Supplementary Material to Robustly leveraging post-randomization information to improve precision in randomized trials

Bingkai Wang[1] and Yu Du[2]

[1]The Statistics and Data Science Department of the Wharton School, University of Pennsylvania, PA, USA
[2]Statistics, Data and Analytics, Eli Lilly and Company, IN, USA

In Section A, we give some useful notations. In Section B, we provide the regularity conditions for Theorem 1 of the main paper. In Section C, we define the variance estimators for $\widetilde{V}^{(\mathrm{est})}$ and $V^{(\mathrm{est})}$ and provide consistent estimators. In Section D, we introduce a few lemmas for proving our main results. In Section E, we prove Theorem 1, Corollary 1 and Corollary 2 of the main paper. In Section F, we give an example where MMRM-II is less precise than ANCOVA. In Section G, we provide the missing data mechanism for MAR in the simulation study. In Section H, we provide additional simulations under homogeneity and homoscedasticity.

A Notations

Throughout the proofs, we use $\boldsymbol{Y}, \boldsymbol{Y}(j), Y_t, Y_t(j)$ instead of $\boldsymbol{Y}_i, \boldsymbol{Y}_i(j), Y_{it}, Y_{it}(j)$ to represent random variables from the distribution P for conciseness. The same notation is used for $\boldsymbol{M}, \boldsymbol{M}(j), M_t, M_t(j), \boldsymbol{X}, A$.

Let $\mathbf{1}_K$ be the column vector of length K with each component equal to 1, $\mathbf{0}_K$ be the column vector of length K with each component equal to 0, and $\boldsymbol{e}_t$ be the column vector of length K with the t-th entry 1 and the rest 0. Let $\mathbf{I}_K$ be the $K \times K$ identity matrix. Let $\otimes$ be the Kronecker product. Let I be the indicator function, i.e. $I\{A\} = 1$ if event A is true and 0 otherwise. For any random vector $\boldsymbol{W}$ with finite second-order moment, we define $\widetilde{\boldsymbol{W}} = \boldsymbol{W} - E[\boldsymbol{W}]$ and $Var(\boldsymbol{W}) = E[\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{W}}^\top]$. For two random vectors $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$, we define $Cov(\boldsymbol{W}_1, \boldsymbol{W}_2) = E[\widetilde{\boldsymbol{W}}_1\widetilde{\boldsymbol{W}}_2^\top]$. Let $\{0,1\}^K$ be the set of K-dimensional binary vectors, i.e. $\{0,1\}^K = \{(x_1, \ldots, x_K)^\top : x_t \in \{0,1\}, t = 1, \ldots K\}$. For any matrix (or vector), we use $||\cdot||$ to denote its L_2 matrix (or vector) norm. For any vector $\boldsymbol{v} = (v_1, \ldots, v_L)$, we use $diag\{\boldsymbol{v}\}$ or $diag\{v_l : l = 1, \ldots, L\}$ to denote an $L \times L$ diagonal matrix with the diagonal entries being $(v_1, \ldots, v_L)$. For any sequence $x_1, \ldots, x_n, \ldots$, we define $P_n x = n^{-1}\sum_{i=1}^n x_i$.

B Regularity conditions

The regularity conditions for Theorem 1 are assumed on estimating equations $\boldsymbol{\psi}^{(\mathrm{ANCOVA})}$, $\boldsymbol{\psi}^{(\mathrm{MMRM})}$ and $\boldsymbol{\psi}^{(\mathrm{IMMRM})}$, which are defined by Equations (1), (3), and (5) below, respectively. Each of $\boldsymbol{\psi}^{(\mathrm{ANCOVA})}$, $\boldsymbol{\psi}^{(\mathrm{MMRM})}$ and $\boldsymbol{\psi}^{(\mathrm{IMMRM})}$ is a q-dimensional function of random variables $(A, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{M})$ and a set of parameters $\boldsymbol{\theta} \in \mathbb{R}^q$, where q and $\boldsymbol{\theta}$ vary among estimators, and $\boldsymbol{\Delta}$ are embedded in $\boldsymbol{\theta}$. Without causing confusion, we use $\boldsymbol{\psi}(A, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{M}; \boldsymbol{\theta})$ to represent any of the above three estimating equations. As we show in the proof of Theorem 1 below, each of the ANCOVA, MMRM-I, MMRM-II and IMMRM estimators is an M-estimator, which is defined as the solution of $\boldsymbol{\Delta}$ to the equations $P_n\boldsymbol{\psi}(A, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{M}; \boldsymbol{\theta}) = \mathbf{0}$.

The regularity conditions are similar to those used in Section 5.3 of van der Vaart

(1998) in their theorem on estimating equations $\boldsymbol{\psi}$ for showing asymptotic normality of M-estimators for independent, identically distributed data. The regularity conditions are given below:

(1) $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, a compact set in $\mathbb{R}^q$.

(2) $E[||\boldsymbol{\psi}(j, \boldsymbol{X}, \boldsymbol{Y}(j), \boldsymbol{M}(j); \boldsymbol{\theta})||^2] < \infty$ for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $j \in \{0, \ldots, J\}$.

(3) There exists a unique solution in the interior of $\boldsymbol{\Theta}$, denoted as $\underline{\boldsymbol{\theta}}$, to the equations

$$\sum_{j=0}^{J} \pi_j E[\boldsymbol{\psi}(j, \boldsymbol{X}, Y(j), M(j); \boldsymbol{\theta})] = \mathbf{0}.$$

(4) For each $j \in \{0, \ldots, J\}$, the function $\boldsymbol{\theta} \mapsto \boldsymbol{\psi}(j, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{m}; \boldsymbol{\theta})$ is twice continuously differentiable for every $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{m})$ in the support of $(\boldsymbol{X}, \boldsymbol{Y}(j), \boldsymbol{M}(j))$ and is dominated by an integrable function $\boldsymbol{u}(\boldsymbol{X}, \boldsymbol{Y}(j), \boldsymbol{M}(j))$.

(5) There exist a $C > 0$ and and integrable function $v(\boldsymbol{X}, \boldsymbol{Y}(j), \boldsymbol{M}(j))$, such that, for each entry $\psi_r, r = 1, \ldots, q$, of $\boldsymbol{\psi}$, $||\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} \psi_r(j, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{m}; \boldsymbol{\theta})|| < v(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{m})$ for every $(j, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{m})$ in the support of $(A, \boldsymbol{X}, \boldsymbol{Y}(j), \boldsymbol{M}(j))$ and $\boldsymbol{\theta}$ with $||\boldsymbol{\theta} - \underline{\boldsymbol{\theta}}|| < C$.

(6) $E\left[\left|\left|\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\psi}(j, \boldsymbol{X}, \boldsymbol{Y}(j), \boldsymbol{M}(j); \boldsymbol{\theta})\Big|_{\boldsymbol{\theta} = \underline{\boldsymbol{\theta}}}\right|\right|^2\right] < \infty$ for $j \in \{0, \ldots, J\}$ and

$$\sum_{j=0}^{J} \pi_j E\left[\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\psi}(j, \boldsymbol{X}, \boldsymbol{Y}(j), \boldsymbol{M}(j); \boldsymbol{\theta})\Big|_{\boldsymbol{\theta} = \underline{\boldsymbol{\theta}}}\right]$$

is invertible.

C Variance estimators in Theorem 1

For an M-estimator $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^q$ defined by $P_n \boldsymbol{\psi}(A, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{M}; \boldsymbol{\theta}) = \mathbf{0}$, its sandwich variance estimator under simple randomization is defined as

$$\widetilde{\mathbf{V}}_n(\boldsymbol{\psi}, \widehat{\boldsymbol{\theta}}) = \frac{1}{n} \left\{ P_n \left[\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\psi}(A, \boldsymbol{X}, Y, M; \boldsymbol{\theta}) \bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \right] \right\}^{-1} \left\{ P_n \left[\boldsymbol{\psi}(A, \boldsymbol{X}, Y, M; \widehat{\boldsymbol{\theta}}) \boldsymbol{\psi}(A, \boldsymbol{X}, Y, M; \widehat{\boldsymbol{\theta}})^t \right] \right\}$$
$$\left\{ P_n \left[\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\psi}(A, \boldsymbol{X}, Y, M; \boldsymbol{\theta}) \bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \right] \right\}^{-1,t}.$$

Since $\boldsymbol{\Delta}$ is embedded in $\boldsymbol{\theta}$, we can find $\mathbf{C} \in \mathbf{R}^{J \times q}$ such that $\boldsymbol{\Delta} = \mathbf{C}\boldsymbol{\theta}$. Then the variance estimator of $\widehat{\boldsymbol{\Delta}}$ under simple randomization is defined as $\mathbf{C}\widetilde{\mathbf{V}}_n(\boldsymbol{\psi}, \widehat{\boldsymbol{\theta}})\mathbf{C}^\top$.

For each est $\in \{\text{ANCOVA}, \text{MMRM-I}, \text{MMRM-II}, \text{IMMRM}\}$, the variance estimator $\widetilde{\mathbf{V}}_n^{(\text{est})}$ is calculated by $\mathbf{C}\widetilde{\mathbf{V}}_n(\boldsymbol{\psi}^{(\text{est})}, \widehat{\boldsymbol{\theta}})\mathbf{C}^\top$. We note that the MMRM-I estimator and MMRM-II estimator share the same estimating equations $\boldsymbol{\psi}^{(\text{MMRM})}$ with different specifications of $\mathbf{u}(\boldsymbol{X})$ as described in Equation (2).

We next define $\mathbf{V}_n^{(\text{est})}$ for est $= \text{ANCOVA}, \text{MMRM-I}, \text{MMRM-II}$. Define

$$\begin{aligned}
\widehat{Var}(\widehat{E}[\boldsymbol{X}|S]) &= \sum_{s\in\mathcal{S}} \frac{(P_n I\{S=s\}\boldsymbol{X})(P_n I\{S=s\}\boldsymbol{X})^\top}{P_n I\{S=s\}} - (P_n\boldsymbol{X})(P_n\boldsymbol{X})^\top,\\
\widehat{Var}(\boldsymbol{X}) &= P_n\boldsymbol{X}\boldsymbol{X}^\top - (P_n\boldsymbol{X})(P_n\boldsymbol{X})^\top,\\
\widehat{Cov}(\boldsymbol{X}, Y_K(j)) &= P_n\frac{I\{A=j\}}{\pi_j}\boldsymbol{X}Y_K - P_n\boldsymbol{X}P_n\frac{I\{A=j\}}{\pi_j}Y_K,\\
\widehat{\boldsymbol{b}}_{Kj} &= \widehat{Var}(\boldsymbol{X})^{-1}\widehat{Cov}(\boldsymbol{X}, Y_K(j)),\\
\widehat{\boldsymbol{b}}_K &= \sum_{j=0}^{J}\pi_j\widehat{Var}(\boldsymbol{X})^{-1}\widehat{Cov}(\boldsymbol{X}, Y_K(j)),\\
\widehat{\mathbf{z}} &= (\widehat{\boldsymbol{b}}_{K0} - \widehat{\boldsymbol{b}}_K, \dots, \widehat{\boldsymbol{b}}_{KJ} - \widehat{\boldsymbol{b}}_K),\\
\widehat{\mathbf{v}} &= (\widehat{\boldsymbol{b}}_{K0} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{X}}, \dots, \widehat{\boldsymbol{b}}_{KJ} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{X}}),\\
\mathbf{L} &= (-\mathbf{1}_J \quad \mathbf{I}_J),
\end{aligned}$$

where $\widehat{\boldsymbol{\beta}}_{\boldsymbol{X}}$ is the MLE of $\boldsymbol{\beta}_{\boldsymbol{X}}$ in the MMRM-I working model (2).

Following Equation (16), we define

$$\begin{aligned}
\mathbf{V}_n^{(\text{ANCOVA})} = \widetilde{\mathbf{V}}_n^{(\text{ANCOVA})} - \frac{1}{n}\mathbf{L}[&diag\{\pi_j^{-1}(\widehat{\boldsymbol{b}}_{Kj} - \widehat{\boldsymbol{b}}_K)^\top\widehat{Var}(\widehat{E}[\boldsymbol{X}|S])(\widehat{\boldsymbol{b}}_{Kj} - \widehat{\boldsymbol{b}}_K) : j = 0, \dots, J\}\\
&- \widehat{\mathbf{z}}^\top\widehat{Var}(\widehat{E}[\boldsymbol{X}|S])\widehat{\mathbf{z}}]\mathbf{L}^\top.
\end{aligned}$$

Following Equation (15), we define

$$\begin{aligned}
\mathbf{V}_n^{(\text{MMRM-I})} = \widetilde{\mathbf{V}}_n^{(\text{MMRM-I})} - \frac{1}{n}\mathbf{L}[&diag\{\pi_j^{-1}(\widehat{\boldsymbol{b}}_{Kj} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{X}})^\top\widehat{Var}(\widehat{E}[\boldsymbol{X}|S])(\widehat{\boldsymbol{b}}_{Kj} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{X}}) : j = 0, \dots, J\}\\
&- \widehat{\mathbf{v}}^\top\widehat{Var}(\widehat{E}[\boldsymbol{X}|S])\widehat{\mathbf{v}}]\mathbf{L}^\top.
\end{aligned}$$

Similarly, we define

$$\begin{aligned}
\mathbf{V}_n^{(\text{MMRM-II})} = \widetilde{\mathbf{V}}_n^{(\text{MMRM-II})} - \frac{1}{n}\mathbf{L}[&diag\{\pi_j^{-1}(\widehat{\boldsymbol{b}}_{Kj} - \widehat{\boldsymbol{b}}_K)^\top\widehat{Var}(\widehat{E}[\boldsymbol{X}|S])(\widehat{\boldsymbol{b}}_{Kj} - \widehat{\boldsymbol{b}}_K) : j = 0, \dots, J\}\\
&- \widehat{\mathbf{z}}^\top\widehat{Var}(\widehat{E}[\boldsymbol{X}|S])\widehat{\mathbf{z}}]\mathbf{L}^\top.
\end{aligned}$$

D Lemmas

Lemma 1. *Let $\mathbf{\Sigma} \in \mathbb{R}^{K \times K}$ be a positive definite matrix. For each $\boldsymbol{m} \in \{0,1\}^K \setminus \{\mathbf{0}_K\}$, let $n_{\boldsymbol{m}} = \sum_{t=1}^K m_t$ be the number of ones in $\boldsymbol{m}$ and $t_{\boldsymbol{m},1} < \cdots < t_{\boldsymbol{m},n_{\boldsymbol{m}}}$ denote the ordered list of locations of ones in $\boldsymbol{m}$, i.e., $m_t = 1$ if $t \in \{t_{\boldsymbol{m},1}, \ldots, t_{\boldsymbol{m},n_{\boldsymbol{m}}}\}$ and 0 otherwise. We define $\mathbf{D}_{\boldsymbol{m}} = [\boldsymbol{e}_{t_{\boldsymbol{m},1}} \ \ \ldots \ \ \boldsymbol{e}_{t_{\boldsymbol{m},n_{\boldsymbol{m}}}}] \in \mathbb{R}^{K \times n_{\boldsymbol{m}}}$ and*

$$\mathbf{V}_{\boldsymbol{m}}(\mathbf{\Sigma}) = I\{\boldsymbol{m} \in \{0,1\}^K \setminus \{\mathbf{0}_K\}\}\mathbf{D}_{\boldsymbol{m}}(\mathbf{D}_{\boldsymbol{m}}^\top \mathbf{\Sigma} \mathbf{D}_{\boldsymbol{m}})^{-1}\mathbf{D}_{\boldsymbol{m}}^\top \in \mathbb{R}^{K \times K},$$

which are deterministic functions of $\boldsymbol{m}$ and $\mathbf{\Sigma}$.

