[\operatorname{cov}(X_i, Y_i^{(t)}) - \Sigma_X \beta\right] \pi_t \\
&= \sum_{t=1}^{k} \operatorname{cov}(X_i, Y_i^{(t)})\pi_t - \Sigma_X \beta
\end{aligned}
$$

and, thus, $\beta = \Sigma_X^{-1} \sum_{t=1}^{k} \operatorname{cov}(X_i, Y_i^{(t)})\pi_t = \sum_{t=1}^{k} \pi_t \beta_t$. □

**Lemma 3.** *Under conditions (C1)-(C2), for* $t = 1, \ldots, k$, $\hat{\beta}_t = \beta_t + o_p(1)$ *and* $\hat{\beta} = \beta + o_p(1)$*;*

*Proof.* (i) We prove the result for $\hat{\beta}_t$. The proof for $\hat{\beta}$ is analogous and omitted. Notice that

$$
\frac{1}{n_t} \sum_{i:A_i=a_t} (X_i - \bar{X}_t) Y_i = \frac{1}{n_t} \sum_{i=1}^{n} I(A_i = a_t) X_i Y_i - \frac{1}{n_t} \sum_{i=1}^{n} I(A_i = a_t) X_i \frac{1}{n_t} \sum_{i=1}^{n} I(A_i = a_t) Y_i
$$

Let $\mathcal{A} = \{A_1, \ldots, A_n\}$ and $\mathcal{F} = \{Z_1, \ldots, Z_n\}$. Note that

$$
\begin{aligned}
E\left\{\frac{1}{n} \sum_{i=1}^{n} I(A_i = a_t) X_i Y_i \mid \mathcal{A}, \mathcal{F}\right\} &= \frac{1}{n} \sum_{i=1}^{n} I(A_i = a_t) E(X_i Y_i^{(t)} \mid \mathcal{A}, \mathcal{F}) \\
&= \frac{1}{n} \sum_{i=1}^{n} I(A_i = a_t) E(X_i Y_i^{(t)} \mid Z_i),
\end{aligned}
$$

where the second line holds because $E(X_i Y_i^{(t)} \mid \mathcal{A}, \mathcal{F}) = E(X_i Y_i^{(t)} \mid \mathcal{F}) = E(X_i Y_i^{(t)} \mid Z_i)$ from (C1) and (C2) (i). Moreover, $n^{-1} \sum_{i=1}^{n} I(A_i = a_t) X_i Y_i^{(t)}$ is an average of independent random variables once conditional on $\{\mathcal{A}, \mathcal{F}\}$. From the existence of second moment of $XY^{(t)}$, and the weak law of large numbers for independent random variables, we conclude that, for any $\epsilon > 0$,

$$
\lim_{n \to \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^{n} I(A_i = a_t) X_i Y_i - \frac{1}{n} \sum_{i=1}^{n} I(A_i = a_t) E(X_i Y_i^{(t)} \mid Z_i)\right| \geq \epsilon \mid \mathcal{A}, \mathcal{F}\right) = 0
$$

From the bounded convergence theorem, the above equation also holds unconditionally. In other words,

$$\frac{1}{n}\sum_{i=1}^{n} I(A_i = a_t)X_iY_i - \frac{1}{n}\sum_{i=1}^{n} I(A_i = a_t)E(X_iY_i^{(t)} \mid Z_i) = o_p(1).$$

Furthermore,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} I(A_i = a_t)E(X_iY_i^{(t)} \mid Z_i) &= \frac{1}{n}\sum_{z}\sum_{i=1}^{n} I(Z_i = z)I(A_i = a_t)E(X_iY_i^{(t)} \mid Z_i = z) \\
&= \frac{1}{n}\sum_{z} E(X_iY_i^{(t)} \mid Z_i = z)\sum_{i=1}^{n} I(Z_i = z)I(A_i = a_t) \\
&= \frac{1}{n}\sum_{z} E(X_iY_i^{(t)} \mid Z_i = z)n_t(z) \\
&= \sum_{z} E(X_iY_i^{(t)} \mid Z_i = z)\frac{n_t(z)}{n(z)}\frac{n(z)}{n} \\
&= \sum_{z} E(X_iY_i^{(t)} \mid Z_i = z)\pi_t P(Z_i = z) + o_p(1) \\
&= \pi_t E(X_iY_i^{(t)}) + o_p(1)
\end{aligned}$$

This together with the fact that $n_t/n = \sum_z n_t(z)/\{\sum_z n(z)\} = \pi_t + o_p(1)$, we have

$$\frac{1}{n_t}\sum_{i=1}^{n} I(A_i = a_t)X_iY_i = E(X_iY_i^{(t)}) + o_p(1)$$

Similarly, we can show the result with $X_iY_i$ replaced by $X_i$ or $Y_i$ also holds, i.e.,

$$\begin{aligned}
\frac{1}{n_t}\sum_{i=1}^{n} I(A_i = a_t)X_i &= E(X_i) + o_p(1) \\
\frac{1}{n_t}\sum_{i=1}^{n} I(A_i = a_t)Y_i &= E(Y_i^{(t)}) + o_p(1)
\end{aligned}$$

The denominator of $\hat{\beta}_t$ can be treated similarly, which leads to

$$\frac{1}{n_t}\sum_{i:A_i=a_t} \{X_i - \bar{X}_t\}\{X_i - \bar{X}_t\}^\top = \Sigma_X + o_p(1).$$

The proof is completed by using the definition of $\beta_t$.

□

# 3 Technical Proofs

## 3.1 Proof of (9)

Under simple randomization, $A_1, \ldots, A_n$ are independent with other variables. Let $\bar{X}_{-t} = (n-n_t)^{-1}\sum_{i:A_i\neq a_t} X_i$. Then $\bar{X}_t - \bar{X} = n^{-1}(n-n_t)(\bar{X}_t - \bar{X}_{-t})$. Note that $\bar{X}_t$ and $\bar{X}_{-t}$ are uncorrelated conditional on $\mathcal{A} = (A_1, ..., A_n)$, as

$$\operatorname{cov}(\bar{X}_t, \bar{X}_{-t} \mid \mathcal{A}) = \frac{1}{(n-n_t)n_t}\sum_{i=1}^{n}\sum_{j=1}^{n} I(A_i = a_t)I(A_j \neq a_t)\operatorname{cov}(X_i, X_j \mid \mathcal{A}) = 0,$$

where the last equality is from $\operatorname{cov}(X_i, X_j \mid \mathcal{A}) = \operatorname{cov}(X_i, X_j) = 0$ for $i \neq j$. Similarly, we can show that $\bar{Y}_t$ and $\bar{X}_{-t}$ are uncorrelated conditional on $\mathcal{A}$.

Then,

$$\begin{aligned}
\operatorname{cov}\{\beta_t^\top(\bar{X}_t - \bar{X}), \bar{Y}_t\} &= \beta_t^\top \operatorname{cov}\left(\frac{n-n_t}{n}\bar{X}_t - \frac{n-n_t}{n}\bar{X}_{-t}, \bar{Y}_t\right) \\
&= \beta_t^\top E\left\{\operatorname{cov}\left(\frac{n-n_t}{n}\bar{X}_t, \bar{Y}_t \mid \mathcal{A}\right)\right\} \\
&= \beta_t^\top E\left\{\frac{n-n_t}{nn_t^2}\operatorname{cov}\left(\sum_{i:A_i=a_t} X_i, \sum_{i:A_i=a_t} Y_i \mid \mathcal{A}\right)\right\} \\
&= \beta_t^\top E\left\{\frac{n-n_t}{nn_t^2}\sum_{i:A_i=a_t}\operatorname{cov}\left(X_i, Y_i^{(t)}\right)\right\} \\
&= \beta_t^\top E\left\{\frac{n-n_t}{nn_t}\right\}\operatorname{cov}\left(X_i, Y_i^{(t)}\right) \\
&= E\left\{\frac{n-n_t}{nn_t}\right\}\beta_t^\top \Sigma_X \beta_t
\end{aligned}$$

where the second equality is from $\operatorname{cov}(\bar{X}_{-t}, \bar{Y}_t \mid \mathcal{A}) = 0$, $E(\bar{Y}_t \mid \mathcal{A}) = E(Y^{(t)})$ and the identity that $\operatorname{cov}(X, Y) = E\{\operatorname{cov}(X, Y \mid Z)\} + \operatorname{cov}\{E(X \mid Z), E(Y \mid Z)\}$. Also note that

$$\begin{aligned}
\operatorname{var}\{\beta_t^\top(\bar{X}_t - \bar{X})\} &= \beta_t^\top \operatorname{var}\left(\frac{n-n_t}{n}(\bar{X}_t - \bar{X}_{-t})\right)\beta_t \\
&= \beta_t^\top E\left(\frac{(n-n_t)^2}{n^2}\operatorname{var}(\bar{X}_t - \bar{X}_{-t} \mid \mathcal{A})\right)\beta_t
\end{aligned}$$

$$
\begin{aligned}
&= \beta_t^\top E\left(\frac{(n-n_t)^2}{n^2}\{\mathrm{var}(\bar{X}_t \mid \mathcal{A}) + \mathrm{var}(\bar{X}_{-t} \mid \mathcal{A})\}\right)\beta_t \\
&= \beta_t^\top E\left(\frac{(n-n_t)^2}{n^2}\left\{\frac{\mathrm{var}(X_i)}{n_t} + \frac{\mathrm{var}(X_i)}{n-n_t}\right\}\right)\beta_t \\
&= E\left\{\frac{n-n_t}{nn_t}\right\}\beta_t^\top \Sigma_X \beta_t
\end{aligned}
$$

where the second equality uses the identity that $\mathrm{var}(X) = E\{\mathrm{var}(X \mid Z)\} + \mathrm{var}\{E(X \mid Z)\}$, and $E\left(\bar{X}_t - \bar{X}_{-t} \mid \mathcal{A}\right) = E(X_i) - E(X_i) = 0$.

## 3.2 Proof of Lemma 1

For any fixed $k$-dimensional vector $\ell = (\ell_1, \ldots, \ell_k)^\top$, we have

$$
\begin{aligned}
& \ell^\top\{\mathrm{diag}(\pi_t^{-1} m_t^\top m_t) - M^\top M\}\ell \\
= \; & \sum_{t=1}^{k} \pi_t^{-1}\ell_t^2 m_t^\top m_t - \left\{\sum_{t=1}^{k} \ell_t m_t^\top\right\}\left\{\sum_{t=1}^{k} \ell_t m_t\right\} \\
= \; & E(Q^\top Q) - E(Q^\top)E(Q) \\
= \; & \mathrm{tr}\{E(QQ^\top)\} - \mathrm{tr}\{E(Q)E(Q^\top)\} \\
\geq \; & 0,
\end{aligned}
$$

where tr denotes the trace of a matrix, $Q$ denotes a $p$-dimensional random vector that takes value $\pi_t^{-1}\ell_t m_t$ with probability $\pi_t$, $t = 1, \ldots, k$, and the last equality follows from the fact that the covariance matrix $\mathrm{var}(Q) = E(QQ^\top) - E(Q)E(Q^\top)$ is positive semidefinite.

## 3.3 Proof of Theorem 1

(i) First, from $\bar{X}_t - \bar{X} = O_p(n^{-1/2})$ and $\hat{b}_t = b_t + o_p(1)$, we have

$$
\begin{aligned}
\hat{\theta}(\hat{b}_1, \ldots, \hat{b}_k) &= \hat{\theta}(b_1, \ldots, b_k) + \{(\bar{X}_1 - \bar{X})(b_1 - \hat{b}_1), \ldots, (\bar{X}_k - \bar{X})(b_k - \hat{b}_k)\}^\top \\
&= \hat{\theta}(b_1, \ldots, b_k) + o_p(n^{-1/2}).
\end{aligned}
$$

Write the sample average as $\mathbb{E}_n[\mu(X)] = n^{-1}\sum_{i=1}^n \mu(X_i)$. Then,

$$\bar{X} - \mu_X = \sum_{t=1}^k \frac{1}{n}\sum_{i=1}^n I(A_i = a_t)(X_i - \mu_X) = \sum_{t=1}^k \mathbb{E}_n\left[I(A = a_t)(X - \mu_X)\right],$$

and

$$\begin{aligned}
&\bar{Y}_t - \theta_t - (\bar{X}_t - \mu_X)^\top b_t \\
&= \frac{1}{n_t}\sum_{i=1}^n I(A_i = a_t)\left\{Y_i - \theta_t - (X_i - \mu_X)^\top b_t\right\} \\
&= \pi_t^{-1}\frac{1}{n}\sum_{i=1}^n I(A_i = a_t)\left\{Y_i - \theta_t - (X_i - \mu_X)^\top b_t\right\} \\
&\qquad + \left(\frac{1}{n_t/n} - \frac{1}{\pi_t}\right)\frac{1}{n}\sum_{i=1}^n I(A_i = a_t)\left\{Y_i - \theta_t - (X_i - \mu_X)^\top b_t\right\} \\
&= \pi_t^{-1}\mathbb{E}_n\left[I(A = a_t)\{Y - \theta_t - (X - \mu_X)^\top b_t\}\right] + o_p(n^{-1/2}),
\end{aligned}$$

where the last equality holds because $\mathbb{E}_n\left[I(A = a_t)\{Y - \theta_t - (X - \mu_X)^\top b_t\}\right] = O_p(n^{-1/2})$ from the central limit theorem, and $n/n_t - \pi_t^{-1} = o_p(1)$ from condition (C2) (ii). Hence, we can decompose $\hat{\theta}(b_1, \ldots, b_k)$ as

$$\begin{aligned}
&\hat{\theta}(b_1, \ldots, b_k) - \theta \\
&= \begin{pmatrix} \bar{Y}_1 - \theta_1 - (\bar{X}_1 - \mu_X)^\top b_1 \\ \vdots \\ \bar{Y}_k - \theta_k - (\bar{X}_k - \mu_X)^\top b_k \end{pmatrix} + \begin{pmatrix} b_1^\top(\bar{X} - \mu_X) \\ \vdots \\ b_k^\top(\bar{X} - \mu_X) \end{pmatrix} \\
&= \underbrace{\begin{pmatrix} \pi_1^{-1}\mathbb{E}_n\left[I(A = a_1)\{Y - \theta_1 - (X - \mu_X)^\top b_1\}\right] \\ \vdots \\ \pi_k^{-1}\mathbb{E}_n\left[I(A = a_k)\{Y - \theta_k - (X - \mu_X)^\top b_k\}\right] \end{pmatrix}}_{M_1} + \underbrace{\begin{pmatrix} b_1^\top \sum_{t=1}^k \mathbb{E}_n\left[I(A = a_t)(X - \mu_X)\right] \\ \vdots \\ b_k^\top \sum_{t=1}^k \mathbb{E}_n\left[I(A = a_t)(X - \mu_X)\right] \end{pmatrix}}_{M_2} \\
&\qquad + o_p(n^{-1/2})
\end{aligned}$$

$$
= \begin{pmatrix} \pi_1^{-1}a_1^\top & -\pi_1^{-1}b_1^\top & 0_p^\top & \cdots & 0_p^\top \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \pi_k^{-1}a_k^\top & 0_p^\top & 0_p^\top & \cdots & -\pi_k^{-1}b_k^\top \end{pmatrix}_{k\times(k+kp)} \underbrace{\begin{pmatrix} \mathbb{E}_n\left[I(A=a_1)(Y-\theta_1)\right] \\ \vdots \\ \mathbb{E}_n\left[I(A=a_k)(Y-\theta_k)\right] \\ \mathbb{E}_n\left[I(A=a_1)(X-\mu_X)\right] \\ \vdots \\ \mathbb{E}_n\left[I(A=a_k)(X-\mu_X)\right] \end{pmatrix}}_{V_{(k+kp)\times 1}}
$$

$$
+ \begin{pmatrix} b_1^\top & b_1^\top & \dots & b_1^\top \\ \vdots & \vdots & \ddots & \vdots \\ b_k^\top & b_k^\top & \dots & b_k^\top \end{pmatrix}_{k\times(kp)} \begin{pmatrix} \mathbb{E}_n\left[I(A=a_1)(X-\mu_X)\right] \\ \vdots \\ \mathbb{E}_n\left[I(A=a_k)(X-\mu_X)\right] \end{pmatrix}_{(kp)\times 1} + o_p(n^{-1/2}),
$$

where $a_t$ denotes the $k$-dimensional vector whose $t$th component is 1 and other components are 0, $p$ is the dimension of $X$, $0_p$ denotes a $p$-dimensional vector of zeros. From the central limit theorem, we have that the random vector $\sqrt{n}V$ is asymptotically normal with mean 0. This implies that $\sqrt{n}\{\hat{\theta}(b_1,\dots,b_k)-\theta\}$ is asymptotically normal with mean 0 from the Cramér-Wold device.

It remains to calculate the asymptotic variance of $\sqrt{n}\{\hat{\theta}(b_1,\dots,b_k)-\theta\}$. In the following, we consider $M_1$ and $M_2$ separately.

Consider $M_1$, where the $t$th component equals

$$
M_{1t} = \pi_t^{-1}\mathbb{E}_n\left[I(A=a_t)\{Y-\theta_t-(X-\mu_X)^\top b_t\}\right].
$$

We have that $(M_{1t}, t=1,\dots,k)$ are mutually independent and

$$
\mathrm{var}(M_{1t}) = (n\pi_t)^{-1}\mathrm{var}\{Y^{(t)} - X^\top b_t\},
$$

Hence, $\mathrm{var}(M_1)$ is a diagonal matrix, with the diagonal elements being $\mathrm{var}(M_{1t}), t = 1,\dots,k$. That is, $n\mathrm{var}(M_1) = \mathrm{diag}\{\pi_t^{-1}\mathrm{var}(Y^{(t)} - X^\top b_t)\}$.

Next, consider $M_2$, which can be reformulated as

$$M_2 = \begin{pmatrix} b_1^\top & & \\ & \ddots & \\ & & b_k^\top \end{pmatrix}_{k\times(kp)} \begin{pmatrix} \sum_{t=1}^k \mathbb{E}_n\left[I(A=a_t)(X-\mu_X)\right] \\ \cdots \\ \sum_{t=1}^k \mathbb{E}_n\left[I(A=a_t)(X-\mu_X)\right] \end{pmatrix}_{(kp)\times 1}$$

whose variance can be easily derived as $n\text{var}(M_2) = B^\top \Sigma_X B$.