Let $\boldsymbol{M} = (M_1, \ldots, M_K)$ be a K-dimensional binary random vector taking values in $\{0,1\}^K$. We assume that $P(\boldsymbol{M} = \mathbf{1}_K) > 0$. Then the following statements hold.

(1) $E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{\Sigma})]$ is well-defined and positive definite.

(2) $\boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{\Sigma})]^{-1}\boldsymbol{e}_K \leq P(M_K = 1)^{-1}\boldsymbol{e}_K^\top \mathbf{\Sigma} \boldsymbol{e}_K$. The equality holds if and only if either of the following conditions holds: (i) $K = 1$ or (ii) $P(M_t = 1, M_K = 0)\sigma_{t,K} = 0$ for $t = 1, \ldots, K-1$, where $\sigma_{t,K} = \boldsymbol{e}_t^\top \mathbf{\Sigma} \boldsymbol{e}_K$ is the (t,K)-th entry of $\mathbf{\Sigma}$.

(3) Let $\mathbf{A} \in \mathbb{R}^{K \times K}$ be a positive definite matrix. Then $\boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{\Sigma})]^{-1}\boldsymbol{e}_K \leq \boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{A})]^{-1} E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{A})\mathbf{\Sigma}\mathbf{V}_{\boldsymbol{M}}(\mathbf{A})] E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{A})]^{-1}\boldsymbol{e}_K$. The equality holds if and only if $P(\boldsymbol{M} = \boldsymbol{m})\boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{A})]^{-1}\mathbf{V}_{\boldsymbol{m}}(\mathbf{A}) = P(\boldsymbol{M} = \boldsymbol{m})\boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{\Sigma})]^{-1}\mathbf{V}_{\boldsymbol{m}}(\mathbf{\Sigma})$ for all $\boldsymbol{m} \in \{0,1\}^K$.

(4) Let $\mathbf{B} \in \mathbb{R}^{K \times K}$ be a positive semi-definite matrix. Then $\boldsymbol{e}_K^\top \mathbf{B} \boldsymbol{e}_K \leq \boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{\Sigma})]^{-1} E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{\Sigma})\mathbf{B}\mathbf{V}_{\boldsymbol{M}}(\mathbf{\Sigma})] E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{\Sigma})]^{-1}\boldsymbol{e}_K$. The equality holds if and only if $P(\boldsymbol{M} = \boldsymbol{m})\boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{\Sigma})]^{-1}\mathbf{V}_{\boldsymbol{m}}(\mathbf{\Sigma})\mathbf{B}$ does not vary across $\boldsymbol{m} \in \{0,1\}^K$.

(5) Letting $\mathbf{C} \in \mathbb{R}^{K\times K}$ be a positive definite matrix such that $\mathbf{C}-\boldsymbol{\Sigma}$ is positive semi-definite, then $\boldsymbol{e}_K^\top(E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{C})]^{-1} - E[\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})]^{-1})\boldsymbol{e}_K \geq \boldsymbol{e}_K^\top(\mathbf{C}-\boldsymbol{\Sigma})\boldsymbol{e}_K$.

Proof. (1) By definition, for any $\boldsymbol{m} \in \{0,1\}^K \setminus \{\mathbf{0}_K\}$, $\mathbf{D}_{\boldsymbol{m}}$ is full column rank. Since $\boldsymbol{\Sigma}$ is positive definite and $n_i \leq K$, then $\mathbf{D}_{\boldsymbol{m}}^\top\boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{m}}$ is positive definite and $\mathbf{V}_{\boldsymbol{m}} = \mathbf{D}_{\boldsymbol{m}}(\mathbf{D}_{\boldsymbol{m}}^\top\boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{m}})^{-1}\mathbf{D}_{\boldsymbol{m}}^\top$ is positive semi-definite. In particular, for the case $\boldsymbol{m} = \mathbf{1}_K$, $\mathbf{V}_{\boldsymbol{m}} = \boldsymbol{\Sigma}^{-1}$ and so is positive definite. Since $E[\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})] = \sum_{\boldsymbol{m}\in\{0,1\}^K\setminus\{\mathbf{0}_K\}} \mathbf{V}_{\boldsymbol{m}} P(\boldsymbol{M} = \boldsymbol{m})$ and we assume $P(\boldsymbol{M} = \mathbf{1}_K) > 0$, then for any $\boldsymbol{x} \in \mathbb{R}^K$,

$$\begin{aligned}
\boldsymbol{x}^\top E[\mathbf{V}_{\boldsymbol{M}}]\boldsymbol{x} &= \boldsymbol{x}^\top\mathbf{V}_{\mathbf{1}_K}\boldsymbol{x}P(\boldsymbol{M} = \mathbf{1}_K) + \sum_{\boldsymbol{m}\in\{0,1\}^K\setminus\{\mathbf{0}_K,\mathbf{1}_K\}} \boldsymbol{x}^\top\mathbf{V}_{\boldsymbol{m}}\boldsymbol{x}P(\boldsymbol{M} = \boldsymbol{m}) \\
&\geq \boldsymbol{x}^\top\mathbf{V}_{\mathbf{1}_K}\boldsymbol{x}P(\boldsymbol{M} = \mathbf{1}_K) \\
&> 0,
\end{aligned}$$

which implies that $E[\mathbf{V}_{\boldsymbol{M}}]$ is positive definite and so invertible.

(2) Consider the case $K = 1$. we have that $\boldsymbol{\Sigma}$ reduces to a positive number σ and $V_m(\sigma) = I\{m = 1\}\sigma^{-1}$. Then $\boldsymbol{e}_K^\top E[V_m(\sigma)]^{-1}\boldsymbol{e}_K = \frac{1}{P(m=1)}\sigma = \frac{1}{P(M_K=1)}\boldsymbol{e}_K^\top\boldsymbol{\Sigma}\boldsymbol{e}_K$. We next consider the case that $K \geq 2$. Define $\boldsymbol{\Omega}_K = \{\boldsymbol{m} \in \{0,1\}^K \setminus \{\boldsymbol{e}_K\} : m_K = 1\}$ and $\boldsymbol{\Omega}_{-K} = \{\boldsymbol{m} \in \{0,1\}^K \setminus \{\mathbf{0}_K\} : m_K = 0\}$. Then we have

$$\mathbf{D}_{\boldsymbol{m}} = \begin{pmatrix} \widetilde{\mathbf{D}}_{\boldsymbol{m}} & \mathbf{0}_{K-1} \\ \mathbf{0}^\top_{\sum_{t=1}^K m_t - 1} & 1 \end{pmatrix} \text{ if } \boldsymbol{m} \in \boldsymbol{\Omega}_K \text{ and } \mathbf{D}_{\boldsymbol{m}} = \begin{pmatrix} \widetilde{\mathbf{D}}_{\boldsymbol{m}} \\ \mathbf{0}^\top_{\sum_{t=1}^K m_t} \end{pmatrix} \text{ if } \boldsymbol{m} \in \boldsymbol{\Omega}_{-K},$$

where $\widetilde{\mathbf{D}}_{\boldsymbol{m}} \in \mathbb{R}^{(K-1)\times(\sum_{t=1}^K m_t - 1)}$ is a matrix taking the first $K-1$ rows and first $(\sum_{t=1}^K m_t - 1)$ columns of $\mathbf{D}_{\boldsymbol{m}}$ if $\boldsymbol{m} \in \boldsymbol{\Omega}_K$ and $\widetilde{\mathbf{D}}_{\boldsymbol{m}} \in \mathbb{R}^{(K-1)\times(\sum_{t=1}^K m_t)}$ is the first $K-1$ rows of $\widetilde{\mathbf{D}}_{\boldsymbol{m}}$ if $\boldsymbol{m} \in \boldsymbol{\Omega}_{-K}$. We further define $\boldsymbol{\Sigma}_{-K,-K} \in \mathbb{R}^{(K-1)\times(K-1)}$, $\boldsymbol{\Sigma}_{-K,K} \in \mathbb{R}^{K-1}$ and $\sigma = \boldsymbol{e}_K^\top\boldsymbol{\Sigma}\boldsymbol{e}_K$

such that

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{-K,-K} & \boldsymbol{\Sigma}_{-K,K} \\ \boldsymbol{\Sigma}_{-K,K}^\top & \sigma \end{pmatrix}.$$

Using the formula of block matrix inversion, we can compute that, if $\boldsymbol{m} \in \boldsymbol{\Omega}_K$, then

$$\mathbf{V}_{\boldsymbol{m}}(\boldsymbol{\Sigma}) = \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{m}} & -\sigma^{-1}\boldsymbol{A}_{\boldsymbol{m}}\boldsymbol{\Sigma}_{-K,K} \\ -\sigma^{-1}\boldsymbol{\Sigma}_{-K,K}^\top\boldsymbol{A}_{\boldsymbol{m}} & \sigma^{-1} + \sigma^{-2}\boldsymbol{\Sigma}_{-K,K}^\top\boldsymbol{A}_{\boldsymbol{m}}\boldsymbol{\Sigma}_{-K,K} \end{pmatrix},$$

and, if $\boldsymbol{m} \in \boldsymbol{\Omega}_{-K}$, then

$$\mathbf{V}_{\boldsymbol{m}}(\boldsymbol{\Sigma}) = \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{m}} & \mathbf{0}_{K-1} \\ \mathbf{0}_{K-1}^\top & 0 \end{pmatrix},$$

where $\boldsymbol{A}_{\boldsymbol{m}} = \widetilde{\mathbf{D}}_{\boldsymbol{m}}\{\widetilde{\mathbf{D}}_{\boldsymbol{m}}^\top(\boldsymbol{\Sigma}_{-K,-K} - \boldsymbol{\Sigma}_{-K,K}\boldsymbol{\Sigma}_{-K,K}^\top/\sigma)\widetilde{\mathbf{D}}_{\boldsymbol{m}}\}^{-1}\widetilde{\mathbf{D}}_{\boldsymbol{m}}^\top \in \mathbb{R}^{(K-1)\times(K-1)}$ if $\boldsymbol{m} \in \Omega_K$ and $\boldsymbol{A}_{\boldsymbol{m}} = \widetilde{\mathbf{D}}_{\boldsymbol{m}}\{\widetilde{\mathbf{D}}_{\boldsymbol{m}}^\top\boldsymbol{\Sigma}_{-K,-K}\widetilde{\mathbf{D}}_{\boldsymbol{m}}\}^{-1}\widetilde{\mathbf{D}}_{\boldsymbol{m}}^\top \in \mathbb{R}^{(K-1)\times(K-1)}$ if $\boldsymbol{m} \in \Omega_{-K}$. Here $\boldsymbol{A}_{\boldsymbol{m}}$ is well defined for each $\boldsymbol{m} \in \Omega_K \bigcup \Omega_{-K}$ since $\widetilde{\mathbf{D}}_{\boldsymbol{m}}^\top\boldsymbol{\Sigma}\widetilde{\mathbf{D}}_{\boldsymbol{m}}$ is positive definite. In addition, if $\boldsymbol{m} = \boldsymbol{e}_K$, then $\mathbf{D}_{\boldsymbol{m}} = \boldsymbol{e}_K$ and $\mathbf{V}_{\boldsymbol{m}} = \sigma^{-1}\boldsymbol{e}_K\boldsymbol{e}_K^\top$. Hence

$$E[\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})] = \begin{pmatrix} \sum_{\boldsymbol{m}\in\Omega_K\bigcup\Omega_{-K}} p_{\boldsymbol{m}}\boldsymbol{A}_{\boldsymbol{m}} & -\sum_{\boldsymbol{m}\in\Omega_K} p_{\boldsymbol{m}}\sigma^{-1}\boldsymbol{A}_{\boldsymbol{m}}\boldsymbol{\Sigma}_{-K,K} \\ -\sum_{\boldsymbol{m}\in\Omega_K} p_{\boldsymbol{m}}\sigma^{-1}\boldsymbol{\Sigma}_{-K,K}^\top\boldsymbol{A}_{\boldsymbol{m}} & \sum_{\boldsymbol{m}\in\Omega_K} p_{\boldsymbol{m}}(\sigma^{-1} + \sigma^{-2}\boldsymbol{\Sigma}_{-K,K}^\top\boldsymbol{A}_{\boldsymbol{m}}\boldsymbol{\Sigma}_{-K,K}) + p_{\boldsymbol{e}_K}\sigma^{-1} \end{pmatrix}.$$

Since we have shown that $E[\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})]$ is positive definite and $P(M_K = 1) = p_{\boldsymbol{e}_K} + \sum_{\boldsymbol{m}\in\Omega_K} p_{\boldsymbol{m}}$, by using the formula of block matrix inversion again, we have

$$\begin{aligned}
&(\boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{m}}(\boldsymbol{\Sigma})]^{-1}\boldsymbol{e}_K)^{-1} \\
&= P(M_K = 1)\sigma^{-1} + \sum_{\boldsymbol{m}\in\Omega_K} p_{\boldsymbol{m}}\sigma^{-2}\boldsymbol{\Sigma}_{-K,K}^\top\boldsymbol{A}_{\boldsymbol{m}}\boldsymbol{\Sigma}_{-K,K} \\
&\quad - (\sum_{\boldsymbol{m}\in\Omega_K} p_{\boldsymbol{m}}\sigma^{-1}\boldsymbol{\Sigma}_{-K,K}^\top\boldsymbol{A}_{\boldsymbol{m}})(\sum_{\boldsymbol{m}\in\Omega_K\bigcup\Omega_{-K}} p_{\boldsymbol{m}}\boldsymbol{A}_{\boldsymbol{m}})^{-1}(\sum_{\boldsymbol{m}\in\Omega_K} p_{\boldsymbol{m}}\sigma^{-1}\boldsymbol{A}_{\boldsymbol{m}}\boldsymbol{\Sigma}_{-K,K}) \\
&= P(M_K = 1)\sigma^{-1} + \sigma^{-2}\boldsymbol{\Sigma}_{-K,K}^\top\{\mathbf{G}_{\Omega_K} - \mathbf{G}_{\Omega_K}(\mathbf{G}_{\Omega_K} + \mathbf{G}_{\Omega_{-K}})^{-1}\mathbf{G}_{\Omega_K}\}\boldsymbol{\Sigma}_{-K,K},
\end{aligned}$$

where $\mathbf{G}_{\Omega_K} = \sum_{\boldsymbol{m}\in\Omega_K} p_{\boldsymbol{m}} \boldsymbol{A}_{\boldsymbol{m}}$ and $\mathbf{G}_{\Omega_{-K}} = \sum_{\boldsymbol{m}\in\Omega_{-K}} p_{\boldsymbol{m}} \boldsymbol{A}_{\boldsymbol{m}}$.