Finally, consider $\text{cov}(M_1, M_2)$ whose $(t,s)$ element equals

$$\begin{aligned}
&\text{cov}\left\{\pi_t^{-1}\mathbb{E}_n\left[I(A=a_t)\{Y-\theta_t-(X-\mu_X)^\top b_t\}\right], b_s^\top \sum_{t=1}^k \mathbb{E}_n\left[I(A=a_t)(X-\mu_X)\right]\right\} \\
&= \text{cov}\left\{\pi_t^{-1}\mathbb{E}_n\left[I(A=a_t)\{Y-\theta_t-(X-\mu_X)^\top b_t\}\right], b_s^\top \mathbb{E}_n\left[I(A=a_t)(X-\mu_X)\right]\right\} \\
&= n^{-1}\pi_t^{-1}\text{cov}\left\{I(A=a_t)(Y-X^\top b_t)\}, b_s^\top I(A=a_t)(X-\mu_X)\right\} \\
&= n^{-1}\pi_t^{-1}E\left\{I(A=a_t)(Y-X^\top b_t)b_s^\top(X-\mu_X)\right\} \\
&= n^{-1}E\left\{(Y^{(t)}-X^\top b_t)b_s^\top(X-\mu_X)\right\} \\
&= n^{-1}\left\{\text{cov}(Y^{(t)}, b_s^\top X) - \text{cov}(X^\top b_t, b_s^\top X)\right\} \\
&= n^{-1}\{\beta_t^\top \Sigma_X b_s - b_t^\top \Sigma_X b_s\} \\
&= n^{-1}(\beta_t - b_t)^\top \Sigma_X b_s.
\end{aligned}$$

Thus, $n\text{cov}(M_1, M_2) = (\mathscr{B} - B)^\top \Sigma_X B$ and $n\text{cov}(M_2, M_1) = B^\top \Sigma_X(\mathscr{B} - B)$. Combining the above results, we conclude that $\sqrt{n}\{\hat{\theta}(b_1, \ldots, b_k) - \theta\}$ is asymptotically normal with mean 0 and variance $V_{\text{SR}}(B)$,

$$\begin{aligned}
V_{\text{SR}}(B) &= \text{diag}\{\pi_t^{-1}\text{var}(Y^{(t)} - X^\top b_t)\} + (\mathscr{B} - B)^\top \Sigma_X B + B^\top \Sigma_X(\mathscr{B} - B) + B^\top \Sigma_X B \\
&= \text{diag}\{\pi_t^{-1}\text{var}(Y^{(t)} - X^\top b_t)\} + \mathscr{B}^\top \Sigma_X B + B^\top \Sigma_X \mathscr{B} - B^\top \Sigma_X B.
\end{aligned}$$

(ii) Note that

$$\begin{aligned}
\text{var}(Y^{(t)} - b_t^\top X) =& \text{var}(Y^{(t)} - \beta_t^\top X + \beta_t^\top X - b_t^\top X) \\
=& \text{var}\{Y^{(t)} - \beta_t^\top X\} + \text{var}\{(\beta_t - b_t)^\top X\} + 2\text{cov}\{Y^{(t)} - \beta_t^\top X, (\beta_t - b_t)^\top X\}
\end{aligned}$$

$$=\text{var}\{Y^{(t)} - \beta_t^\top X\} + (\beta_t - b_t)^\top \Sigma_X (\beta_t - b_t).$$

Then simple algebra shows that

$$\begin{aligned}
& V_{\text{SR}}(B) - V_{\text{SR}}(\mathscr{B}) \\
&= \text{diag}\{\pi_t^{-1}\text{var}(Y^{(t)} - b_t^\top X)\} - \text{diag}\{\pi_t^{-1}\text{var}(Y^{(t)} - \beta_t^\top X)\} - (\mathscr{B} - B)^\top \Sigma_X (\mathscr{B} - B) \\
&= \text{diag}\{\pi_t^{-1}(\beta_t - b_t)^\top \Sigma_X (\beta_t - b_t)\} - (\mathscr{B} - B)^\top \Sigma_X (\mathscr{B} - B).
\end{aligned}$$

The rest follows from applying Lemma 1 with $M = \Sigma_X^{1/2}(\mathscr{B} - B)$.

## 3.4 Proof of Corollary 1

From Lemma 3, we know that $\hat{\beta} = \beta + o_p(1)$ and $\hat{\beta}_t = \beta_t + o_p(1)$, $t = 1, \ldots, k$. Let $\sigma_A^2, \sigma_B^2, \sigma_U^2$ respectively be the asymptotic variance of $\sqrt{n} c_{ts}^\top \hat{\theta}_{\text{ANHC}}$, $\sqrt{n} c_{ts}^\top \hat{\theta}_{\text{ANC}}$ and $\sqrt{n} c_{ts}^\top \hat{\theta}_{\text{AN}}$, where from Theorem 1,

$$\begin{aligned}
\sigma_A^2 =& \frac{\text{var}(Y^{(t)} - X^\top \beta_t)}{\pi_t} + \frac{\text{var}(Y^{(s)} - X^\top \beta_s)}{\pi_s} + (\beta_t - \beta_s)^\top \Sigma_X (\beta_t - \beta_s) \\
\sigma_B^2 =& \frac{\text{var}(Y^{(t)} - X^\top \beta)}{\pi_t} + \frac{\text{var}(Y^{(s)} - X^\top \beta)}{\pi_s} \\
\sigma_U^2 =& \frac{\text{var}(Y^{(t)})}{\pi_t} + \frac{\text{var}(Y^{(s)})}{\pi_s}
\end{aligned}$$

The results in Corollary 1(i) follows from

$$\begin{aligned}
& \sigma_A^2 - \sigma_U^2 \\
=& \frac{\beta_t^\top \Sigma_X \beta_t - 2\text{cov}(X^\top \beta_t, Y^{(t)})}{\pi_t} + \frac{\beta_s^\top \Sigma_X \beta_s - 2\text{cov}(X^\top \beta_s, Y^{(s)})}{\pi_s} + \{\beta_t - \beta_s\}^\top \Sigma_X \{\beta_t - \beta_s\} \\
=& \frac{\beta_t^\top \Sigma_X \beta_t - 2\beta_t^\top \Sigma_X \beta_t}{\pi_t} + \frac{\beta_s^\top \Sigma_X \beta_s - 2\beta_s^\top \Sigma_X \beta_s}{\pi_s} + \{\beta_t - \beta_s\}^\top \Sigma_X \{\beta_t - \beta_s\} \\
=& -\frac{\beta_t^\top \Sigma_X \beta_t}{\pi_t} - \frac{\beta_s^\top \Sigma_X \beta_s}{\pi_s} + \{\beta_t - \beta_s\}^\top \Sigma_X \{\beta_t - \beta_s\} \\
=& -\frac{\{\pi_s \beta_t + \pi_t \beta_s\}^\top \Sigma_X \{\pi_s \beta_t + \pi_t \beta_s\}}{\pi_t \pi_s (\pi_t + \pi_s)} - \{\beta_t - \beta_s\}^\top \Sigma_X \{\beta_t - \beta_s\} \left( \frac{1 - \pi_t - \pi_s}{\pi_t + \pi_s} \right)
\end{aligned}$$

where the second equality follows from $\beta_t = \Sigma_X^{-1}\text{cov}(X, Y^{(t)})$. This also proves that $\sigma_A^2 \leq \sigma_U^2$, because $\Sigma_X$ is positive definite and $\pi_t + \pi_s \leq 1$. If $\sigma_A^2 = \sigma_U^2$, then we must have $\pi_s\beta_t + \pi_t\beta_s = 0$ and $(1-\pi_t-\pi_s)\{\beta_t - \beta_s\} = 0$.

To show the results in Corollary 1(ii), notice that

$$\begin{aligned}\sigma_B^2 =& \frac{\text{var}\{Y^{(t)} - X^\top\beta_t + X^\top\beta_t - X^\top\beta\}}{\pi_t} + \frac{\text{var}\{Y^{(s)} - X^\top\beta_s + X^\top\beta_s - X^\top\beta\}}{\pi_s} \\ =& \frac{\text{var}\{Y^{(t)} - X^\top\beta_t\} + \text{var}\{X^\top\beta_t - X^\top\beta\}}{\pi_t} + \frac{\text{var}\{Y^{(s)} - X^\top\beta_s\} + \text{var}\{X^\top\beta_s - X^\top\beta\}}{\pi_s}\end{aligned}$$

where the second equality holds because

$$\begin{aligned}&\text{cov}\{Y^{(t)} - \beta_t^\top X, \beta_t^\top X - \beta^\top X\} = \text{cov}\{Y^{(t)} - \beta_t^\top X, X\}\{\beta_t - \beta\} \\ &= \{\text{cov}(Y^{(t)}, X) - \beta_t^\top\Sigma_X\}\{\beta_t - \beta\} = 0\end{aligned}$$

Then,

$$\sigma_A^2 - \sigma_B^2 = \{\beta_t - \beta_s\}^\top\Sigma_X\{\beta_t - \beta_s\} - \frac{\{\beta_t - \beta\}^\top\Sigma_X\{\beta_t - \beta\}}{\pi_t} - \frac{\{\beta_s - \beta\}^\top\Sigma\{\beta_s - \beta\}}{\pi_s}$$

In order to show that $\sigma_A^2 - \sigma_B^2 \leq 0$, we prove a stronger statement: it is true that for any $\tilde{\beta}$,

$$\{\beta_t - \beta_s\}^\top\Sigma_X\{\beta_t - \beta_s\} - \frac{\{\beta_t - \tilde{\beta}\}^\top\Sigma_X\{\beta_t - \tilde{\beta}\}}{\pi_t} - \frac{\{\beta_s - \tilde{\beta}\}^\top\Sigma_X\{\beta_s - \tilde{\beta}\}}{\pi_s} \leq 0. \tag{S3}$$

As a consequence, setting $\tilde{\beta}$ as $\beta = \sum_{t=1}^k \pi_t\beta_t$, the statement in (S3) also holds. This proves $\sigma_A^2 - \sigma_B^2 \leq 0$.