For $\mathbf{G}_{\Omega_K}$, we note that $\mathbf{1}_K \in \Omega_K$ with $p_{\mathbf{1}_K} > 0$ and $\mathbf{G}_{\mathbf{1}_K} = \{\boldsymbol{\Sigma}_{-K,-K} - \boldsymbol{\Sigma}_{-K,K}\boldsymbol{\Sigma}_{-K,K}^\top/\sigma\}^{-1}$ is positive definite. Hence $\mathbf{G}_{\Omega_K}$ is positive definite. Furthermore, $\mathbf{G}_{\Omega_{-K}} \succeq \mathbf{0}$ by definition. Then

$$
\begin{aligned}
&\mathbf{G}_{\Omega_K}^{-1} - (\mathbf{G}_{\Omega_K} + \mathbf{G}_{\Omega_{-K}})^{-1} \\
&= (\mathbf{G}_{\Omega_K} + \mathbf{G}_{\Omega_{-K}})^{-1}\{(\mathbf{G}_{\Omega_K} + \mathbf{G}_{\Omega_{-K}})\mathbf{G}_{\Omega_K}^{-1}(\mathbf{G}_{\Omega_K} + \mathbf{G}_{\Omega_{-K}}) - (\mathbf{G}_{\Omega_K} + \mathbf{G}_{\Omega_{-K}})\}(\mathbf{G}_{\Omega_K} + \mathbf{G}_{\Omega_{-K}})^{-1} \\
&= (\mathbf{G}_{\Omega_K} + \mathbf{G}_{\Omega_{-K}})^{-1}(\mathbf{G}_{\Omega_{-K}} + \mathbf{G}_{\Omega_{-K}}\mathbf{G}_{\Omega_K}^{-1}\mathbf{G}_{\Omega_{-K}})(\mathbf{G}_{\Omega_K} + \mathbf{G}_{\Omega_{-K}})^{-1} \\
&\succeq \mathbf{0}.
\end{aligned}
$$

Hence

$$
\begin{aligned}
&(\boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{m}}(\boldsymbol{\Sigma})]^{-1}\boldsymbol{e}_K)^{-1} \\
&= P(M_K = 1)\sigma^{-1} + \sigma^{-2}\boldsymbol{\Sigma}_{-K,K}^\top\{\mathbf{G}_{\Omega_K} - \mathbf{G}_{\Omega_K}(\mathbf{G}_{\Omega_K} + \mathbf{G}_{\Omega_{-K}})^{-1}\mathbf{G}_{\Omega_K}\}\boldsymbol{\Sigma}_{-K,K} \\
&= P(M_K = 1)\sigma^{-1} + \sigma^{-2}\boldsymbol{\Sigma}_{-K,K}^\top\mathbf{G}_{\Omega_K}\{\mathbf{G}_{\Omega_K}^{-1} - (\mathbf{G}_{\Omega_K} + \mathbf{G}_{\Omega_{-K}})^{-1}\}\mathbf{G}_{\Omega_K}\boldsymbol{\Sigma}_{-K,K} \\
&\geq P(M_K = 1)\sigma^{-1},
\end{aligned}
$$

which completes the proof of $\boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})]^{-1}\boldsymbol{e}_K \leq \frac{1}{P(M_K=1)}\boldsymbol{e}_K^\top\boldsymbol{\Sigma}\boldsymbol{e}_K$.

We next examine when $\boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})]^{-1}\boldsymbol{e}_K = \frac{1}{P(M_K=1)}\boldsymbol{e}_K^\top\boldsymbol{\Sigma}\boldsymbol{e}_K$. Since $\mathbf{G}_{\Omega_K}^{-1}$ is positive semi-definite, then the above derivation shows that the equality holds if and only if $\boldsymbol{\Sigma}_{-K,K}^\top\mathbf{G}_{\Omega_K}(\mathbf{G}_{\Omega_K} + \mathbf{G}_{\Omega_{-K}})^{-1}\mathbf{G}_{\Omega_{-K}} = \mathbf{0}$. Noting that

$$
\mathbf{G}_{\Omega_K}(\mathbf{G}_{\Omega_K} + \mathbf{G}_{\Omega_{-K}})^{-1}\mathbf{G}_{\Omega_{-K}} = \mathbf{G}_{\Omega_K} - \mathbf{G}_{\Omega_K}(\mathbf{G}_{\Omega_K} + \mathbf{G}_{\Omega_{-K}})^{-1}\mathbf{G}_{\Omega_K}
$$

is symmetric, the equality holds if and only if $\boldsymbol{\Sigma}_{-K,K}^\top\mathbf{G}_{\Omega_{-K}}(\mathbf{G}_{\Omega_K} + \mathbf{G}_{\Omega_{-K}})^{-1}\mathbf{G}_{\Omega_K} = \mathbf{0}$. Since

$(\mathbf{G}_{\Omega_K} + \mathbf{G}_{\Omega_{-K}})^{-1}\mathbf{G}_{\Omega_K}$ is positive definite, we get

$$
\begin{aligned}
&\boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})]^{-1}\boldsymbol{e}_K = \frac{1}{P(M_K=1)}\boldsymbol{e}_K^\top \boldsymbol{\Sigma}\boldsymbol{e}_K \\
&\Leftrightarrow \boldsymbol{\Sigma}_{-K,K}^\top \mathbf{G}_{\Omega_{-K}}(\mathbf{G}_{\Omega_K} + \mathbf{G}_{\Omega_{-K}})^{-1}\mathbf{G}_{\Omega_K} = \mathbf{0} \\
&\Leftrightarrow \boldsymbol{\Sigma}_{-K,K}^\top \mathbf{G}_{\Omega_{-K}}\boldsymbol{\Sigma}_{-K,K} = 0 \\
&\Leftrightarrow \sum_{\boldsymbol{m}\in\Omega_{-K}} p_{\boldsymbol{m}}\boldsymbol{\Sigma}_{-K,K}^\top \widetilde{\mathbf{D}}_{\boldsymbol{m}}\{\widetilde{\mathbf{D}}_{\boldsymbol{m}}^\top \boldsymbol{\Sigma}_{-K,-K}\widetilde{\mathbf{D}}_{\boldsymbol{m}}\}^{-1}\widetilde{\mathbf{D}}_{\boldsymbol{m}}^\top \boldsymbol{\Sigma}_{-K,K} = 0 \\
&\Leftrightarrow p_{\boldsymbol{m}}\boldsymbol{\Sigma}_{-K,K}^\top \widetilde{\mathbf{D}}_{\boldsymbol{m}} = 0 \text{ for each } \boldsymbol{m}\in\Omega_{-K} \\
&\Leftrightarrow P(M_t=1, M_K=0)\sigma_{t,K} = 0 \text{ for each } t=1,\ldots,K-1,
\end{aligned}
$$

where $\sigma_{j,K} = \boldsymbol{e}_j^\top \boldsymbol{\Sigma}\boldsymbol{e}_K$, which completes the proof.

(3) Denote $\boldsymbol{c}_{\boldsymbol{M}}^\top(\mathbf{A}) = \boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{A})]^{-1}\mathbf{V}_{\boldsymbol{M}}(\mathbf{A})$ and $\boldsymbol{c}_{\boldsymbol{M}}^\top(\boldsymbol{\Sigma}) = \boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})]^{-1}\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})$. We have the following derivation:

$$
\begin{aligned}
&E[\{\boldsymbol{c}_{\boldsymbol{M}}^\top(\mathbf{A}) - \boldsymbol{c}_{\boldsymbol{M}}^\top(\boldsymbol{\Sigma})\}\boldsymbol{\Sigma}\{\boldsymbol{c}_{\boldsymbol{M}}(\mathbf{A}) - \boldsymbol{c}_{\boldsymbol{M}}(\boldsymbol{\Sigma})\}] \\
&= E[\boldsymbol{c}_{\boldsymbol{M}}^\top(\mathbf{A})\boldsymbol{\Sigma}\boldsymbol{c}_{\boldsymbol{M}}(\mathbf{A})] - E[\boldsymbol{c}_{\boldsymbol{M}}^\top(\mathbf{A})\boldsymbol{\Sigma}\boldsymbol{c}_{\boldsymbol{M}}(\boldsymbol{\Sigma})] - E[\boldsymbol{c}_{\boldsymbol{M}}^\top(\boldsymbol{\Sigma})\boldsymbol{\Sigma}\boldsymbol{c}_{\boldsymbol{M}}(\mathbf{A})] + E[\boldsymbol{c}_{\boldsymbol{M}}^\top(\boldsymbol{\Sigma})\boldsymbol{\Sigma}\boldsymbol{c}_{\boldsymbol{M}}(\boldsymbol{\Sigma})] \\
&= E[\boldsymbol{c}_{\boldsymbol{M}}^\top(\mathbf{A})\boldsymbol{\Sigma}\boldsymbol{c}_{\boldsymbol{M}}(\mathbf{A})] - \boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})]^{-1}\boldsymbol{e}_K,
\end{aligned}
$$

where the last equation comes from the fact that

$$
\begin{aligned}
\mathbf{V}_{\boldsymbol{M}}(\mathbf{A})\boldsymbol{\Sigma}\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma}) &= \mathbf{D}_{\boldsymbol{m}}(\mathbf{D}_{\boldsymbol{m}}^\top \mathbf{A}\mathbf{D}_{\boldsymbol{m}})^{-1}\mathbf{D}_{\boldsymbol{m}}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{m}}(\mathbf{D}_{\boldsymbol{m}}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{m}})^{-1}\mathbf{D}_{\boldsymbol{m}}^\top = \mathbf{V}_{\boldsymbol{M}}(\mathbf{A}) \\
\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})\boldsymbol{\Sigma}\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma}) &= \mathbf{D}_{\boldsymbol{m}}(\mathbf{D}_{\boldsymbol{m}}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{m}})^{-1}\mathbf{D}_{\boldsymbol{m}}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{m}}(\mathbf{D}_{\boldsymbol{m}}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{m}})^{-1}\mathbf{D}_{\boldsymbol{m}}^\top = \mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma}).
\end{aligned}
$$

Since $\boldsymbol{\Sigma}$ is positive definite, then we have $E[\boldsymbol{c}_{\boldsymbol{M}}^\top(\mathbf{A})\boldsymbol{\Sigma}\boldsymbol{c}_{\boldsymbol{M}}(\mathbf{A})] \geq \boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})]^{-1}\boldsymbol{e}_K$, which is the desired inequality. The equality holds if and only if $p_{\boldsymbol{m}}\{\boldsymbol{c}_{\boldsymbol{m}}^\top(\mathbf{A}) - \boldsymbol{c}_{\boldsymbol{M}}^\top(\boldsymbol{\Sigma})\} = \mathbf{0}_K^\top$ for all $\boldsymbol{m}\in\{0,1\}^K \setminus \{\mathbf{0}_K\}$.

(4) Define $\boldsymbol{x}_{\boldsymbol{M}}^\top = \boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})]^{-1}\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})\mathbf{B}^{\frac{1}{2}}$. Here $\mathbf{B}^{\frac{1}{2}}$ is well-defined since $\mathbf{B}$ is positive semi-definite. Then we have

$$
\begin{aligned}
&\boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})]^{-1}E[\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})\mathbf{B}\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})]E[\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})]^{-1}\boldsymbol{e}_K - \boldsymbol{e}_K^\top\mathbf{B}\boldsymbol{e}_K \\
&= E[\boldsymbol{x}_{\boldsymbol{M}}^\top\boldsymbol{x}_{\boldsymbol{M}}] - E[\boldsymbol{x}_{\boldsymbol{M}}]^\top E[\boldsymbol{x}_{\boldsymbol{M}}] \\
&= \sum_{t=1}^{K} Var(\boldsymbol{x}_{\boldsymbol{M}}^\top\boldsymbol{e}_t) \\
&\geq 0.
\end{aligned}
$$

The equality holds if and only if $Var(\boldsymbol{x}_{\boldsymbol{M}}^\top\boldsymbol{e}_t) = 0$ for $t = 1, \ldots, K$, which is equivalent to $\boldsymbol{x}_{\boldsymbol{M}}^\top$ being a constant vector.

(5) Define $\mathbf{B} = \mathbf{C} - \boldsymbol{\Sigma}$. The statement is proved by the following derivation:

$$
\begin{aligned}
&\boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{C})]^{-1}\boldsymbol{e}_K \\
&= \boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{C})]^{-1}E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{C})\mathbf{C}\mathbf{V}_{\boldsymbol{M}}(\mathbf{C})]E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{C})]^{-1}\boldsymbol{e}_K \\
&= \boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{C})]^{-1}E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{C})(\mathbf{B}+\boldsymbol{\Sigma})\mathbf{V}_{\boldsymbol{M}}(\mathbf{C})]E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{C})]^{-1}\boldsymbol{e}_K \\
&\geq \boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{C})]^{-1}E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{C})\mathbf{B}\mathbf{V}_{\boldsymbol{M}}(\mathbf{C})]E[\mathbf{V}_{\boldsymbol{M}}(\mathbf{C})]^{-1}\boldsymbol{e}_K + \boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})]^{-1}\boldsymbol{e}_K \\
&\geq \boldsymbol{e}_K^\top\mathbf{B}\boldsymbol{e}_K + \boldsymbol{e}_K^\top E[\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})]^{-1}\boldsymbol{e}_K,
\end{aligned}
$$

where the first inequality results from Lemma 1 (3), and the second inequality comes from Lemma 1 (4). □

Lemma 2 (Kronecker product)**.** *Let* $\mathbf{A} \in \mathbb{R}^{n_1\times n_2}, \mathbf{B} \in \mathbb{R}^{n_3\times n_4}, \mathbf{C} \in \mathbb{R}^{n_2\times n_5}, \mathbf{D} \in \mathbb{R}^{n_4\times n_6}$ *be random matrices. Then*

(1) $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = (\mathbf{AC})\otimes(\mathbf{BD})$,

(2) If $\mathbf{A}$ *is independent of* $(\mathbf{B}, \mathbf{C})$*, then* $E[\mathbf{A} \otimes \mathbf{B}] = E[\mathbf{A}] \otimes E[\mathbf{B}]$ *and* $E[(\mathbf{AC}) \otimes \mathbf{B}] = E[(E[\mathbf{A}]\mathbf{C}) \otimes \mathbf{B}]$*.*