In what follows, we prove the claim in (S3). Note that the gradient of the left hand side of (S3) is

$$-2\left[\frac{\{\tilde{\beta} - \beta_t\}^\top\Sigma_X}{\pi_t} + \frac{\{\tilde{\beta} - \beta_s\}^\top\Sigma_X}{\pi_s}\right],$$

which equals zero when $\tilde{\beta} = \{\pi_s\beta_t + \pi_t\beta_s\}/(\pi_t + \pi_s)$. This is also the unique solution from the positive definiteness of $\Sigma_X$. It is also easy to see that the Hessian of the left hand side of (S3) is negative definite, which means that $\tilde{\beta} = \{\pi_s\beta_t + \pi_t\beta_s\}/(\pi_t + \pi_s)$ is the global and unique maximizer of the left hand side of (S3). The statement in (S3) is true because when evaluated at $\tilde{\beta} = \{\pi_s\beta_t + \pi_t\beta_s\}/(\pi_t + \pi_s)$, the left hand side of (S3) equals

$$\begin{aligned}
&\{\beta_t - \beta_s\}^\top \Sigma_X \{\beta_t - \beta_s\} \\
&- \left\{\beta_t - \frac{\pi_s\beta_t + \pi_t\beta_s}{\pi_t + \pi_s}\right\}^\top \Sigma_X \left\{\beta_t - \frac{\pi_s\beta_t + \pi_t\beta_s}{\pi_t + \pi_s}\right\} \frac{1}{\pi_t} \\
&- \left\{\beta_s - \frac{\pi_s\beta_t + \pi_t\beta_s}{\pi_t + \pi_s}\right\}^\top \Sigma_X \left\{\beta_s - \frac{\pi_s\beta_t + \pi_t\beta_s}{\pi_t + \pi_s}\right\} \frac{1}{\pi_s} \\
= &- \{\beta_t(z) - \beta_s(z)\}^\top \Sigma_X \{\beta_t(z) - \beta_s(z)\} \left(\frac{1 - \pi_t - \pi_s}{\pi_t + \pi_s}\right) \leq 0
\end{aligned}$$

This completes the proof for $\sigma_A^2 \leq \sigma_B^2$, where the equality holds if and only if $\{\beta_t - \beta_s\}(1 - \pi_t - \pi_s) = 0$ and $\sum_{t=1}^k \pi_t\beta_t = \{\pi_s\beta_t + \pi_t\beta_s\}/(\pi_t + \pi_s)$.

## 3.5 Proof of Theorem 2

First, from $\bar{X}_t - \bar{X} = O_p(n^{-1/2})$ and $\hat{\beta}_t = \beta_t + o_p(1)$ from Lemma 3, we have $\hat{\theta}(\hat{\beta}_1, \ldots, \hat{\beta}_k) = \hat{\theta}(\beta_1, \ldots, \beta_k) + o_p(n^{-1/2})$. By using the definition $\beta_t = \Sigma_X^{-1}\text{cov}(X_i, Y_i^{(t)})$, we have

$$E\big[X_i^\top\{Y_i^{(t)} - \theta_t - (X_i - \mu_X)^\top\beta_t\}\big] = \text{cov}(X_i, Y_i^{(t)}) - \text{cov}(X_i, Y_i^{(t)}) = 0.$$

Because $Z_i$ is discrete and $X_i$ contains all joint levels of $Z_i$ as a sub-vector, according to the estimation equations from the least squares, we have that

$$E\left[I(Z_i = z)\{Y_i^{(t)} - \theta_t - (X_i - \mu_X)^\top\beta_t\}\right] = 0,\ \forall z \in \mathcal{Z},$$

and thus,

$$E\left\{Y_i^{(t)} - \theta_t - (X_i - \mu_X)^\top\beta_t \mid Z_i\right\} = 0,\ \text{a.s.}. \tag{S4}$$

Moreover, recall that $\mathcal{A} = \{A_1, \ldots, A_n\}$ and $\mathcal{F} = \{Z_1, \ldots, Z_n\}$, then

$$E\left\{\bar{Y}_t - \bar{X}_t^\top\beta_t \mid \mathcal{A}, \mathcal{F}\right\} = E\left\{\frac{\sum_{i=1}^n I(A_i = a_t)(Y_i^{(t)} - X_i^\top\beta_t)}{n_t} \mid \mathcal{A}, \mathcal{F}\right\}$$

$$= \frac{\sum_{i=1}^{n} I(A_i = a_t) E\left\{Y_i^{(t)} - X_i^\top \beta_t \mid Z_i\right\}}{n_t} = \theta_t - \mu_X^\top \beta_t, \text{ a.s..}$$

This implies that $\bar{Y}_t - \theta_t - (\bar{X}_t - \mu_X)^\top \beta_t = \bar{Y}_t - \bar{X}_t^\top \beta_t - E(\bar{Y}_t - \bar{X}_t^\top \beta_t \mid \mathcal{A}, \mathcal{F})$ a.s..

We decompose $\hat{\theta}(\beta_1, \ldots, \beta_k)$ as

$$\hat{\theta}(\beta_1, \ldots, \beta_k) - \theta = \begin{pmatrix} \bar{Y}_1 - \theta_1 - (\bar{X}_1 - \mu_X)^\top \beta_1 \\ \cdots \\ \bar{Y}_k - \theta_k - (\bar{X}_k - \mu_X)^\top \beta_k \end{pmatrix} + \begin{pmatrix} \beta_1^\top (\bar{X} - \mu_X) \\ \cdots \\ \beta_k^\top (\bar{X} - \mu_X) \end{pmatrix}$$

$$= \underbrace{\begin{pmatrix} \bar{Y}_1 - \theta_1 - (\bar{X}_1 - \mu_X)^\top \beta_1 \\ \cdots \\ \bar{Y}_k - \theta_k - (\bar{X}_k - \mu_X)^\top \beta_k \end{pmatrix}}_{M_1} + \underbrace{\begin{pmatrix} \beta_1^\top (\bar{X} - E(\bar{X} \mid \mathcal{A}, \mathcal{F})) \\ \cdots \\ \beta_k^\top (\bar{X} - E(\bar{X} \mid \mathcal{A}, \mathcal{F})) \end{pmatrix}}_{M_{21}} + \underbrace{\begin{pmatrix} \beta_1^\top (E(\bar{X} \mid \mathcal{A}, \mathcal{F}) - \mu_X) \\ \cdots \\ \beta_k^\top (E(\bar{X} \mid \mathcal{A}, \mathcal{F}) - \mu_X) \end{pmatrix}}_{M_{22}}$$

$$= \begin{pmatrix} a_1^\top & -\beta_1^\top & 0_p^\top & \cdots & 0_p^\top \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_k^\top & 0_p^\top & 0_p^\top & \cdots & -\beta_k^\top \end{pmatrix} \underbrace{\begin{pmatrix} \bar{Y}_1 - E(\bar{Y}_1 \mid \mathcal{A}, \mathcal{F}) \\ \cdots \\ \bar{Y}_k - E(\bar{Y}_k \mid \mathcal{A}, \mathcal{F}) \\ \bar{X}_1 - E(\bar{X}_1 \mid \mathcal{A}, \mathcal{F}) \\ \cdots \\ \bar{X}_k - E(\bar{X}_k \mid \mathcal{A}, \mathcal{F}) \end{pmatrix}}_{\tilde{V}}$$

$$+ \begin{pmatrix} \beta_1^\top & & \\ & \ddots & \\ & & \beta_k^\top \end{pmatrix} \begin{pmatrix} n^{-1}n_1 I_p & n^{-1}n_2 I_p & \ldots & n^{-1}n_k I_p \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1}n_1 I_p & n^{-1}n_2 I_p & \ldots & n^{-1}n_k I_p \end{pmatrix} \begin{pmatrix} \bar{X}_1 - E(\bar{X}_1 \mid \mathcal{A}, \mathcal{F}) \\ \vdots \\ \bar{X}_k - E(\bar{Y}_k \mid \mathcal{A}, \mathcal{F}) \end{pmatrix} + M_{22}.$$

Conditioned on $\mathcal{A}, \mathcal{F}$, every component in $\tilde{V}$ is an average of independent terms. We verify at the end of this section that Lindeberg's condition holds for $(\bar{Y}_t - E(\bar{Y}_t \mid \mathcal{A}, \mathcal{F}), \bar{X}_t^\top - E(\bar{X}_t^\top \mid \mathcal{A}, \mathcal{F}))c$ for any $(p+1)$-dimensional real-valued vector $c$. From Lindeberg's Central Limit Theorem, as $n \to \infty$, $\sqrt{n}(\bar{Y}_t - E(\bar{Y}_t \mid \mathcal{A}, \mathcal{F}), \bar{X}_t^\top - E(\bar{X}_t^\top \mid \mathcal{A}, \mathcal{F}))c$ is asymptotically normal with mean 0 conditional on $\mathcal{A}, \mathcal{F}$. From the Cramér-Wold device, we have that $\sqrt{n}(\bar{Y}_t - E(\bar{Y}_t \mid \mathcal{A}, \mathcal{F}), \bar{X}_t^\top - E(\bar{X}_t^\top \mid \mathcal{A}, \mathcal{F}))^\top$ is asymptotically normal with mean 0 conditional on $\mathcal{A}, \mathcal{F}$. As $\{(\bar{Y}_t - E(\bar{Y}_t \mid \mathcal{A}, \mathcal{F}), \bar{X}_t^\top - E(\bar{X}_t^\top \mid \mathcal{A}, \mathcal{F}))^\top, t = 1, \ldots, k\}$ are mutually independent conditional on $\mathcal{A}, \mathcal{F}$, it immediately follows that $\sqrt{n}\tilde{V}$ and thus $\sqrt{n}(M_1 + M_{21})$ are asymptotically normal with mean 0 conditional on $\mathcal{A}, \mathcal{F}$.

Next, we calculate the variance. For $M_1$, the variance of its $t$th component is

$$\begin{aligned} n\text{var}(M_{1t} \mid \mathcal{A}, \mathcal{F}) &= \frac{n}{n_t^2}\text{var}\left\{\sum_{i:A_i=a_t} Y_i^{(t)} - (X_i - \mu_X)^\top \beta_t \mid \mathcal{A}, \mathcal{F}\right\} \\ &= \frac{n}{n_t^2}\sum_{i:A_i=a_t} \text{var}\left\{Y_i^{(t)} - (X_i - \mu_X)^\top \beta_t \mid Z_i\right\} \\ &= \frac{n}{n_t^2}\sum_z \sum_{i:A_i=a_t, Z_i=z} \text{var}\left\{Y_i^{(t)} - (X_i - \mu_X)^\top \beta_t \mid Z_i = z\right\} \\ &= \frac{n}{n_t}\sum_z \frac{n_t(z)}{n_t}\text{var}\left\{Y_i^{(t)} - (X_i - \mu_X)^\top \beta_t \mid Z_i = z\right\} \\ &= \frac{1}{\pi_t}\sum_z P(Z_i = z)\text{var}\left\{Y_i^{(t)} - (X_i - \mu_X)^\top \beta_t \mid Z_i = z\right\} + o_p(1) \\ &= \frac{1}{\pi_t}E\left[\text{var}\left\{Y_i^{(t)} - (X_i - \mu_X)^\top \beta_t \mid Z_i\right\}\right] + o_p(1), \end{aligned}$$

where the second line and the fifth line are respectively from (C2) (i) and (C2) (ii). Moreover, $M_{1t}$ and $M_{1s}$ are independent conditional on $\mathcal{A}, \mathcal{F}$, for $t \neq s$. Hence,

$$\text{var}(\sqrt{n}M_1 \mid \mathcal{A}, \mathcal{F}) = \text{diag}\left\{\pi_t^{-1}E\left[\text{var}\{Y_i^{(t)} - (X_i - \mu_X)^\top \beta_t \mid Z_i\}\right]\right\} + o_p(1), \tag{S5}$$

which does not depend on the randomization scheme. For $M_{21}$, we have that

$$n\text{var}(\bar{X} - E(\bar{X} \mid \mathcal{A}, \mathcal{F}) \mid \mathcal{A}, \mathcal{F}) = \frac{1}{n}\sum_{i=1}^n \text{var}(X_i \mid Z_i) = E\{\text{var}(X_i \mid Z_i)\} + o_p(1)$$

$$n\text{var}(M_{21} \mid \mathcal{A}, \mathcal{F}) = \mathscr{B}^\top E\{\text{var}(X_i \mid Z_i)\}\mathscr{B} + o_p(1).$$

For the covariance, consider $n\text{cov}(M_1, M_{21} \mid \mathcal{A}, \mathcal{F})$ whose $(t, s)$ element equals

$$n\text{cov}(M_{1t}, \bar{X}^\top \beta_s \mid \mathcal{A}, \mathcal{F}) \tag{S6}$$

$$
\begin{aligned}
&= n\mathrm{cov}\left(\bar{Y}_t - \bar{X}_t^\top \beta_t, \sum_{j=1}^{k} \frac{n_j}{n} \bar{X}_j^\top \beta_s \mid \mathcal{A}, \mathcal{F}\right) \\
&= n\mathrm{cov}\left(\bar{Y}_t - \bar{X}_t^\top \beta_t, \frac{n_t}{n} \bar{X}_t^\top \beta_s \mid \mathcal{A}, \mathcal{F}\right) \\
&= \frac{1}{n_t} \sum_{i:A_i=a_t} \mathrm{cov}\left(Y_i^{(t)} - X_i^\top \beta_t, X_i^\top \beta_s \mid Z_i\right) \\
&= \frac{1}{n_t} \sum_{i:A_i=a_t} \sum_{z\in\mathcal{Z}} I(Z_i = z)\mathrm{cov}\left(Y_i^{(t)} - X_i^\top \beta_t, X_i^\top \beta_s \mid Z_i = z\right) \\
&= \sum_{z\in\mathcal{Z}} \frac{n_t(z)}{n_t} \mathrm{cov}\left(Y_i^{(t)} - X_i^\top \beta_t, X_i^\top \beta_s \mid Z_i = z\right) \\
&= \sum_{z\in\mathcal{Z}} P(Z = z)\mathrm{cov}\left(Y_i^{(t)} - X_i^\top \beta_t, X_i^\top \beta_s \mid Z_i = z\right) + o_p(1) \\
&= E\left\{\mathrm{cov}\left(Y_i^{(t)} - X_i^\top \beta_t, X_i^\top \beta_s \mid Z_i\right)\right\} + o_p(1) \\
&= o_p(1),
\end{aligned}
$$

where the last equality holds because $E(Y_i^{(t)} - X_i^\top \beta_t \mid Z_i) = \theta_t - \mu_X^\top \beta_t$ and, thus, $\mathrm{cov}\{E(Y_i^{(t)} - X_i^\top \beta_t \mid Z_i), E(X_i^\top \beta_s \mid Z_i)\} = 0$ and $E\{\mathrm{cov}(Y_i^{(t)} - X_i^\top \beta_t, X_i^\top \beta_s \mid Z_i)\} = \mathrm{cov}(Y_i^{(t)} - X_i^\top \beta_t, X_i^\top \beta_s) = 0$ according to the definition of $\beta_t$.

Combining the above derivations and from the Slutsky's theorem, we have shown that

$$
\begin{aligned}
&\sqrt{n}(M_1 + M_{21}) \mid \mathcal{A}, \mathcal{F} \\
&\xrightarrow{d} N\left(0, \mathrm{diag}\left\{\pi_t^{-1} E\left[\mathrm{var}\{Y_i^{(t)} - (X_i - \mu_X)^\top \beta_t \mid Z_i\}\right]\right\} + \mathscr{B}^\top E\{\mathrm{var}(X_i \mid Z_i)\}\mathscr{B}\right).
\end{aligned}
$$

From the bounded convergence theorem, this result also holds unconditionally, i.e.,

$$
\begin{aligned}
&\sqrt{n}(M_1 + M_{21}) \\
&\xrightarrow{d} N\left(0, \mathrm{diag}\left\{\pi_t^{-1} E\left[\mathrm{var}\{Y_i^{(t)} - (X_i - \mu_X)^\top \beta_t \mid Z_i\}\right]\right\} + \mathscr{B}^\top E\{\mathrm{var}(X_i \mid Z_i)\}\mathscr{B}\right).
\end{aligned}
$$

Moreover, since $E(\bar{X} \mid \mathcal{A}, \mathcal{F})$ is an average of identically and independently distributed terms, by the central limit theorem,

$$
\sqrt{n}\{E(\bar{X} \mid \mathcal{A}, \mathcal{F}) - \mu_X\} = n^{-1/2} \sum_{i=1}^{n} \{E(X_i \mid Z_i) - \mu_X\} \xrightarrow{d} N(0, \mathrm{var}(E(X_i \mid Z_i))),
$$

and

$$\sqrt{n}M_{22} \xrightarrow{d} N(0, \mathscr{B}^\top \mathrm{var}(E(X_i \mid Z_i))\mathscr{B}).$$

Next, we show that $(\sqrt{n}(M_1 + M_{21}), \sqrt{n}M_{22}) \xrightarrow{d} (\xi_1, \xi_2)$, where $(\xi_1, \xi_2)$ are mutually independent. This can be seen from

$$\begin{aligned}
&P(\sqrt{n}(M_1 + M_{21}) \le t_1, \sqrt{n}M_{22} \le t_2) \\
&= E\{I(\sqrt{n}(M_1 + M_{21}) \le t_1)I(\sqrt{n}M_{22} \le t_2)\} \\
&= E\{P(\sqrt{n}(M_1 + M_{21}) \le t_1 \mid \mathcal{A}, \mathcal{F})I(\sqrt{n}M_{22} \le t_2)\} \\
&= E\{\{P(\sqrt{n}(M_1 + M_{21}) \le t_1 \mid \mathcal{A}, \mathcal{F}) - P(\xi_1 \le t_1)\}I(\sqrt{n}M_{22} \le t_2)\} \\
&\qquad + P(\xi_1 \le t_1)P(\sqrt{n}M_{22} \le t_2) \\
&\to P(\xi_1 \le t_1)P(\xi_2 \le t_2),
\end{aligned}$$

where the last step follows from the bounded convergence theorem.