(3) $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$ *if* $\mathbf{A}$ *and* $\mathbf{B}$ *are invertible.*

(4) $(\mathbf{A} \otimes \mathbf{B})^\top = \mathbf{A}^\top \otimes \mathbf{B}^\top$*.*

(5) Suppose $n_1 = n_2$*,* $n_3 = n_4$*,* $\mathbf{A}$ *has eigenvalues* $\lambda_1, \ldots, \lambda_{n_1}$*, and* $\mathbf{B}$ *has eigenvalues* $\mu_1, \ldots, \mu_{n_3}$*. Then* $\mathbf{A} \otimes \mathbf{B}$ *has eigenvalues* $\lambda_i \mu_j$ *for each* $i = 1, \ldots, n_1$ *and* $j = 1, \ldots, n_3$*.*

Lemma 3. *Given Assumption 1, for each* $j = 0, \ldots, J$*, let* $\boldsymbol{Z}_i(j) = \boldsymbol{h}_j(\boldsymbol{Y}_i(j), \boldsymbol{M}_i(j), \boldsymbol{X}_i) \in \mathbb{R}^q$ *for some function* $\boldsymbol{h}_j$ *such that* $E[||\boldsymbol{Z}_i(j)\boldsymbol{Z}_i(j)^\top||] < \infty$*. Then under stratified randomization,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=0}^{J} \left(I\{A_i = j\} \boldsymbol{Z}_i(j) - \pi_j E[\boldsymbol{Z}(j)] \right) \xrightarrow{d} N(\mathbf{0}, \mathbf{G}),$$

where

$$\mathbf{G} = \sum_{j=0}^{J} \pi_j E[Var\{\boldsymbol{Z}(j)|S\}] + Var\left(\sum_{j=0}^{J} \pi_j E[\boldsymbol{Z}(j)|S] \right).$$

Furthermore,

$$\sum_{j=0}^{J} \pi_j E[\boldsymbol{Z}(j)\boldsymbol{Z}(j)^\top] - E\left[\sum_{j=0}^{J} \pi_j \boldsymbol{Z}(j)\right] E\left[\sum_{j=0}^{J} \pi_j \boldsymbol{Z}(j)\right]^\top - \mathbf{G} = E[\mathbf{U}(diag\{\boldsymbol{\pi}\} - \boldsymbol{\pi}\boldsymbol{\pi}^\top)\mathbf{U}^\top]$$

is positive semi-definite, where

$$\mathbf{U} = (E[\boldsymbol{Z}(0)|S], \ldots, E[\boldsymbol{Z}(J)|S])$$
$$\boldsymbol{\pi} = (\pi_0, \ldots, \pi_J)^\top.$$

Proof. Let $\mathcal{S} = \{1, \dots, R\}$ denote the levels in S. Using the fact that $E[\boldsymbol{Z}_i(j)|S = S_i] = \sum_{s\in\mathcal{S}} I\{S_i = s\}E[\boldsymbol{Z}(j)|S = s]$ and $E[\boldsymbol{Z}_i(j)] = \sum_{s\in\mathcal{S}} P(S = s)E[\boldsymbol{Z}(j)|S = s]$, we have

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=0}^{J}\Big(I\{A_i = j\}\boldsymbol{Z}_i(j) - \pi_j E[\boldsymbol{Z}(j)]\Big) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=0}^{J}\sum_{s\in\mathcal{S}} I\{A_i = j, S_i = s\}\Big(\boldsymbol{Z}_i(j) - E[\boldsymbol{Z}(j)|S = S_i]\Big) \\
&\quad + \sum_{s\in\mathcal{S}}\sqrt{n}\left(\frac{\sum_{i=1}^n I\{S_i = s\}}{n} - P(S = s)\right)\sum_{j=0}^{J}\pi_j E[\boldsymbol{Z}(j)|S = s] \\
&\quad + \sum_{s\in\mathcal{S}}\sum_{j=0}^{J}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(I\{A_i = j, S_i = s\} - \pi_j I\{S_i = s\}\right)E[\boldsymbol{Z}(j)|S = s] \\
&= (\mathbf{1}_{(J+1)L}\otimes \mathbf{I}_q)^\top \mathbb{L}_n^{(1)} + \mathbf{u}^\top\mathbb{L}_n^{(2)} + \mathbf{v}_n^\top\mathbb{L}_n^{(3)},
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbb{L}_n^{(1)} &= \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I\{A_i = j, S_i = s\}\{\boldsymbol{Z}_i(j) - E[\boldsymbol{Z}(j)|S = S_i]\} : (j, s) \in \{0, \dots, J\}\times\mathcal{S}\right), \\
\mathbb{L}_n^{(2)} &= \left(\sqrt{n}\left\{\frac{\sum_{i=1}^n I\{S_i = s\}}{n} - P(S = s)\right\} : s\in\mathcal{S}\right), \\
\mathbb{L}_n^{(3)} &= \left(\sqrt{n}\left\{\frac{\sum_{i=1}^n I\{A_i = j, S_i = s\}}{\sum_{i=1}^n I\{S_i = s\}} - \pi_j\right\} : (j, s) \in \{0, \dots, J\}\times\mathcal{S}\right), \\
\mathbf{u} &= \left(\sum_{j=0}^{J}\pi_j E[\boldsymbol{Z}(j)|S = 1], \dots, \sum_{j=0}^{J}\pi_j E[\boldsymbol{Z}(j)|S = R]\right)^\top, \\
\mathbf{v}_n &= \left(\frac{\sum_{i=1}^n I\{S_i = s\}}{n}E[\boldsymbol{Z}(j)|S = s] : (j, s) \in \{0, \dots, J\}\times\mathcal{S}\right)^\top,
\end{aligned}
$$

where $(\boldsymbol{x}_{js} : (j, s) \in \{0, \dots, J\}\times\mathcal{S}) = (\boldsymbol{x}_{01}^\top, \dots, \boldsymbol{x}_{0R}^\top, \dots, \boldsymbol{x}_{J1}^\top, \dots, \boldsymbol{x}_{JR}^\top)^\top$ and $(\boldsymbol{x}_{js} : s \in \mathcal{S}) = (\boldsymbol{x}_{j1}^\top, \dots, \boldsymbol{x}_{jR}^\top)^\top$ for any vectors $\boldsymbol{x}_{js}\in\mathbb{R}^q$.

We next show that

$$\begin{pmatrix} \mathbb{L}_n^{(1)} \\ \mathbb{L}_n^{(2)} \\ \mathbb{L}_n^{(3)} \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & diag\{p_{\mathcal{S}}\} - p_{\mathcal{S}} p_{\mathcal{S}}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_{(J+1)L \times (J+1)L} \end{pmatrix}\right),$$

where

$$\boldsymbol{\Sigma}_1 = bdiag\{\pi_j P(S=s) Var\{Z(j)|S=s\} : (j,s) \in \{0,\ldots,J\} \times \mathcal{S}\},$$
$$p_{\mathcal{S}} = (P(S=1),\ldots,P(S=R))^\top,$$

where $bdiag\{\mathbf{V}_{js} : (j,s) \in \{0,\ldots,J\} \times \mathcal{S}\}$ represents a block diagonal matrix with $\mathbf{V}_{js}$ being the $\{(s-1)R+j+1\}$-th diagonal block. The proof can be found in Lemma C.1 and C.2 of Appendix C of Bugni et al. (2019). The only difference is that $\boldsymbol{Z}_i(j)$ is substituted for $Y_i(j)$ and all the arguments still hold.

By the delta method, we have $(\mathbf{1}_{(J+1)L} \otimes \mathbf{I}_q)^\top \mathbb{L}_n^{(1)} + \mathbf{u}^\top \mathbb{L}_n^{(2)} \xrightarrow{d} N(0, \mathbf{G})$ and $\mathbf{v}_n \xrightarrow{P} \mathbf{v}$ with $\mathbf{v} = (P(S=s)E[\boldsymbol{Z}(j)|S=s] : (j,s) \in \{0,\ldots,J\} \times \mathcal{S})$. Using Slutsky's theorem twice, we get the desired asymptotic normal distribution.

Finally, we have the following derivation:

$$
\begin{aligned}
&\sum_{j=0}^{J} \pi_j E[\boldsymbol{Z}(j)\boldsymbol{Z}(j)^\top] - E\left[\sum_{j=0}^{J} \pi_j \boldsymbol{Z}(j)\right] E\left[\sum_{j=0}^{J} \pi_j \boldsymbol{Z}(j)\right]^\top - \mathbf{G} \\
&= \sum_{j=0}^{J} \pi_j E[\boldsymbol{Z}(j)\boldsymbol{Z}(j)^\top] - E\left[\sum_{j=0}^{J} \pi_j \boldsymbol{Z}(j)\right] E\left[\sum_{j=0}^{J} \pi_j \boldsymbol{Z}(j)\right]^\top \\
&- \sum_{j=0}^{J} \pi_j E[\boldsymbol{Z}(j)\boldsymbol{Z}(j)^\top] + \sum_{j=0}^{J} \pi_j E[E[\boldsymbol{Z}(j)|S]E[\boldsymbol{Z}(j)|S]^\top] \\
&- E\left[E\left[\sum_{j=0}^{J} \pi_j \boldsymbol{Z}(j)\middle|S\right] E\left[\sum_{j=0}^{J} \pi_j \boldsymbol{Z}(j)\middle|S\right]^\top\right] + E\left[\sum_{j=0}^{J} \pi_j \boldsymbol{Z}(j)\right] E\left[\sum_{j=0}^{J} \pi_j \boldsymbol{Z}(j)\right]^\top \\
&= \sum_{j=0}^{J} \pi_j E[E[\boldsymbol{Z}(j)|S]E[\boldsymbol{Z}(j)|S]^\top] - E\left[E\left[\sum_{j=0}^{J} \pi_j \boldsymbol{Z}(j)\middle|S\right] E\left[\sum_{j=0}^{J} \pi_j \boldsymbol{Z}(j)\middle|S\right]^\top\right] \\
&= E[\mathbf{U}(diag\{\boldsymbol{\pi}\} - \boldsymbol{\pi}\boldsymbol{\pi}^\top)\mathbf{U}^\top].
\end{aligned}
$$

Since $diag\{\boldsymbol{\pi}\} \succeq \boldsymbol{\pi}\boldsymbol{\pi}^\top$, then we get $E[\mathbf{U}(diag\{\boldsymbol{\pi}\} - \boldsymbol{\pi}\boldsymbol{\pi}^\top)\mathbf{U}^\top]$ is positive semi-definite. □

Lemma 4. *Given Assumption 1, under simple or stratified randomization, each data vector* $(A_i, \boldsymbol{Y}_i, \boldsymbol{M}_i, \boldsymbol{X}_i)$ *is identically distributed and, for* $i = 1, \ldots, n$ *,* A_i *is independent of* $\boldsymbol{W}_i$ *and* $P(A_i = j) = \pi_j$*.*

Let P^* *denote the distribution of* $(A_i, \boldsymbol{Y}_i, \boldsymbol{M}_i, \boldsymbol{X}_i)$ *and* E^* *be the associated expectation. Define* $Z = f(\boldsymbol{Y}, \boldsymbol{M}, \boldsymbol{X})$ *and* $Z(j) = f(\boldsymbol{Y}(j), \boldsymbol{M}(j), \boldsymbol{X})$ *such that* $E[Z(j)^2] < \infty$ *for* $j = 0, \ldots, J$*. Then* $E^*[I\{A = j\}Z] = \pi_j E[Z(j)]$ *and* $E^*[I\{A = j\}Z|S] = \pi_j E[Z(j)|S]$ *for* $j = 0, \ldots, J$*.*

Proof. See Lemma 4 and Lemma 3 in the Supplementary of Wang et al. (2023). The only difference of proof is that $A = 1$ is substituted by $A = j$ for $j = 1, \ldots, J$, and $(\boldsymbol{Y}, \boldsymbol{M})$ are substituted for (Y, M). □

E Proofs

E.1 Proof of Theorem 1

Outline of the proof: Consider the estimator $\widehat{\boldsymbol{\Delta}}^{(\text{est})}$ for each est $\in$ {ANCOVA, MMRM-I, MMRM-II, IMMRM}. We first show that $\boldsymbol{\Delta}^{(\text{est})}$ is an M-estimator. We then apply Theorem 1 of Wang et al. (2023) to show that $\boldsymbol{\Delta}^{(\text{est})}$ is model-robust and asymptotically linear with influence function $IF^{(\text{est})}$. The influence function $IF^{(\text{est})}$ is the same under simple and stratified randomization. Next, we prove the asymptotic normality by Lemma 3, which is a central limit theorem for sums of random vectors under stratified randomization that generalizes Lemma B.2 of Bugni et al. (2019). Next, we calculate $IF^{(\text{est})}$ and derive the asymptotic covariance matrix i.e., $\mathbf{V}^{(\text{est})}$ and $\widetilde{\mathbf{V}}^{(\text{est})}$, for which the detailed algebra is given in Lemma 5 and Lemma 6. Finally, we compare the asymptotic covariance matrices, where Lemma 1 is used to handle missing data.

Proof of Theorem 1. The ANCOVA estimator can be computed by solving $P_n\boldsymbol{\psi}^{(\text{ANCOVA})}(A, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{M}; \boldsymbol{\theta}) = \mathbf{0}$, where

$$\boldsymbol{\psi}^{(\text{ANCOVA})}(A, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{M}; \boldsymbol{\theta}) = I\{M_K = 1\}(Y_K - \beta_0 - \sum_{j=1}^{J} \beta_{Aj} I\{A = j\} - \boldsymbol{\beta}_{\boldsymbol{X}}^\top \boldsymbol{X}) \begin{pmatrix} 1 \\ \boldsymbol{A} \\ \boldsymbol{X} \end{pmatrix}, \tag{1}$$

where $\boldsymbol{A} = (I\{A = 1\}, \ldots, I\{A = J\})^\top$ is a vector of treatment assignment indicator and $\boldsymbol{\theta} = (\beta_0, \beta_{A1}, \ldots, \beta_{AJ}, \boldsymbol{\beta}_{\boldsymbol{X}}^\top)^\top$. Hence $\widehat{\boldsymbol{\Delta}}^{(\text{ANCOVA})}$ is an M-estimator.