Finally, from $\sqrt{n}\{\hat{\theta}(\beta_1, \ldots, \beta_k) - \theta\} = \sqrt{n}(M_1 + M_{21} + M_{22})$, we have

$$\begin{aligned}
&\sqrt{n}\{\hat{\theta}(\beta_1, \ldots, \beta_k) - \theta\} \\
&\xrightarrow{d} N\left(0, \mathrm{diag}\left\{\pi_t^{-1}E\left[\mathrm{var}\{Y_i^{(t)} - (X_i - \mu_X)^\top \beta_t \mid Z_i\}\right]\right\} + \mathscr{B}^\top \Sigma_X \mathscr{B}\right).
\end{aligned}$$

Note that we have also shown that the asymptotic distribution of $\sqrt{n}\{\hat{\theta}(\beta_1, \ldots, \beta_k) - \theta\}$ is invariant under randomization schemes satisfying (C2). The above asymptotic distribution of $\sqrt{n}\{\hat{\theta}(\beta_1, \ldots, \beta_k) - \theta\}$ is the same as the result in Theorem 2 because $E\{Y_i^{(t)} - \theta_t - (X_i - \mu_X)^\top \beta_t \mid Z_i\} = 0$ a.s., and thus, $E[\mathrm{var}\{Y_i^{(t)} - (X_i - \mu_X)^\top \beta_t \mid Z_i\}] = \mathrm{var}(Y_i^{(t)} - X_i^\top \beta_t)$.

We conclude the proof by verifing the Lindeberg's condition for $(\bar{Y}_t - E(\bar{Y}_t \mid \mathcal{A}, \mathcal{F}), \bar{X}_t^\top - E(\bar{X}_t^\top \mid \mathcal{A}, \mathcal{F}))c$ conditional on $\mathcal{A}, \mathcal{F}$, which we rearrange and write as

$$\sum_{i=1}^n K_i, \text{ where } K_i = \frac{1}{n_t} I(A_i = a_t)\big(Y_i^{(t)} - E(Y_i^{(t)} \mid Z_i), X_i^\top - E(X_i^\top \mid Z_i)\big)c.$$

From $E(K_i \mid \mathcal{A}, \mathcal{F}) = 0$ and

$$\mathrm{var}(K_i \mid \mathcal{A}, \mathcal{F}) = \frac{1}{n_t^2} I(A_i = a_t) c^\top \mathrm{var}\left( (Y_i^{(t)}, X_i^\top)^\top \mid Z_i \right) c.$$

Then, the Lindeberg's condition holds with probability 1 because for any $\epsilon > 0$,

$$\begin{aligned}
&\sum_{i=1}^n E\left[ \frac{K_i^2}{\mathrm{var}(\sum_{i=1}^n K_i \mid \mathcal{A}, \mathcal{F})} I\left\{ \frac{K_i^2}{\mathrm{var}(\sum_{i=1}^n K_i \mid \mathcal{A}, \mathcal{F})} > \epsilon \right\} \mid \mathcal{A}, \mathcal{F} \right] \\
&= \sum_{i=1}^n \frac{\mathrm{var}(K_i \mid \mathcal{A}, \mathcal{F})}{\mathrm{var}(\sum_{i=1}^n K_i \mid \mathcal{A}, \mathcal{F})} E\left[ \frac{K_i^2}{\mathrm{var}(K_i \mid \mathcal{A}, \mathcal{F})} I\left\{ \frac{K_i^2}{\mathrm{var}(\sum_{i=1}^n K_i \mid \mathcal{A}, \mathcal{F})} > \epsilon \right\} \mid \mathcal{A}, \mathcal{F} \right] \\
&\le \max_i E\left[ \frac{K_i^2}{\mathrm{var}(K_i \mid \mathcal{A}, \mathcal{F})} I\left\{ \frac{K_i^2}{\mathrm{var}(K_i \mid \mathcal{A}, \mathcal{F})} > \epsilon \frac{\mathrm{var}(\sum_{i=1}^n K_i \mid \mathcal{A}, \mathcal{F})}{\mathrm{var}(K_i \mid \mathcal{A}, \mathcal{F})} \right\} \mid \mathcal{A}, \mathcal{F} \right] \\
&= o(1)
\end{aligned}$$

where the third line is because $\sum_{i=1}^n \mathrm{var}(K_i \mid \mathcal{A}, \mathcal{F}) = \mathrm{var}(\sum_{i=1}^n K_i \mid \mathcal{A}, \mathcal{F})$, and the last line is because $K_i/\sqrt{\mathrm{var}(K_i \mid \mathcal{A}, \mathcal{F})}$ has zero expectation and unit variance, and that $\max_i \mathrm{var}(K_i \mid \mathcal{A}, \mathcal{F})/\mathrm{var}(\sum_{i=1}^n K_i \mid \mathcal{A}, \mathcal{F}) \le C \max(\{n_t(z)\}^{-1}, z \in \mathcal{Z}) = o(1)$ with probability 1, where $C$ is a generic constant.

## 3.6 Proof of Theorem 3

(i) First, from $\bar{X}_t - \bar{X} = O_p(n^{-1/2})$ and $\hat{b}_t = b_t + o_p(1)$, we have $\hat{\theta}(\hat{b}_1, \ldots, \hat{b}_k) = \hat{\theta}(b_1, \ldots, b_k) + o_p(n^{-1/2})$. Also note that

$$\begin{aligned}
&E(\bar{Y}_t - \theta_t - (\bar{X}_t - \mu_X)^\top b_t \mid \mathcal{A}, \mathcal{F}) \\
&= E\left( \frac{\sum_{i=1}^n I(A_i = a_t)(Y_i^{(t)} - \theta_t - (X_i - \mu_X)^\top b_t)}{n_t} \mid \mathcal{A}, \mathcal{F} \right) \\
&= \frac{\sum_{i=1}^n (I(A_i = a_t) - \pi_t) E(Y_i^{(t)} - \theta_t - (X_i - \mu_X)^\top b_t \mid Z_i)}{n_t} \\
&\quad + \frac{\pi_t}{n_t} \sum_{i=1}^n E(Y_i^{(t)} - \theta_t - (X_i - \mu_X)^\top b_t \mid Z_i) \\
&= \sum_{z \in \mathcal{Z}} \left( \frac{n_t(z)}{n(z)} - \pi_t \right) E(Y_i^{(t)} - \theta_t - (X_i - \mu_X)^\top b_t \mid Z_i = z) \frac{n(z)}{n_t}
\end{aligned}$$

$$
\begin{aligned}
&+ \frac{\pi_t}{n_t} \sum_{i=1}^{n} E(Y_i^{(t)} - \theta_t - (X_i - \mu_X)^\top b_t \mid Z_i) \\
=& \sum_{z \in \mathcal{Z}} \left( \frac{n_t(z)}{n(z)} - \pi_t \right) E(Y_i^{(t)} - \theta_t - (X_i - \mu_X)^\top b_t \mid Z_i = z) P(Z = z) \pi_t^{-1} \\
&+ \frac{1}{n} \sum_{i=1}^{n} E(Y_i^{(t)} - \theta_t - (X_i - \mu_X)^\top b_t \mid Z_i) + o_p(n^{-1/2}),
\end{aligned}
$$

where the last equality is from $n(z)/n = P(Z=z) + o_p(1)$, $n_t/n = \pi_t + o_p(1)$, $\left(\frac{n_t(z)}{n(z)} - \pi_t\right) = O_p(n^{-1/2})$ due to condition (C3), and $n^{-1}\sum_{i=1}^{n} E(Y_i^{(t)} - \theta_t - (X_i - \mu_X)^\top b_t \mid Z_i) = O_p(n^{-1/2})$.

Thus, we can decompose $\hat{\theta}(b_1, \dots, b_k)$ as

$$
\begin{aligned}
&\hat{\theta}(b_1, \dots, b_k) - \theta \\
&= \begin{pmatrix} \bar{Y}_1 - \theta_1 - (\bar{X}_1 - \mu_X)^\top b_1 \\ \cdots \\ \bar{Y}_k - \theta_k - (\bar{X}_k - \mu_X)^\top b_k \end{pmatrix} + \begin{pmatrix} b_1^\top (\bar{X} - \mu_X) \\ \cdots \\ b_k^\top (\bar{X} - \mu_X) \end{pmatrix} \\
&= \underbrace{\begin{pmatrix} \bar{Y}_1 - E(\bar{Y}_1 \mid \mathcal{A}, \mathcal{F}) - (\bar{X}_1 - E(\bar{X}_1 \mid \mathcal{A}, \mathcal{F}))^\top b_1 \\ \cdots \\ \bar{Y}_k - E(\bar{Y}_k \mid \mathcal{A}, \mathcal{F}) - (\bar{X}_k - E(\bar{X}_k \mid \mathcal{A}, \mathcal{F}))^\top b_k \end{pmatrix}}_{M_{11}} + \underbrace{\begin{pmatrix} b_1^\top (\bar{X} - E(\bar{X} \mid \mathcal{A}, \mathcal{F})) \\ \cdots \\ b_k^\top (\bar{X} - E(\bar{X} \mid \mathcal{A}, \mathcal{F})) \end{pmatrix}}_{M_{21}} \\
&+ \underbrace{\begin{pmatrix} \sum_{z \in \mathcal{Z}} \left( \frac{n_1(z)}{n(z)} - \pi_1 \right) E(Y_i^{(1)} - \theta_1 - (X_i - \mu_X)^\top b_1 \mid Z_i = z) P(Z = z) \pi_1^{-1} \\ \cdots \\ \sum_{z \in \mathcal{Z}} \left( \frac{n_k(z)}{n(z)} - \pi_k \right) E(Y_i^{(k)} - \theta_k - (X_i - \mu_X)^\top b_k \mid Z_i = z) P(Z = z) \pi_k^{-1} \end{pmatrix}}_{M_{12}} \\
&+ \underbrace{\begin{pmatrix} n^{-1} \sum_{i=1}^{n} E(Y_i^{(1)} - \theta_1 - (X_i - \mu_X)^\top b_1 \mid Z_i) \\ \cdots \\ n^{-1} \sum_{i=1}^{n} E(Y_i^{(k)} - \theta_k - (X_i - \mu_X)^\top b_k \mid Z_i) \end{pmatrix}}_{M_{31}} + \underbrace{\begin{pmatrix} n^{-1} \sum_{i=1}^{n} b_1^\top E(X_i - \mu_X \mid Z_i) \\ \cdots \\ n^{-1} \sum_{i=1}^{n} b_k^\top E(X_i - \mu_X \mid Z_i) \end{pmatrix}}_{M_{32}} \\
&+ o_p(n^{-1/2})
\end{aligned}
$$