The MMRM-I and MMRM-II working models can be rewritten in one formula below:

$$\boldsymbol{Y} = \boldsymbol{\beta}_0 + (\mathbf{I}_K \otimes \boldsymbol{A})^\top \boldsymbol{\beta}_{\boldsymbol{A}} + \mathbf{u}(\boldsymbol{X})^\top \boldsymbol{\beta}_{\mathbf{u}(\boldsymbol{X})} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\boldsymbol{\beta}_0 = (\beta_{01}, \ldots, \beta_{0K})^\top \in \mathbb{R}^K, \boldsymbol{\beta}_A = (\beta_{A11}, \ldots, \beta_{AJ1}, \ldots, \beta_{A1K}, \ldots \beta_{AJK})^\top \in \mathbb{R}^{JK}, \boldsymbol{\beta}_{\mathbf{u}(\boldsymbol{X})} \in \mathbb{R}^q$ are column vectors of parameters, $\mathbf{u}(\boldsymbol{X}) \in \mathbb{R}^{q_{\mathbf{u}} \times K}$ is a matrix function of $\boldsymbol{X}$, and the error terms $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} \in \mathbb{R}^{K \times K}$ is a positive-definite covariance matrix. For MMRM-I, $q_{\mathbf{u}} = p$ (the dimension of $\boldsymbol{X}$), $\mathbf{u}(\boldsymbol{X}) = \boldsymbol{X}\mathbf{1}_K^\top$ and $\boldsymbol{\beta}_{\mathbf{u}(\boldsymbol{X})} = \boldsymbol{\beta}_{\boldsymbol{X}}$. For MMRM-II, $q = pK$, $\mathbf{u}(\boldsymbol{X}) = \mathbf{I}_K \otimes \boldsymbol{X}$ and $\boldsymbol{\beta}_{\mathbf{u}(\boldsymbol{X})} = (\boldsymbol{\beta}_{\boldsymbol{X}1}^\top, \ldots, \boldsymbol{\beta}_{\boldsymbol{X}K}^\top)^\top$.

Under the working model, the random error vectors $\boldsymbol{\varepsilon}_i, i = 1, \ldots, n$, are assumed to be independent (of each other and of $\{(A_i, \boldsymbol{X}_i)\}_{i=1}^n$), identically distributed draws from $N(\mathbf{0}, \boldsymbol{\Sigma})$. Denote $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_L)^\top \in \mathbb{R}^L$ is the vector of unknown parameters in $\boldsymbol{\Sigma}$. For example, $\boldsymbol{\alpha}$ consists of the lower triangular and diagonal entries of $\boldsymbol{\Sigma}$ if no structure is assumed on $\boldsymbol{\Sigma}$.

For each i, let $n_i = \sum_{t=1}^K M_{it}$ be the number of non-missing outcomes and $\boldsymbol{Y}_i^o \in \mathbb{R}^{n_i}$ be the observed outcomes if $n_i > 0$. Let $t_{i,1} < \cdots < t_{i,n_i}$ denote the ordered list of visits when the outcomes are not missing. For example, $t_{i,1}$ is the first non-missing visit for subject i. We define $\mathbf{D}_{\boldsymbol{M}_i} = [\boldsymbol{e}_{t_{i,1}} \ \boldsymbol{e}_{t_{i,2}} \ \ldots \ \boldsymbol{e}_{t_{i,n_i}}] \in \mathbb{R}^{K \times n_i}$. We use the subscript $\boldsymbol{M}_i$ to note that $\mathbf{D}_{\boldsymbol{M}_i}$ is a deterministic function of $\boldsymbol{M}_i$. Then $\boldsymbol{Y}_i^o = \mathbf{D}_{\boldsymbol{M}_i}^\top \boldsymbol{Y}_i$. The observed data vector for each i is $(\boldsymbol{Y}_i^o, \boldsymbol{M}_i, A_i, \boldsymbol{X}_i)$.

Denote the full set of parameters as $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}_0^\top, \boldsymbol{\beta}_A^\top, \boldsymbol{\beta}_{\mathbf{u}(\boldsymbol{X})}^\top)^\top$. We further define $\mathbf{Q} = [\mathbf{I}_K \ (\mathbf{I}_K \otimes \boldsymbol{A})^\top \ \mathbf{u}(\boldsymbol{X})^\top]^\top$. We let $\mathbf{Q}_i$ denote $\mathbf{Q}$ with $\boldsymbol{A}_i, \boldsymbol{X}_i$ substituted for $\boldsymbol{A}, \boldsymbol{X}$. It follows that $\mathbf{Q}_i^\top \boldsymbol{\beta} = \boldsymbol{\beta}_0 + (\mathbf{I}_K \otimes \boldsymbol{A}_i)^\top \boldsymbol{\beta}_{\boldsymbol{A}} + \mathbf{u}(\boldsymbol{X}_i)^\top \boldsymbol{\beta}_{\mathbf{u}(\boldsymbol{X})}$. Then we have $\boldsymbol{Y}_i^o | (A_i, \boldsymbol{X}_i, \boldsymbol{M}_i, \boldsymbol{M}_i \neq \mathbf{0}_K) \sim N(\mathbf{D}_{\boldsymbol{M}_i}^\top \mathbf{Q}_i^\top \boldsymbol{\beta}, \mathbf{D}_{\boldsymbol{M}_i}^\top \boldsymbol{\Sigma} \mathbf{D}_{\boldsymbol{M}_i})$ under the MMRM-I or MMRM-II working model assumptions and missing completely at random (MCAR). The corresponding log likelihood function conditional on $\{A_i, \boldsymbol{X}_i, \boldsymbol{M}_i\}_{i=1}^n$ is a constant independent of the parameter vector $\boldsymbol{\theta}$ plus the following:

$$
\begin{aligned}
&-\frac{1}{2}\sum_{i=1}^{n} I\{\boldsymbol{M}_i \neq \mathbf{0}_K\}\left\{\log|\mathbf{D}_{\boldsymbol{M}_i}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{M}_i}| + (\boldsymbol{Y}_i^o - \mathbf{D}_{\boldsymbol{M}_i}^\top \mathbf{Q}_i^\top \boldsymbol{\beta})^\top (\mathbf{D}_{\boldsymbol{M}_i}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{M}_i})^{-1}(\boldsymbol{Y}_i^o - \mathbf{D}_{\boldsymbol{M}_i}^\top \mathbf{Q}_i^\top \boldsymbol{\beta})\right\} \\
&= -\frac{1}{2}\sum_{i=1}^{n} I\{\boldsymbol{M}_i \neq \mathbf{0}_K\}\left\{\log|\mathbf{D}_{\boldsymbol{M}_i}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{M}_i}| + (\boldsymbol{Y}_i - \mathbf{Q}_i^\top \boldsymbol{\beta})^\top \mathbf{D}_{\boldsymbol{M}_i}(\mathbf{D}_{\boldsymbol{M}_i}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{M}_i})^{-1}\mathbf{D}_{\boldsymbol{M}_i}^\top(\boldsymbol{Y}_i - \mathbf{Q}_i^\top \boldsymbol{\beta})\right\} \\
&= \frac{n}{2} P_n l(\boldsymbol{\theta}; \boldsymbol{Y}^o | A, \boldsymbol{X}, \boldsymbol{M}),
\end{aligned}
$$

where we define

$$
\begin{aligned}
&l(\boldsymbol{\theta}; \boldsymbol{Y}^o | A, \boldsymbol{X}, \boldsymbol{M}) \\
&= -I\{\boldsymbol{M} \neq \mathbf{0}_K\}\left\{\log|\mathbf{D}_{\boldsymbol{M}}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{M}}| + (\boldsymbol{Y} - \mathbf{Q}^\top \boldsymbol{\beta})^\top \mathbf{D}_{\boldsymbol{M}}(\mathbf{D}_{\boldsymbol{M}}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{M}})^{-1}\mathbf{D}_{\boldsymbol{M}}^\top(\boldsymbol{Y} - \mathbf{Q}^\top \boldsymbol{\beta})\right\}.
\end{aligned}
$$

To derive the estimating functions for the corresponding maximum likelihood estimator, we use the following results to compute the differential of $l(\boldsymbol{\theta}; \boldsymbol{Y}^o | A, \boldsymbol{X}, \boldsymbol{M})$ with respect to $\boldsymbol{\theta}$. By Equation (8.7) of Dwyer (1967), we have $\frac{\partial \log(|\mathbf{D}_{\boldsymbol{M}}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{M}}|)}{\partial \boldsymbol{\Sigma}} = \mathbf{D}_{\boldsymbol{M}}(\mathbf{D}_{\boldsymbol{M}}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{M}})^{-1}\mathbf{D}_{\boldsymbol{M}}^\top$. Using the chain rule of matrix derivatives (MacRae et al., 1974, Theorem 8), we have

$$
\frac{\partial \log(|\mathbf{D}_{\boldsymbol{M}}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{M}}|)}{\partial \alpha_j} = \operatorname{tr}\left(\frac{\partial \log(|\mathbf{D}_{\boldsymbol{M}}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{M}}|)}{\partial \boldsymbol{\Sigma}}\frac{\partial \boldsymbol{\Sigma}}{\partial \alpha_j}\right) = \operatorname{tr}\left(\mathbf{D}_{\boldsymbol{M}}(\mathbf{D}_{\boldsymbol{M}}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{M}})^{-1}\mathbf{D}_{\boldsymbol{M}}^\top \frac{\partial \boldsymbol{\Sigma}}{\partial \alpha_j}\right).
$$

By Theorem 5 of MacRae et al. (1974), we have

$$
\frac{\partial (\mathbf{D}_{\boldsymbol{M}}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{M}})^{-1}}{\partial \alpha_j} = -(\mathbf{D}_{\boldsymbol{M}}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{M}})^{-1}\mathbf{D}_{\boldsymbol{M}}^\top \frac{\partial \boldsymbol{\Sigma}}{\partial \alpha_j}\mathbf{D}_{\boldsymbol{M}}(\mathbf{D}_{\boldsymbol{M}}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{M}})^{-1}.
$$

Denoting $\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma}) = I\{\boldsymbol{M} \neq \mathbf{0}_K\}\mathbf{D}_{\boldsymbol{M}}(\mathbf{D}_{\boldsymbol{M}}^\top \boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{M}})^{-1}\mathbf{D}_{\boldsymbol{M}}^\top$, we have shown that

$$
\frac{\partial \mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})}{\partial \alpha_j} = -\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})\frac{\partial \boldsymbol{\Sigma}}{\partial \alpha_j}\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma}).
$$

Using the above results, the estimating functions for the MLE $\widehat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ under MMRM-I or MMRM-II are

$$\boldsymbol{\psi}^{(\mathrm{MMRM})}(A, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{M}; \boldsymbol{\theta})$$
$$= \left(\begin{array}{c} \mathbf{Q}\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{Y} - \mathbf{Q}^\top \boldsymbol{\beta}) \\ -\mathrm{tr}(\mathbf{V}_{\boldsymbol{M}} \frac{\partial \boldsymbol{\Sigma}}{\partial \alpha_l}) + (\boldsymbol{Y} - \mathbf{Q}^\top \boldsymbol{\beta})^\top \mathbf{V}_{\boldsymbol{M}} \frac{\partial \boldsymbol{\Sigma}}{\partial \alpha_l} \mathbf{V}_{\boldsymbol{M}} (\boldsymbol{Y} - \mathbf{Q}^\top \boldsymbol{\beta}),\ l = 1, \dots, L \end{array} \right), \tag{3}$$

which implies that $\widehat{\boldsymbol{\Delta}}^{(\mathrm{MMRM\text{-}I})}$ and $\widehat{\boldsymbol{\Delta}}^{(\mathrm{MMRM\text{-}II})}$ are M-estimators. In the above expression of $\boldsymbol{\psi}^{(\mathrm{MMRM})}$, we omit $(\boldsymbol{\Sigma})$ from $\mathbf{V}_{\boldsymbol{M}}(\boldsymbol{\Sigma})$ for conciseness. We note that $\mathbf{V}_{\boldsymbol{M}}$ is a random matrix taking values in $\mathbf{R}^{K \times K}$ and defined in the same way as in Lemma 1.

The IMMRM working model (3) can be written as

$$\boldsymbol{Y} = \boldsymbol{\beta}_0 + (\mathbf{I}_K \otimes \boldsymbol{A})^\top \boldsymbol{\beta}_{\boldsymbol{A}} + (\mathbf{I}_K \otimes \boldsymbol{X})^\top \boldsymbol{\beta}_{\mathbf{I}_K \otimes \boldsymbol{X}} + (\mathbf{I}_K \otimes \boldsymbol{X} \otimes \boldsymbol{A})^\top \boldsymbol{\beta}_{\boldsymbol{A}\boldsymbol{X}} + \boldsymbol{\varepsilon}_{\boldsymbol{A}}, \tag{4}$$

where $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_A$, $\boldsymbol{A}$ are defined in Equation (2), $\boldsymbol{\beta}_{\mathbf{I}_K \otimes \boldsymbol{X}} = (\boldsymbol{\beta}_{\boldsymbol{X}1}^\top, \dots, \boldsymbol{\beta}_{\boldsymbol{X}K}^\top)^\top$, $\boldsymbol{\beta}_{\boldsymbol{A}\boldsymbol{X}}^\top \in \mathbb{R}^{JpK}$ with the $\{Jp(k-1) + J(m-1) + j\}$-th entry being $\beta_{\boldsymbol{A}X_m jk}$ for $j = 1, \dots, J, k = 1, \dots, K$ and $m = 1, \dots, p$, and $\boldsymbol{\varepsilon}_{\boldsymbol{A}} = \sum_{j=0}^J I\{A = j\} \boldsymbol{\varepsilon}_j$, where $\boldsymbol{\varepsilon}_j \sim N(\mathbf{0}, \boldsymbol{\Sigma}_j)$ and $(\boldsymbol{\varepsilon}_0, \dots, \boldsymbol{\varepsilon}_J)$ are independent of each other. Let $\boldsymbol{\alpha}_j \in \mathbb{R}^L$ be the unknown parameters in $\boldsymbol{\Sigma}_j$ for $j = 0, \dots, J$. We define $\boldsymbol{\gamma} = (\boldsymbol{\beta}_0^\top, \boldsymbol{\beta}_{\boldsymbol{A}}^\top, \boldsymbol{\beta}_{\mathbf{I}_K \otimes \boldsymbol{X}}^\top, \boldsymbol{\beta}_{\boldsymbol{A}\boldsymbol{X}}^\top)^\top$ and $\boldsymbol{\theta} = (\boldsymbol{\Delta}, \boldsymbol{\gamma}^\top, \boldsymbol{\alpha}_0^\top, \dots, \boldsymbol{\alpha}_J^\top)^\top$. Following a similar procedure to MMRM-I, the estimating functions for the MLE $\widehat{\boldsymbol{\theta}}^{(\mathrm{IMMRM})}$ under IMMRM are

$$\boldsymbol{\psi}^{(\mathrm{IMMRM})}(A, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{M}; \boldsymbol{\theta})$$
$$= \left(\begin{array}{c} \beta_{AjK} + \boldsymbol{X}^\top \boldsymbol{\beta}_{A\boldsymbol{X}jK} - \Delta_j,\ j = 1, \dots, J \\ \mathbf{R}\mathbf{V}_{A\boldsymbol{M}}(\mathbf{Y} - \mathbf{R}^\top \boldsymbol{\gamma}) \\ I\{A = j\} \left(-\mathrm{tr}(\mathbf{V}_{A\boldsymbol{M}} \frac{\partial \boldsymbol{\Sigma}_j}{\partial \alpha_{jl}}) + (\boldsymbol{Y} - \mathbf{R}^\top \boldsymbol{\gamma})^\top \mathbf{V}_{A\boldsymbol{M}} \frac{\partial \boldsymbol{\Sigma}_j}{\partial \alpha_{jl}} \mathbf{V}_{A\boldsymbol{M}} (\boldsymbol{Y} - \mathbf{R}^\top \boldsymbol{\gamma}) \right), \\ \hfill j = 0, \dots, J,\ l = 1, \dots, L \end{array} \right), \tag{5}$$

where $\mathbf{R} = [\mathbf{I}_K \quad (\mathbf{I}_K \otimes \boldsymbol{A})^\top \quad (\mathbf{I}_K \otimes \boldsymbol{X})^\top \quad (\mathbf{I}_K \otimes \boldsymbol{X} \otimes \boldsymbol{A})^\top]^\top$ and
$\mathbf{V}_{A\boldsymbol{M}} = \mathbf{V}_{\boldsymbol{M}}(\sum_{j=0}^J I\{A = j\} \boldsymbol{\Sigma}_j) = I\{\boldsymbol{M} \neq \mathbf{0}_K\} \mathbf{D}_{\boldsymbol{M}} (\sum_{j=0}^J I\{A = j\} \mathbf{D}_{\boldsymbol{M}}^\top \boldsymbol{\Sigma}_j \mathbf{D}_{\boldsymbol{M}})^{-1} \mathbf{D}_{\boldsymbol{M}}^\top$. Hence $\widehat{\boldsymbol{\Delta}}^{(\mathrm{IMMRM})}$, as the first J entries of $\widehat{\boldsymbol{\theta}}^{(\mathrm{IMMRM})}$, is an M-estimator.

For each est $\in$ {ANCOVA, MMRM-I, MMRM-II, IMMRM}, we have just shown that $\widehat{\boldsymbol{\Delta}}^{(\text{est})}$ is an M-estimator. By Assumption 1 and regularity conditions, we apply Theorem 1 of Wang et al. (2023) and get, under simple or stratified randomization,

$$\sqrt{n}(\widehat{\boldsymbol{\Delta}}^{(\text{est})} - \underline{\boldsymbol{\Delta}}^{(\text{est})}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} IF^{(\text{est})}(A_i, \boldsymbol{X}_i, \boldsymbol{Y}_i, \boldsymbol{M}_i) + o_p(\mathbf{1}), \tag{6}$$

where $\underline{\boldsymbol{\Delta}}^{(\text{est})}$ satisfies $E^*[\boldsymbol{\psi}^{(\text{est})}(A, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{M}; \boldsymbol{\theta})] = \mathbf{0}$ with E^* defined in Lemma 4, and $IF^{(\text{est})}$ represents the J-dimensional influence function. We note that Theorem 1 of Wang et al. (2023) is developed for binary treatment (i.e. $J = 1$) and scalar outcome (i.e. $K = 1$), but their proof can be easily generalized to accommodate multiple treatment arms and repeated measured outcomes (as in Example 3 of Wang et al., 2023).

We next show $\widehat{\boldsymbol{\Delta}}^{(\text{MMRM-I})}$ is model-robust. By $E^*[\boldsymbol{\psi}^{(\text{MMRM})}(A, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{M}; \underline{\boldsymbol{\theta}})] = \mathbf{0}$, we have

$$E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}\{\boldsymbol{Y} - \underline{\boldsymbol{\beta}}_0 - (\mathbf{I}_K \otimes \boldsymbol{A})^\top \underline{\boldsymbol{\beta}}_{\boldsymbol{A}} - \mathbf{u}(\boldsymbol{X})^\top \underline{\boldsymbol{\beta}}_{\mathbf{u}(\boldsymbol{X})}\}] = \mathbf{0},$$
$$E^*[(\mathbf{I}_K \otimes \boldsymbol{A})\underline{\mathbf{V}}_{\boldsymbol{M}}\{\boldsymbol{Y} - \underline{\boldsymbol{\beta}}_0 - (\mathbf{I}_K \otimes \boldsymbol{A})^\top \underline{\boldsymbol{\beta}}_{\boldsymbol{A}} - \mathbf{u}(\boldsymbol{X})^\top \underline{\boldsymbol{\beta}}_{\mathbf{u}(\boldsymbol{X})}\}] = \mathbf{0},$$

which are first $2K$ equations in $E^*[\boldsymbol{\psi}(A, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{M}; \underline{\boldsymbol{\theta}})] = \mathbf{0}$, where $\underline{\mathbf{V}}_{\boldsymbol{M}} = \mathbf{V}_{\boldsymbol{M}}(\underline{\boldsymbol{\Sigma}}^{(\text{MMRM-I})})$ with $\underline{\boldsymbol{\Sigma}}^{(\text{MMRM-I})}$ being the probability limit of $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\alpha}})$ under MMRM-I. By the MCAR assumption and Lemma 4, the first K equations imply that

$$E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]\sum_{j=0}^{J}\{E[\boldsymbol{Y}(j)] - \underline{\boldsymbol{\beta}}_0 - \underline{\boldsymbol{\beta}}_{Aj} - E[\mathbf{u}(\boldsymbol{X})^\top \underline{\boldsymbol{\beta}}_{\mathbf{u}(\boldsymbol{X})}]\}\pi_j = \mathbf{0},$$

where $\boldsymbol{\beta}_{Aj} = (\boldsymbol{\beta}_{Aj1}, \ldots, \boldsymbol{\beta}_{AjK})^\top$ for $j = 1, \ldots, J$ and $\boldsymbol{\beta}_{Aj} = \mathbf{0}_K$. Similarly, the $(K+1) - 2K$th equations imply that, for $j = 1, \ldots, K$ and $t = 1, \ldots, K$,

$$E^*[\boldsymbol{e}_t^\top \underline{\mathbf{V}}_{\boldsymbol{M}}]\{E[\boldsymbol{Y}(j)] - \underline{\boldsymbol{\beta}}_0 - \underline{\boldsymbol{\beta}}_{Aj} - E[\mathbf{u}(\boldsymbol{X})^\top \underline{\boldsymbol{\beta}}_{\mathbf{u}(\boldsymbol{X})}]\}\pi_j = \mathbf{0}.$$

The above two sets of equations and the positivtiy assumption imply that, for $j = 0, \ldots, J$,

$$E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]\{E[\boldsymbol{Y}(j)] - \underline{\boldsymbol{\beta}}_0 - \underline{\boldsymbol{\beta}}_{Aj} - E[\mathbf{u}(\boldsymbol{X})^\top \underline{\boldsymbol{\beta}}_{\mathbf{u}(\boldsymbol{X})}]\} = \mathbf{0}.$$

The assumption $P(\boldsymbol{M}(j) = \mathbf{1}_K) > 0$ and Lemma 1 (1) implies that $E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]$ is invertible. Then the above equations imply that, for $j = 1, \ldots, J$,

$$\underline{\boldsymbol{\beta}}_{Aj} = \underline{\boldsymbol{\beta}}_{Aj} - \underline{\boldsymbol{\beta}}_{A0} = E[\boldsymbol{Y}(j)] - E[\boldsymbol{Y}(0)]$$

and hence $\underline{\boldsymbol{\Delta}}^{(\text{MMRM-I})} = (\underline{\beta}_{A1K}, \ldots, \underline{\beta}_{AJK})^\top = \boldsymbol{\Delta}^*$. The above proof also applies to the MMRM-II estimator by substituting $\mathbf{V}_{\boldsymbol{M}}(\underline{\boldsymbol{\Sigma}}^{(\text{MMRM-II})})$ for $\underline{\mathbf{V}}_{\boldsymbol{M}}$, which implies that $\widehat{\boldsymbol{\Delta}}^{(\text{MMRM-II})}$ is model-robust. Also, since the ANCOVA estimator is a special case of the MMRM-II estimator setting $K = 1$, we get that the ANCOVA estimator is model-robust.

Following a similar procedure, we next show that $\widehat{\boldsymbol{\Delta}}^{(\text{IMMRM})}$ is model-robust. We have that, for $j = 1, \ldots, J$,

$$E[\underline{\mathbf{V}}_{j\boldsymbol{M}}]\{E[\boldsymbol{Y}(j)] - \underline{\boldsymbol{\beta}}_0 - \underline{\boldsymbol{\beta}}_{Aj} - E[(\mathbf{I}_K \otimes \boldsymbol{X})^\top \underline{\boldsymbol{\beta}}_{\mathbf{I}_K \otimes \boldsymbol{X}}] - E[(\mathbf{I}_K \otimes \boldsymbol{X} \otimes \boldsymbol{e}_j)^\top \underline{\boldsymbol{\beta}}_{\boldsymbol{A}\boldsymbol{X}}]\}\pi_j = \mathbf{0},$$

and

$$E[\underline{\mathbf{V}}_{0\boldsymbol{M}}]\{E[\boldsymbol{Y}(0)] - \underline{\boldsymbol{\beta}}_0 - E[(\mathbf{I}_K \otimes \boldsymbol{X})^\top \underline{\boldsymbol{\beta}}_{\mathbf{I}_K \otimes \boldsymbol{X}}]\}\pi_0 = \mathbf{0},$$

where $\underline{\mathbf{V}}_{j\boldsymbol{M}} = I\{\boldsymbol{M}(j) \neq \mathbf{0}_K\}\mathbf{D}_{\boldsymbol{M}(j)}(\mathbf{D}_{\boldsymbol{M}(j)}^\top \underline{\boldsymbol{\Sigma}}_j^{(\text{IMMRM})} \mathbf{D}_{\boldsymbol{M}(j)})^{-1}\mathbf{D}_{\boldsymbol{M}(j)}^\top$ with $\underline{\boldsymbol{\Sigma}}_j^{(\text{IMMRM})}$ being the probability limit of $\boldsymbol{\Sigma}_j(\widehat{\boldsymbol{\alpha}}_j)$ in the IMMRM model (3) for $j = 0, \ldots, J$. The assumption $P(\boldsymbol{M}(j) = \mathbf{1}_K) > 0$ and Lemma 1 (1) implies that $E[\underline{\mathbf{V}}_{j\boldsymbol{M}}]$ is invertible. Thus, for $j = 1, \ldots, J$,

$$E[\boldsymbol{Y}(j)] - E[\boldsymbol{Y}(0)] = \underline{\boldsymbol{\beta}}_{Aj} + E[(\mathbf{I}_K \otimes \boldsymbol{X} \otimes \boldsymbol{e}_j)^\top \underline{\boldsymbol{\beta}}_{\boldsymbol{A}\boldsymbol{X}}],$$

which implies $E[Y_K(j)] - E[Y_K(0)] = \underline{\beta}_{AjK} + E[\boldsymbol{X}]^\top \underline{\boldsymbol{\beta}}_{\boldsymbol{AX}jK}$. Since the first equation of $\boldsymbol{\psi}^{(\text{IMMRM})}$ indicates that $\underline{\beta}_{AjK} + E[\boldsymbol{X}]^\top \underline{\boldsymbol{\beta}}_{\boldsymbol{AX}jK} = \underline{\Delta}_j$, we get $\underline{\Delta}_j = E[Y_K(j)] - E[Y_K(0)] = \Delta_j^*$, which completes the proof of model-robustness of $\widehat{\boldsymbol{\Delta}}^{(\text{IMMRM})}$.

We next prove that $\sqrt{n}(\widehat{\boldsymbol{\Delta}}^{(\text{est})} - \boldsymbol{\Delta}^*)$ weakly converges to a normal distribution, under simple or stratified randomization. Given Equations (6), it suffices to show that $\frac{1}{\sqrt{n}}\sum_{i=1}^n IF^{(\text{est})}(A_i, \boldsymbol{X}_i, \boldsymbol{Y}_i, \boldsymbol{M}_i)$ weakly converges to a normal distribution. Under simple randomization, $(A_i, \boldsymbol{X}_i, \boldsymbol{Y}_i, \boldsymbol{M}_i), i = 1, \ldots, n$ are independent to each other and identically distributed. Since the regularity conditions implied that $IF^{(\text{est})}$ has finite second moment, then the central limit theorem implies the desired weak convergence. Furthermore, we have $\widetilde{\mathbf{V}}^{(\text{est})} = Var^*(IF^{(\text{est})}(A, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{M}))$. Under stratified randomization, we define $\boldsymbol{Z}_i(j) = IF^{(\text{est})}(j, \boldsymbol{X}_i, \boldsymbol{Y}_i(j), \boldsymbol{M}_i(j))$ for $j = 0, \ldots, J$. Then $IF^{(\text{est})}(A_i, \boldsymbol{X}_i, \boldsymbol{Y}_i, \boldsymbol{M}_i) = \sum_{j=0}^J I\{A_i = j\}\boldsymbol{Z}_i(j)$. Since $E^*[IF^{(\text{est})}(A, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{M})] = \mathbf{0}$, Lemma 4 implies that $\sum_{j=0}^J \pi_j E[\boldsymbol{Z}_i(j)] = \mathbf{0}$. By the regularity conditions, we apply Lemma 3 and get

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n IF^{(\text{est})}(A_i, \boldsymbol{X}_i, \boldsymbol{Y}_i, \boldsymbol{M}_i) = \frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_{j=0}^J (I\{A_i = j\}\boldsymbol{Z}_i(j) - \pi_j E[\boldsymbol{Z}_i(j)]) \xrightarrow{d} N(\mathbf{0}, \mathbf{G}),$$

which completes the proof of asymptotic normality. In addition, Lemma 3 also implies that $\widetilde{\mathbf{V}}^{(\text{est})} \succeq \mathbf{V}^{(\text{est})}$.