$$
= \begin{pmatrix} a_1^\top & -b_1^\top & 0_p^\top & \cdots & 0_p^\top \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_k^\top & 0_p^\top & 0_p^\top & \cdots & -b_k^\top \end{pmatrix} \underbrace{\begin{pmatrix} \bar{Y}_1 - E(\bar{Y}_1 \mid \mathcal{A}, \mathcal{F}) \\ \cdots \\ \bar{Y}_k - E(\bar{Y}_k \mid \mathcal{A}, \mathcal{F}) \\ \bar{X}_1 - E(\bar{X}_1 \mid \mathcal{A}, \mathcal{F}) \\ \cdots \\ \bar{X}_k - E(\bar{X}_k \mid \mathcal{A}, \mathcal{F}) \end{pmatrix}}_{\tilde{V}_1}
$$

$$
+ \begin{pmatrix} b_1^\top & & \\ & \ddots & \\ & & b_k^\top \end{pmatrix} \begin{pmatrix} n^{-1}n_1 I_p & n^{-1}n_2 I_p & \dots & n^{-1}n_k I_p \\ \vdots & \vdots & \ddots & \vdots \\ n^{-1}n_1 I_p & n^{-1}n_2 I_p & \dots & n^{-1}n_k I_p \end{pmatrix} \begin{pmatrix} \bar{X}_1 - E(\bar{X}_1 \mid \mathcal{A}, \mathcal{F}) \\ \vdots \\ \bar{X}_k - E(\bar{Y}_k \mid \mathcal{A}, \mathcal{F}) \end{pmatrix}
$$

$$
+ M_{12} + M_{31} + M_{32} + o_p(n^{-1/2}).
$$

Conditioned on $\mathcal{A}, \mathcal{F}$, every component in $\tilde{V}_1$ is an average of independent terms. From the Lindeberg's Central Limit Theorem, as $n \to \infty$, $\sqrt{n}\tilde{V}_1$ is asymptotically normal with mean 0 conditional on $\mathcal{A}, \mathcal{F}$, which combined with the Cramér-Wold device implies that $\sqrt{n}(M_{11} + M_{21})$ is asymptotically normal with mean 0 conditional on $\mathcal{A}, \mathcal{F}$. Following the same steps as in the proof of Theorem 2, we have that

$$
\begin{aligned}
&\sqrt{n}(M_{11} + M_{21}) \mid \mathcal{A}, \mathcal{F} \xrightarrow{d} \\
&N\Big(0, \operatorname{diag}\Big\{\pi_t^{-1} E[\operatorname{var}\{Y_i^{(t)} - X_i^\top b_t \mid Z_i\}]\Big\} + B^\top E\{\operatorname{var}(X_i \mid Z_i)\} B \\
&\qquad + (\mathscr{B} - B)^\top E\{\operatorname{var}(X_i \mid Z_i)\} B + B^\top E\{\operatorname{var}(X_i \mid Z_i)\}(\mathscr{B} - B)\Big),
\end{aligned}
$$

and

$$
\begin{aligned}
&\sqrt{n}(M_{11} + M_{21}) \xrightarrow{d} \\
&N\Big(0, \operatorname{diag}\Big\{\pi_t^{-1} E[\operatorname{var}\{Y_i^{(t)} - X_i^\top b_t \mid Z_i\}]\Big\} + B^\top E\{\operatorname{var}(X_i \mid Z_i)\} B \\
&\qquad + (\mathscr{B} - B)^\top E\{\operatorname{var}(X_i \mid Z_i)\} B + B^\top E\{\operatorname{var}(X_i \mid Z_i)\}(\mathscr{B} - B)\Big).
\end{aligned}
$$

Next, notice that $\sqrt{n}M_{12}$ is asymptotically normal conditional on $\mathcal{F}$ with mean 0 from condition (C3). Let $\omega_{ts}(z)$ be the $(t,s)$ element in the matrix $\Omega(z)$, then the conditional variance of $\sqrt{n}M_{12t}$, the $t$th component of $\sqrt{n}M_{12}$, equals

$$
\begin{aligned}
& \mathrm{var}(\sqrt{n}M_{12t} \mid \mathcal{F}) \\
= \; & \pi_t^{-2}\sum_z \left[E\left\{Y_i^{(t)} - \theta_t - (X_i-\mu_X)^\top b_t \mid Z=z\right\}\right]^2 P(Z_i=z)\mathrm{var}\left\{\frac{n_t(z)-\pi_t n(z)}{\sqrt{n(z)}} \mid \mathcal{F}\right\} \\
& + o_p(1) \\
= \; & \pi_t^{-2}\sum_z \left[E\left\{Y_i^{(t)} - \theta_t - (X_i-\mu_X)^\top b_t \mid Z_i=z\right\}\right]^2 P(Z_i=z)\omega_{tt}(z) + o_p(1) \\
= \; & \pi_t^{-2}E\left[\omega_{tt}(Z)\big[E\{Y_i^{(t)} - \theta_t - (X_i-\mu_X)^\top b_t \mid Z_i\}\big]^2\right] + o_p(1),
\end{aligned}
$$

and the conditional covariance between $\sqrt{n}M_{12t}$ and $\sqrt{n}M_{12s}$ equals

$$
\begin{aligned}
& \mathrm{cov}(\sqrt{n}M_{12t}, \sqrt{n}M_{12s} \mid \mathcal{F}) \\
& = \frac{1}{\pi_t\pi_s}\sum_z \prod_{m\in\{t,s\}} E\left\{Y_i^{(m)} - \theta_i - (X_i-\mu_X)^\top b_m \mid Z=z\right\} P(Z_i=z) \\
& \qquad \mathrm{cov}\left\{\frac{n_t(z)-\pi_t n(z)}{\sqrt{n(z)}}, \frac{n_s(z)-\pi_s n(z)}{\sqrt{n(z)}} \mid \mathcal{F}\right\} + o_p(1) \\
& = \frac{1}{\pi_t\pi_s}E\left[\omega_{ts}(Z)E\left\{Y_i^{(t)} - \theta_t - (X_i-\mu_X)^\top b_t \mid Z_i\right\}E\left\{Y_i^{(s)} - \theta_s - (X_i-\mu_X)^\top b_s \mid Z_i\right\}\right] \\
& \qquad + o_p(1).
\end{aligned}
$$

Therefore, from the Slutsky's theorem,

$$
\sqrt{n}M_{12} \mid \mathcal{F} \xrightarrow{d} N\big(0, E\left\{R(B)\Omega(Z_i)R(B)\right\}\big).
$$

Moreover, $M_{31}+M_{32}$ only involves sums of identically and independently distributed terms, and $E(M_{31}+M_{32})=0$. Again using the Cramér-Wold device similarly to the proof of $M_{11}+M_{21}$, we have that $\sqrt{n}(M_{31}+M_{32})$ is asymptotically normal. Let $\pi=(\pi_1,\ldots,\pi_k)^\top$, it is easy to show that

$$
\mathrm{var}(\sqrt{n}M_{31}) = \mathrm{var}\{R(B)\pi\} = E\{R(B)\pi\pi^\top R(B)\}, \quad \mathrm{var}(\sqrt{n}M_{32}) = B^\top \mathrm{var}\{E(X_i \mid Z_i)\}B,
$$

and the $(t,s)$ component of $\mathrm{cov}(\sqrt{n}M_{31}, \sqrt{n}M_{32})$ is

$$\begin{aligned}\mathrm{cov}(\sqrt{n}M_{31t}, \sqrt{n}M_{32s}) &= \mathrm{cov}\left[E\left\{Y_i^{(t)} - (X_i - \mu_X)^\top b_t \mid Z_i\right\}, E(X_i^\top b_s \mid Z_i)\right] \\ &= (\beta_t - b_t)^\top \mathrm{var}(E(X_i \mid Z_i)) b_s.\end{aligned}$$

Hence,

$$\begin{aligned}&\mathrm{var}(\sqrt{n}(M_{31} + M_{32})) \\ &= E\{R(B)\pi\pi^\top R(B)\} + B^\top \mathrm{var}\{E(X_i \mid Z_i)\}B + (\mathscr{B} - B)^\top \mathrm{var}(E(X_i \mid Z_i))B \\ &\quad + B^\top \mathrm{var}(E(X_i \mid Z_i))(\mathscr{B} - B).\end{aligned}$$

Therefore,

$$\begin{aligned}&\sqrt{n}(M_{31} + M_{32}) \\ &\xrightarrow{d} N\Big(0, E\{R(B)\pi\pi^\top R(B)\} + B^\top \mathrm{var}\{E(X_i \mid Z_i)\}B + (\mathscr{B} - B)^\top \mathrm{var}(E(X_i \mid Z_i))B \\ &\qquad + B^\top \mathrm{var}(E(X_i \mid Z_i))(\mathscr{B} - B)\Big).\end{aligned} \tag{S7}$$