For the ANCOVA, MMRM-I, MMRM-II and IMMRM estimators, the influence functions by Lemmas 5 and 6 are given below:

$$IF^{(\text{ANCOVA})} = \mathbf{LT}^{(\text{ANCOVA})}\left\{\boldsymbol{Y} - \boldsymbol{h}(A, \boldsymbol{X})\right\}, \tag{7}$$

$$IF^{(\text{MMRM-I})} = \mathbf{LT}^{(\text{MMRM-I})}(\boldsymbol{Y} - \mathbf{Q}^\top \underline{\boldsymbol{\beta}}), \tag{8}$$

$$IF^{(\text{MMRM-II})} = \mathbf{LT}^{(\text{MMRM-II})}\left\{\boldsymbol{Y} - \boldsymbol{h}(A, \boldsymbol{X})\right\}, \tag{9}$$

$$IF^{(\text{IMMRM})} = \mathbf{LT}^{(\text{IMMRM})}(\boldsymbol{Y} - \mathbf{R}^\top \underline{\boldsymbol{\gamma}}) + \mathbf{Lr}^\top(\boldsymbol{X} - E[\boldsymbol{X}]), \tag{10}$$

where

$$
\begin{aligned}
\mathbf{L} &= (-\mathbf{1}_J \quad \mathbf{I}_J) \in \mathbb{R}^{J\times(J+1)}, \\
\mathbf{T}^{(\mathrm{ANCOVA})} &= \left(\frac{I\{A=0\}}{\pi_0}\frac{I\{M_K=1\}}{P^*(M_K=1)}\boldsymbol{e}_K, \dots, \frac{I\{A=J\}}{\pi_J}\frac{I\{M_K=1\}}{P^*(M_K=1)}\boldsymbol{e}_K\right)^\top, \\
\mathbf{T}^{(\mathrm{MMRM\text{-}I})} &= \left(\frac{I\{A=0\}}{\pi_0}\underline{\mathbf{V}}_{\boldsymbol{M}}E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1}\boldsymbol{e}_K, \dots, \frac{I\{A=J\}}{\pi_J}\underline{\mathbf{V}}_{\boldsymbol{M}}E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1}\boldsymbol{e}_K\right)^\top, \\
\mathbf{T}^{(\mathrm{MMRM\text{-}II})} &= \left(\frac{I\{A=0\}}{\pi_0}\overline{\underline{\mathbf{V}}}_{\boldsymbol{M}}E^*[\overline{\underline{\mathbf{V}}}_{\boldsymbol{M}}]^{-1}\boldsymbol{e}_K, \dots, \frac{I\{A=J\}}{\pi_J}\overline{\underline{\mathbf{V}}}_{\boldsymbol{M}}E^*[\overline{\underline{\mathbf{V}}}_{\boldsymbol{M}}]^{-1}\boldsymbol{e}_K\right)^\top, \\
\mathbf{T}^{(\mathrm{IMMRM})} &= \left(\frac{I\{A=0\}}{\pi_0}\underline{\mathbf{V}}_{0\boldsymbol{M}}E^*[\underline{\mathbf{V}}_{0\boldsymbol{M}}]^{-1}\boldsymbol{e}_K, \dots, \frac{I\{A=J\}}{\pi_J}\underline{\mathbf{V}}_{J\boldsymbol{M}}E^*[\underline{\mathbf{V}}_{J\boldsymbol{M}}]^{-1}\boldsymbol{e}_K\right)^\top, \\
\boldsymbol{Y} - \boldsymbol{h}(A,\boldsymbol{X}) &= \sum_{j=0}^{J} I\{A=j\}\left\{\boldsymbol{Y} - E[\boldsymbol{Y}(j)] - Cov^*(\boldsymbol{Y},\boldsymbol{X})\,Var(\boldsymbol{X})^{-1}(\boldsymbol{X}-E[\boldsymbol{X}])\right\}, \\
\boldsymbol{Y} - \mathbf{Q}^\top\underline{\boldsymbol{\beta}} &= \sum_{j=0}^{J} I\{A=j\}\left\{\boldsymbol{Y}(j) - E[\boldsymbol{Y}(j)] - \mathbf{1}_K(\boldsymbol{X}-E[\boldsymbol{X}])^\top\underline{\boldsymbol{\beta}}_{\boldsymbol{X}}\right\}, \\
\boldsymbol{Y} - \mathbf{R}^\top\underline{\boldsymbol{\gamma}} &= \sum_{j=0}^{J} I\{A=j\}\left\{\boldsymbol{Y}(j) - E[\boldsymbol{Y}(j)] - Cov(\boldsymbol{Y}(j),\boldsymbol{X})Var(\boldsymbol{X})^{-1}(\boldsymbol{X}-E[\boldsymbol{X}])\right\}, \\
\mathbf{r} &= (\boldsymbol{b}_{K0}, \dots, \boldsymbol{b}_{KJ}),
\end{aligned}
$$

where

$$
\begin{aligned}
\underline{\mathbf{V}}_{\boldsymbol{M}} &= \underline{\mathbf{V}}_{\boldsymbol{M}}(\underline{\boldsymbol{\Sigma}}^{(\mathrm{MMRM\text{-}I})}) = \mathbf{V}_{\boldsymbol{M}}(E^*[(\boldsymbol{Y}-\mathbf{Q}^\top\underline{\boldsymbol{\beta}})(\boldsymbol{Y}-\mathbf{Q}^\top\underline{\boldsymbol{\beta}})^\top]), \\
\overline{\underline{\mathbf{V}}}_{\boldsymbol{M}} &= \underline{\mathbf{V}}_{\boldsymbol{M}}(\underline{\boldsymbol{\Sigma}}^{(\mathrm{MMRM\text{-}II})}) = \mathbf{V}_{\boldsymbol{M}}(E^*[\{\boldsymbol{Y}-h(A,\boldsymbol{X})\}\{\boldsymbol{Y}-h(A,\boldsymbol{X})\}^\top]), \\
\underline{\mathbf{V}}_{j\boldsymbol{M}} &= \underline{\mathbf{V}}_{\boldsymbol{M}}(\underline{\boldsymbol{\Sigma}}_j^{(\mathrm{IMMRM})}) = \mathbf{V}_{\boldsymbol{M}}\left(E^*\left[\frac{I\{A=j\}}{\pi_j}(\boldsymbol{Y}-\mathbf{R}^\top\underline{\boldsymbol{\gamma}})(\boldsymbol{Y}-\mathbf{R}^\top\underline{\boldsymbol{\gamma}})^\top\right]\right),\ ,j=0,\dots,J, \\
\underline{\boldsymbol{\beta}}_{\boldsymbol{X}} &= Var(\boldsymbol{X})^{-1}Cov^*(\boldsymbol{Y},\boldsymbol{X})\frac{E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]\mathbf{1}_K}{\mathbf{1}_K^\top E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]\mathbf{1}_K}, \\
\boldsymbol{b}_{Kj} &= Var(\boldsymbol{X})^{-1}Cov(\boldsymbol{X}, Y_K(j)),\ j=0,\dots,J.
\end{aligned}
$$

Furthermore, we have

$$\widetilde{\mathbf{V}}^{(\text{ANCOVA})} = \frac{1}{P^*(M_K=1)}\mathbf{L}\left(diag\{\pi_j^{-1}\boldsymbol{e}_K^\top\underline{\boldsymbol{\Sigma}}_j^{(\text{ANCOVA})}\boldsymbol{e}_K : j=0,\dots,J\}\right)\mathbf{L}^\top, \tag{11}$$

$$\widetilde{\mathbf{V}}^{(\text{MMRM-I})} = \mathbf{L}\ diag\left\{\boldsymbol{e}_K^\top E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1}E^*\left[\pi_j^{-1}\underline{\mathbf{V}}_{\boldsymbol{M}}\underline{\boldsymbol{\Sigma}}_j^{(\text{MMRM-I})}\underline{\mathbf{V}}_{\boldsymbol{M}}\right]E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1}\boldsymbol{e}_K : j=0,\dots,J\right\}\mathbf{L}^\top, \tag{12}$$

$$\widetilde{\mathbf{V}}^{(\text{MMRM-II})} = \mathbf{L}\ diag\left\{\boldsymbol{e}_K^\top E^*\left[\overline{\underline{\mathbf{V}}}_{\boldsymbol{M}}\right]^{-1}E^*\left[\pi_j^{-1}\overline{\underline{\mathbf{V}}}_{\boldsymbol{M}}\underline{\boldsymbol{\Sigma}}_j^{(\text{MMRM-II})}\overline{\underline{\mathbf{V}}}_{\boldsymbol{M}}\right]E^*\left[\overline{\underline{\mathbf{V}}}_{\boldsymbol{M}}\right]^{-1}\boldsymbol{e}_K : j=0,\dots,J\right\}\mathbf{L}^\top, \tag{13}$$

$$\widetilde{\mathbf{V}}^{(\text{IMMRM})} = \mathbf{L}\left(diag\{\boldsymbol{e}_K^\top E^*[\pi_j\underline{\mathbf{V}}_{j\boldsymbol{M}}]^{-1}\boldsymbol{e}_K : j=0,\dots,J\} + \mathbf{r}^\top Var(\boldsymbol{X})\mathbf{r}\right)\mathbf{L}^\top, \tag{14}$$

where

$$\underline{\boldsymbol{\Sigma}}_j^{(\text{ANCOVA})} = \underline{\boldsymbol{\Sigma}}_j^{(\text{MMRM-II})} = E^*\left[\frac{I\{A=j\}}{\pi_j}\{\boldsymbol{Y}-\boldsymbol{h}(A,\boldsymbol{X})\}\{\boldsymbol{Y}-\boldsymbol{h}(A,\boldsymbol{X})\}^\top\right],$$

$$\underline{\boldsymbol{\Sigma}}_j^{(\text{MMRM-I})} = E^*\left[\frac{I\{A=j\}}{\pi_j}(\boldsymbol{Y}-\mathbf{Q}^\top\underline{\boldsymbol{\beta}})(\boldsymbol{Y}-\mathbf{Q}^\top\underline{\boldsymbol{\beta}})^\top\right].$$

We next compute $\mathbf{V}^{(\text{est})}$. Lemma 6 implies that, for $j=0,\dots,J$, $E[Y_K(j)|S] = E[Y_K(j)] + \boldsymbol{b}_{Kj}^\top(E[\boldsymbol{X}|S]-E[\boldsymbol{X}])$. Then using Equation (8), we get

$$E[IF^{(\text{MMRM-I})}(j,\boldsymbol{X}_i,\boldsymbol{Y}_i(j),\boldsymbol{M}_i(j))|S] = \mathbf{L}\widetilde{\boldsymbol{e}}_{j+1}\pi_j^{-1}(\boldsymbol{b}_{Kj}-\underline{\boldsymbol{\beta}}_{\boldsymbol{X}})^\top(E[\boldsymbol{X}|S]-E[\boldsymbol{X}]),$$

where $\widetilde{\boldsymbol{e}}_{j+1}\in\mathbb{R}^{J+1}$ has the $(j+1)$-th entry 1 and the rest 0. Hence by Lemma 3, we have

$$\begin{aligned}\mathbf{V}^{(\text{MMRM-I})} = \widetilde{\mathbf{V}}^{(\text{MMRM-I})} - \mathbf{L}[&diag\{\pi_j^{-1}(\boldsymbol{b}_{Kj}-\underline{\boldsymbol{\beta}}_{\boldsymbol{X}})^\top Var(E[\boldsymbol{X}|S])(\boldsymbol{b}_{Kj}-\underline{\boldsymbol{\beta}}_{\boldsymbol{X}}) : j=0,\dots,J\}\\ &- \mathbf{v}^\top Var(E[\boldsymbol{X}|S])\mathbf{v}]\mathbf{L}^\top,\end{aligned} \tag{15}$$

where $\mathbf{v} = (\boldsymbol{b}_{K0} - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}, \ldots, \boldsymbol{b}_{KJ} - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}})$. similarly, we have $\mathbf{V}^{(\text{IMMRM})} = \widetilde{\mathbf{V}}^{(\text{IMMRM})}$ and

$$\begin{aligned}
\mathbf{V}^{(\text{ANCOVA})} &= \widetilde{\mathbf{V}}^{(\text{ANCOVA})} - \mathbf{L}[diag\{\pi_j^{-1}(\boldsymbol{b}_{Kj} - \boldsymbol{b}_K)^\top Var(E[\boldsymbol{X}|S])(\boldsymbol{b}_{Kj} - \boldsymbol{b}_K) : j = 0, \ldots, J\} \\
&\quad - \mathbf{z}^\top Var(E[\boldsymbol{X}|S])\mathbf{z}]\mathbf{L}^\top, \qquad (16) \\
\mathbf{V}^{(\text{MMRM-II})} &= \widetilde{\mathbf{V}}^{(\text{MMRM-II})} - \mathbf{L}[diag\{\pi_j^{-1}(\boldsymbol{b}_{Kj} - \boldsymbol{b}_K)^\top Var(E[\boldsymbol{X}|S])(\boldsymbol{b}_{Kj} - \boldsymbol{b}_K) : j = 0, \ldots, J\} \\
&\quad - \mathbf{z} Var(E[\boldsymbol{X}|S])\mathbf{z}^\top]\mathbf{L}^\top, \qquad (17)
\end{aligned}$$

where $\boldsymbol{b}_K = Var(\boldsymbol{X})^{-1}Cov^*(\boldsymbol{X}, Y_K)$ and $\mathbf{z} = (\boldsymbol{b}_{K0} - \boldsymbol{b}_K, \ldots, \boldsymbol{b}_{KJ} - \boldsymbol{b}_K)$.

We next show $\mathbf{V}^{(\text{ANCOVA})} \succeq \mathbf{V}^{(\text{IMMRM})}$. By the definition of $\underline{\boldsymbol{\Sigma}}_j^{(\text{ANCOVA})}$ and $\underline{\boldsymbol{\Sigma}}_j^{(\text{IMMRM})}$, we have

$$\underline{\boldsymbol{\Sigma}}_j^{(\text{ANCOVA})} - \underline{\boldsymbol{\Sigma}}_j^{(\text{IMMRM})} = \{Cov(\boldsymbol{Y}(j), \boldsymbol{X}) - Cov^*(\boldsymbol{Y}, \boldsymbol{X})\}Var(\boldsymbol{X})^{-1}\{Cov(\boldsymbol{X}, \boldsymbol{Y}(j)) - Cov^*(\boldsymbol{X}, \boldsymbol{Y})\}$$

is positive semi-definite, and

$$\boldsymbol{e}_K^\top \underline{\boldsymbol{\Sigma}}_j^{(\text{ANCOVA})}\boldsymbol{e}_K = \boldsymbol{e}_K^\top \underline{\boldsymbol{\Sigma}}_j^{(\text{IMMRM})}\boldsymbol{e}_K + (\boldsymbol{b}_{Kj} - \boldsymbol{b}_K)^\top Var(\boldsymbol{X})(\boldsymbol{b}_{Kj} - \boldsymbol{b}_K).$$