Next, we show that $(\sqrt{n}(M_{11}+M_{21}), \sqrt{n}M_{12}, \sqrt{n}(M_{31}+M_{32})) \xrightarrow{d} (\xi_{M1}, \xi_{M2}, \xi_{M3})$, where $(\xi_{M1}, \xi_{M2}, \xi_{M3})$ are mutually independent. This can be seen from

$$\begin{aligned}&P(\sqrt{n}(M_{11} + M_{21}) \le t_1, \sqrt{n}M_{12} \le t_2, \sqrt{n}(M_{31} + M_{32}) \le t_3) \\ =&E\left\{I(\sqrt{n}(M_{11} + M_{21}) \le t_1)I(\sqrt{n}M_{12} \le t_2)I(\sqrt{n}(M_{31} + M_{32}) \le t_3)\right\} \\ =&E\left\{P(\sqrt{n}(M_{11} + M_{21}) \le t_1 \mid \mathcal{A}, \mathcal{F})I(\sqrt{n}M_{12} \le t_2)I(\sqrt{n}(M_{31} + M_{32}) \le t_3)\right\} \\ =&E\left[\left\{P(\sqrt{n}(M_{11} + M_{21}) \le t_1 \mid \mathcal{A}, \mathcal{F}) - P(\xi_{M1} \le t_1)\right\}I(\sqrt{n}M_{12} \le t_2)I(\sqrt{n}(M_{31} + M_{32}) \le t_3)\right] \\ &\quad + P(\xi_{M1} \le t_1)E\left\{I(\sqrt{n}M_{12} \le t_2)I(\sqrt{n}(M_{31} + M_{32}) \le t_3)\right\} \\ =&E\left[\left\{P(\sqrt{n}(M_{11} + M_{21}) \le t_1 \mid \mathcal{A}, \mathcal{F}) - P(\xi_{M1} \le t_1)\right\}I(\sqrt{n}M_{12} \le t_2)I(\sqrt{n}(M_{31} + M_{32}) \le t_3)\right] \\ &\quad + P(\xi_{M1} \le t_1)E\left\{\left[P(\sqrt{n}M_{12} \le t_2 \mid \mathcal{F}) - P(\xi_{M2} \le t_2)\right] I(\sqrt{n}(M_{31} + M_{32}) \le t_3)\right\} \\ &\quad + P(\xi_{M1} \le t_1)P(\xi_{M2} \le t_2)P(\sqrt{n}(M_{31} + M_{32}) \le t_3)\end{aligned}$$

$\rightarrow P(\xi_{M1} \leq t_1) P(\xi_{M2} \leq t_2) P(\xi_{M3} \leq t_3),$

where the last step follows from the bounded convergence theorem.

Finally, from $\sqrt{n}\{\hat{\theta}(b_1,\dots,b_k)-\theta\} = \sqrt{n}(M_{11}+M_{21}+M_{12}+M_{31}+M_{32})+o_p(n^{-1/2})$, we conclude that $\sqrt{n}(\hat{\theta}(b_1,\dots,b_k)-\theta)$ is asymptotically normal with mean 0 and variance

$$
\begin{aligned}
& \operatorname{diag}\left\{\pi_t^{-1} E\{\operatorname{var}(Y_i^{(t)} - X_i^\top b_t \mid Z_i)\}\right\} + E\left\{R(B)\Omega(Z_i)R(B)\right\} + E\left\{R(B)\pi\pi^\top R(B)\right\} \\
& + B^\top \Sigma_X B + (\mathscr{B} - B)^\top \Sigma_X B + B^\top \Sigma_X (\mathscr{B} - B) \\
= \; & \operatorname{diag}\left\{\pi_t^{-1} E\{\operatorname{var}(Y_i^{(t)} - X_i^\top b_t \mid Z_i)\}\right\} + E\left\{R(B)\Omega(Z_i)R(B)\right\} + E\left\{R(B)\pi\pi^\top R(B)\right\} \\
& - B^\top \Sigma_X B + \mathscr{B}^\top \Sigma_X B + B^\top \Sigma_X \mathscr{B} \\
= \; & \operatorname{diag}\left\{\pi_t^{-1} \operatorname{var}(Y^{(t)} - b_t^\top X)\right\} - \operatorname{diag}\left\{\pi_t^{-1} \operatorname{var}\{E(Y_i^{(t)} - X_i^\top b_t \mid Z_i)\}\right\} \\
& + E\left\{R(B)\Omega(Z_i)R(B)\right\} + E\left\{R(B)\pi\pi^\top R(B)\right\} - B^\top \Sigma_X B + \mathscr{B}^\top \Sigma_X B + B^\top \Sigma_X \mathscr{B} \\
= \; & V_{\mathrm{SR}}(B) - \operatorname{diag}\left\{\pi_t^{-1} \operatorname{var}\{E(Y_i^{(t)} - X_i^\top b_t \mid Z_i)\}\right\} + E\left[R(B)\{\Omega(Z_i) + \pi\pi^\top\}R(B)\right] \\
= \; & V_{\mathrm{SR}}(B) - \operatorname{diag}\left\{R(B)\operatorname{diag}(\pi_t)R(B)\right\} + E\left[R(B)\{\Omega(Z_i) + \pi\pi^\top\}R(B)\right] \\
= \; & V_{\mathrm{SR}}(B) - E\left[R(B)\{\operatorname{diag}(\pi_t) - \pi\pi^\top - \Omega(Z_i)\}R(B)\right] \\
= \; & V_{\mathrm{SR}}(B) - E\left[R(B)\{\Omega_{\mathrm{SR}} - \Omega(Z_i)\}R(B)\right].
\end{aligned}
$$

(ii) By using the definition $\beta_t = \Sigma_X^{-1}\operatorname{cov}(X_i, Y_i^{(t)})$, we have $E[X_i^\top\{Y_i^{(t)} - \theta_t - \beta_t^\top(X - \mu_X)\}] = \operatorname{cov}(X_i, Y_i^{(t)}) - \operatorname{cov}(X_i, Y_i^{(t)}) = 0$. Because $X_i$ contains all dummy variables for the joint levels of $Z_i$, we have $E\{Y_i^{(t)} - \theta_t - \beta_t^\top(X_i - \mu_X) \mid Z_i\} = 0$. Hence $R(\mathscr{B}) = 0$ and $R(B) = \operatorname{diag}\{\pi_t^{-1}(\beta_t - b_t)^\top E(X_i - \mu_X \mid Z_i)\}$. Consequently, the difference in asymptotic variance is

$$
\begin{aligned}
& V(B) - V(\mathscr{B}) = V(B) - V_{\mathrm{SR}}(\mathscr{B}) = V(B) - V_{\mathrm{SR}}(B) + V_{\mathrm{SR}}(B) - V_{\mathrm{SR}}(\mathscr{B}) \\
= & \operatorname{diag}\{\pi_t^{-1}(\beta_t - b_t)^\top \Sigma_X (\beta_t - b_t)\} - (\mathscr{B} - B)^\top \Sigma_X (\mathscr{B} - B) - E\left[R(B)\{\Omega_{\mathrm{SR}} - \Omega(Z_i)\}R(B)\}\right] \\
\geq & \operatorname{diag}\{\pi_t^{-1}(\beta_t - b_t)^\top \Sigma_X (\beta_t - b_t)\} - (\mathscr{B} - B)^\top \Sigma_X (\mathscr{B} - B) - E\left[R(B)\Omega_{\mathrm{SR}}R(B)\}\right] \\
= & \operatorname{diag}[\pi_t^{-1}(\beta_t - b_t)^\top E\{\operatorname{var}(X \mid Z)\}(\beta_t - b_t)] - (\mathscr{B} - B)^\top E\{\operatorname{var}(X \mid Z)\}(\mathscr{B} - B),
\end{aligned}
$$

where $M \geq M'$ means $M - M'$ is positive semidefinite for two square matrices $M$ and $M'$ of the same dimension, and the last line follows from $\Omega_{\mathrm{SR}} = \mathrm{diag}(\pi_t) - \pi\pi^\top$, the expression for $R(B)$, and the identity $\Sigma_X = E\{\mathrm{var}(X \mid Z)\} + \mathrm{var}\{E(X \mid Z)\}$. The positive semidefiniteness of the right hand side is from applying Lemma 1 with $M = [E\{\mathrm{var}(X \mid Z)\}]^{1/2}(\mathscr{B} - B)$.

### 3.7 Proof of Corollary 2

When $X$ *only* contains the dummy variables for the joint levels of $Z$, $R(B) = \mathrm{diag}\{\pi_t^{-1}(\beta_t - b_t)^\top(X_i - \mu_X)\}$. Then, it follows from the proof of Theorem 3(ii) that

$$
\begin{aligned}
& V(B) - V_{\mathrm{SR}}(\mathscr{B}) \\
=& \mathrm{diag}\{\pi_t^{-1}(\beta_t - b_t)^\top \Sigma_X (\beta_t - b_t)\} - (\mathscr{B} - B)^\top \Sigma_X (\mathscr{B} - B) - E\left[R(B)\Omega_{\mathrm{SR}} R(B)\}\right] \\
& \qquad + E\left[R(B)\Omega(Z_i)R(B)\}\right] \\
=& \mathrm{diag}\{\pi_t^{-1}(\beta_t - b_t)^\top \Sigma_X (\beta_t - b_t)\} - (\mathscr{B} - B)^\top \Sigma_X (\mathscr{B} - B) - E\left[R(B)\mathrm{diag}(\pi_t) R(B)\}\right] \\
& \qquad + E\left[R(B)\pi\pi^\top R(B)\}\right] + E\left[R(B)\Omega(Z_i)R(B)\}\right] \\
=& E[R(B)\Omega(Z_i)R(B)]
\end{aligned}
$$

# References


Dawid, A. P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(1):1–31.