Using Equations (11) and (14) and the fact that $Var(\boldsymbol{X}) = E[Var(\boldsymbol{X}|S)] + Var(E[\boldsymbol{X}|S])$, we have

$$\begin{aligned}
&\mathbf{V}^{(\text{ANCOVA})} - \widetilde{\mathbf{V}}^{(\text{IMMRM})} \\
&= \mathbf{V}^{(\text{ANCOVA})} - \widetilde{\mathbf{V}}^{(\text{ANCOVA})} + \widetilde{\mathbf{V}}^{(\text{ANCOVA})} - \widetilde{\mathbf{V}}^{(\text{IMMRM})} \\
&\succeq \mathbf{V}^{(\text{ANCOVA})} - \widetilde{\mathbf{V}}^{(\text{ANCOVA})} + \widetilde{\mathbf{V}}^{(\text{ANCOVA})} \\
&\quad - \mathbf{L}\left(diag\{P^*(M_K = 1)^{-1}\pi_j^{-1}\boldsymbol{e}_K^\top \underline{\boldsymbol{\Sigma}}_j^{(\text{IMMRM})}\boldsymbol{e}_K : j = 0, \ldots, J\} + \mathbf{r}^\top Var(\boldsymbol{X})\mathbf{r}\right)\mathbf{L}^\top \\
&= \mathbf{V}^{(\text{ANCOVA})} - \widetilde{\mathbf{V}}^{(\text{ANCOVA})} \\
&\quad + \frac{1}{P^*(M_K = 1)}\mathbf{L} diag\{\pi_j^{-1}(\boldsymbol{b}_{Kj} - \boldsymbol{b}_K)^\top Var(\boldsymbol{X})(\boldsymbol{b}_{Kj} - \boldsymbol{b}_K) : j = 0, \ldots, J\}\mathbf{L}^\top - \mathbf{L}\mathbf{r}^\top Var(\boldsymbol{X})\mathbf{r}\mathbf{L}^\top \\
&\succeq \mathbf{L}[diag\{\pi_j^{-1}(\boldsymbol{b}_{Kj} - \boldsymbol{b}_K)^\top E[Var(\boldsymbol{X}|S)](\boldsymbol{b}_{Kj} - \boldsymbol{b}_K) : j = 0, \ldots, J\} - \mathbf{z}^\top E[Var(\boldsymbol{X}|S)]\mathbf{z}]\mathbf{L}^\top \\
&= \mathbf{L}\mathbf{U}^\top\{(diag\{\boldsymbol{\pi}\} - \boldsymbol{\pi}\boldsymbol{\pi}^\top) \otimes \mathbf{I}_p\}\mathbf{U}\mathbf{L}^\top, \qquad (18)
\end{aligned}$$

where $\boldsymbol{\pi} = (\pi_0, \ldots, \pi_J)^\top$ and

$$\mathbf{U}^\top = \begin{pmatrix} \pi_0^{-1}(\boldsymbol{b}_{K0} - \boldsymbol{b}_K)^\top E[Var(\boldsymbol{X}|S)]^{\frac{1}{2}} & & \\ & \ddots & \\ & & \pi_J^{-1}(\boldsymbol{b}_{KJ} - \boldsymbol{b}_K)^\top E[Var(\boldsymbol{X}|S)]^{\frac{1}{2}} \end{pmatrix}.$$

In the above derivation, the first "$\succeq$" results from Lemma 1 (2), the second "$\succeq$" comes from $P^*(M_K = 1) \le 1$ and $\mathbf{L}\mathbf{z}^\top = \mathbf{L}\mathbf{r}^\top$. Since $diag\{\boldsymbol{\pi}\} \succeq \boldsymbol{\pi}\boldsymbol{\pi}^\top$, then $(diag\{\boldsymbol{\pi}\} - \boldsymbol{\pi}\boldsymbol{\pi}^\top) \otimes \mathbf{I}_p$ is positive semi-definite (by Lemma 2) and hence $\mathbf{V}^{(\text{ANCOVA})} \succeq \widetilde{\mathbf{V}}^{(\text{IMMRM})}$.

We next show $\mathbf{V}^{(\text{MMRM-I})} \succeq \mathbf{V}^{(\text{IMMRM})}$. Using the definition of $\underline{\boldsymbol{\Sigma}}_j^{(\text{MMRM-I})}$ and $\underline{\boldsymbol{\Sigma}}_j^{(\text{IMMRM})}$, we have $\underline{\boldsymbol{\Sigma}}_j^{(\text{MMRM-I})} - \underline{\boldsymbol{\Sigma}}_j^{(\text{IMMRM})} = \boldsymbol{\Lambda}_j$, where

$$\boldsymbol{\Lambda}_j = \left\{Cov(\boldsymbol{Y}(j), \boldsymbol{X}) - \mathbf{1}_K \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}^\top Var(\boldsymbol{X})\right\} Var(\boldsymbol{X})^{-1} \left\{Cov(\boldsymbol{X}, \boldsymbol{Y}(j)) - Var(\boldsymbol{X})\underline{\boldsymbol{\beta}}_{\boldsymbol{X}} \mathbf{1}_K^\top\right\}$$

is positive semi-definite. By Equations (12) and (14), we have

$$\begin{aligned}
&\widetilde{\mathbf{V}}^{(\text{MMRM-I})} - \widetilde{\mathbf{V}}^{(\text{IMMRM})} \\
&= \mathbf{L}\ diag\left\{\boldsymbol{e}_K^\top E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1} E^*\left[\pi_j^{-1} \underline{\mathbf{V}}_{\boldsymbol{M}} (\underline{\boldsymbol{\Sigma}}_j^{(\text{IMMRM})} + \boldsymbol{\Lambda}_j) \underline{\mathbf{V}}_{\boldsymbol{M}}\right] E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1} \boldsymbol{e}_K : j = 0, \ldots, J\right\} \mathbf{L}^\top \\
&\qquad - \mathbf{L} diag\{\boldsymbol{e}_K^\top E[\pi_j \underline{\mathbf{V}}_{j\boldsymbol{M}}]^{-1} \boldsymbol{e}_K : j = 0, \ldots, J\} \mathbf{L}^\top - \mathbf{L}\mathbf{r}^\top Var(\boldsymbol{X}) \mathbf{r}\mathbf{L}^\top \\
&\succeq \mathbf{L} diag\left\{\boldsymbol{e}_K^\top E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1} E^*\left[\pi_j^{-1} \underline{\mathbf{V}}_{\boldsymbol{M}} \boldsymbol{\Lambda}_j \underline{\mathbf{V}}_{\boldsymbol{M}}\right] E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1} \boldsymbol{e}_K : j = 0, \ldots, J\right\} \mathbf{L}^\top - \mathbf{L}\mathbf{r}^\top Var(\boldsymbol{X})^{-1} \mathbf{r}\mathbf{L}^\top \\
&\succeq \mathbf{L} diag\{\pi_j^{-1} \boldsymbol{e}_K^\top \boldsymbol{\Lambda}_j \boldsymbol{e}_K : j = 0, \ldots, J\} \mathbf{L}^\top - \mathbf{L}\mathbf{r}^\top Var(\boldsymbol{X})^{-1} \mathbf{r}\mathbf{L}^\top \\
&= \mathbf{L} diag\{\pi_j^{-1} (\boldsymbol{b}_{Kj} - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}})^\top Var(\boldsymbol{X}) (\boldsymbol{b}_{Kj} - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}) : j = 0, \ldots, J\} \mathbf{L}^\top - \mathbf{L}\mathbf{v}^\top Var(\boldsymbol{X})^{-1} \mathbf{v}\mathbf{L}^\top,
\end{aligned}$$

where the first "$\succeq$" results from Lemma 1 (3), the second "$\succeq$" results from Lemma 1 (4) and the last equation comes from $\boldsymbol{e}_K^\top \boldsymbol{\Lambda}_j \boldsymbol{e}_K = (\boldsymbol{b}_{Kj} - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}})^\top Var(\boldsymbol{X})(\boldsymbol{b}_{Kj} - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}})$ and

$\mathbf{L}\mathbf{v}^\top = \mathbf{L}\mathbf{r}^\top$. By Equation (15) and $Var(\boldsymbol{X}) = E[Var(\boldsymbol{X}|S)] + Var(E[\boldsymbol{X}|S])$, we have

$$
\begin{aligned}
&\mathbf{V}^{(\text{MMRM-I})} - \widetilde{\mathbf{V}}^{(\text{IMMRM})} \\
&= \mathbf{V}^{(\text{MMRM-I})} - \widetilde{\mathbf{V}}^{(\text{MMRM-I})} + \widetilde{\mathbf{V}}^{(\text{MMRM-I})} - \widetilde{\mathbf{V}}^{(\text{IMMRM})} \\
&\succeq -\mathbf{L}[diag\{\pi_j^{-1}(\boldsymbol{b}_{Kj} - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}})^\top Var(E[\boldsymbol{X}|S])(\boldsymbol{b}_{Kj} - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}) : j = 0, \dots, J\} + \mathbf{v}Var(E[\boldsymbol{X}|S])\mathbf{v}^\top]\mathbf{L}^\top \\
&\quad + \mathbf{L}diag\{\pi_j^{-1}(\boldsymbol{b}_{Kj} - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}})^\top Var(\boldsymbol{X})(\boldsymbol{b}_{Kj} - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}) : j = 0, \dots, J\}\mathbf{L}^\top - \mathbf{L}\mathbf{v}^\top Var(\boldsymbol{X})\mathbf{v}\mathbf{L}^\top \\
&= \mathbf{L}[diag\{\pi_j^{-1}(\boldsymbol{b}_{Kj} - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}})^\top E[Var(\boldsymbol{X}|S)](\boldsymbol{b}_{Kj} - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}) : j = 0, \dots, J\} - \mathbf{v}E[Var(\boldsymbol{X}|S)]\mathbf{v}^\top]\mathbf{L}^\top \\
&= \mathbf{L}\mathbf{Z}^\top\{(diag\{\boldsymbol{\pi}\} - \boldsymbol{\pi}\boldsymbol{\pi}^\top) \otimes \mathbf{I}_p\}\mathbf{Z}\mathbf{L}^\top, \qquad (19)
\end{aligned}
$$

where

$$
\mathbf{Z}^\top = \begin{pmatrix} \pi_0^{-1}(\boldsymbol{b}_{K0} - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}})^\top E[Var(\boldsymbol{X}|S)]^{\frac{1}{2}} & & \\ & \ddots & \\ & & \pi_J^{-1}(\boldsymbol{b}_{KJ} - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}})^\top E[Var(\boldsymbol{X}|S)]^{\frac{1}{2}} \end{pmatrix}.
$$

Since $diag\{\boldsymbol{\pi}\} \succeq \boldsymbol{\pi}\boldsymbol{\pi}^\top$, then we get $\mathbf{V}^{(\text{MMRM-I})} \succeq \widetilde{\mathbf{V}}^{(\text{IMMRM})} = \mathbf{V}^{(\text{IMMRM})}$.

Next, for showing $\mathbf{V}^{(\text{MMRM-II})} \succeq \mathbf{V}^{(\text{IMMRM})}$, we can follow a similar proof as in the previous paragraph, where $\underline{\boldsymbol{\Sigma}}^{(\text{MMRM-I})}$ is substituted by $\underline{\boldsymbol{\Sigma}}^{(\text{MMRM-II})}$, and get

$$
\mathbf{V}^{(\text{MMRM-II})} - \widetilde{\mathbf{V}}^{(\text{IMMRM})} \succeq \mathbf{L}\mathbf{U}^\top\{(diag\{\boldsymbol{\pi}\} - \boldsymbol{\pi}\boldsymbol{\pi}^\top) \otimes \mathbf{I}_p\}\mathbf{U}\mathbf{L}^\top,
$$

which is positive semi-definite.

Finally, we give the necessary and sufficient conditions for $\mathbf{V}^{(\text{est})} = \mathbf{V}^{(\text{IMMRM})}$, $(\text{est}) \in \{\text{ANCOVA}, \text{MMRM-I}, \text{MMRM-II}\}$ in Proposition 1 below. □

Proposition 1. *Assume $K > 1$, Assumption 1 and regularity conditions in the Supplementary Material. For $t = 1, \dots, K$ and $j = 0, \dots, J$, we denote $\boldsymbol{b}_{tj} = Var(\boldsymbol{X})^{-1}Cov\{\boldsymbol{X}, Y_t(j)\}$.*

Then $\mathbf{V}^{(\text{ANCOVA})} = \mathbf{V}^{(\text{IMMRM})}$ if and only if either of the following two sets of conditions (a-b) holds:

(a) for each $j = 0, \ldots, J$, $P(M_K(j) = 1) = 1$ and

$(1 - 2I\{J = 1\}\pi_0)(\boldsymbol{b}_{Kj} - \boldsymbol{b}_{K0})^\top E[Var(\boldsymbol{X}|S)] = \mathbf{0}$;

(b) for each $t = 1, \ldots, K-1$ and $j = 0, \ldots, J$,

$P(M_t(j) = 1, M_K(j) = 0)\ Cov\{Y_t(j) - \boldsymbol{b}_{tj}^\top \boldsymbol{X}, Y_K(j) - \boldsymbol{b}_{Kj}^\top \boldsymbol{X}\} = 0$ *and* $\boldsymbol{b}_{Kj} = \boldsymbol{b}_{K0}$.

In addition, $\mathbf{V}^{(\text{MMRM-I})} = \mathbf{V}^{(\text{IMMRM})}$ if and only if

(a') for $j = 0, \ldots, J$ and $\boldsymbol{m} \in \{0,1\}^K \setminus \{\mathbf{0}_K\}$ with $P^*(\boldsymbol{M} = \boldsymbol{m})$, $\boldsymbol{e}_K^\top\{E[\underline{\mathbf{V}}_{j\boldsymbol{M}}]^{-1}\underline{\mathbf{V}}_{j\boldsymbol{m}} - E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1}\underline{\mathbf{V}}_{\boldsymbol{m}}\} = \mathbf{0}$,

(b') for $j = 0, \ldots, J$ and $\boldsymbol{m} \in \{0,1\}^K \setminus \{\mathbf{0}_K\}$ with $P^*(\boldsymbol{M} = \boldsymbol{m})$, $\boldsymbol{e}_K^\top E[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1}\underline{\mathbf{V}}_{\boldsymbol{m}}\boldsymbol{\Lambda}_j$ is a constant vector,

(c') for $j = 0, \ldots, J$, $\left[\boldsymbol{b}_{Kj} - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}^{(\text{MMRM-I})} - I\{J = 1\}\pi_j(\boldsymbol{b}_{Kj} - \boldsymbol{b}_{K0})\right]^\top E[Var(\boldsymbol{X}|S)] = \mathbf{0}$,

where $\underline{\boldsymbol{\beta}}_{\boldsymbol{X}}^{(\text{MMRM-I})}$ is the probability limit of $\boldsymbol{\beta}_{\boldsymbol{X}}$ in the MMRM-I working model.

Also, $\mathbf{V}^{(\text{MMRM-II})} = \mathbf{V}^{(\text{IMMRM})}$ if and only if

(a") for $j = 0, \ldots, J$ and $\boldsymbol{m} \in \{0,1\}^K \setminus \{\mathbf{0}_K\}$ with $P^*(\boldsymbol{M} = \boldsymbol{m})$, $\boldsymbol{e}_K^\top\{E[\underline{\mathbf{V}}_{j\boldsymbol{M}}]^{-1}\underline{\mathbf{V}}_{j\boldsymbol{m}} - E^*[\underline{\overline{\mathbf{V}}}_{\boldsymbol{M}}]^{-1}\underline{\overline{\mathbf{V}}}_{\boldsymbol{m}}\} = \mathbf{0}$,

(b") for $j = 0, \ldots, J$ and $\boldsymbol{m} \in \{0,1\}^K \setminus \{\mathbf{0}_K\}$ with $P^*(\boldsymbol{M} = \boldsymbol{m})$,

$\boldsymbol{e}_K^\top E[\underline{\overline{\mathbf{V}}}_{\boldsymbol{M}}]^{-1}\underline{\overline{\mathbf{V}}}_{\boldsymbol{m}}\{Cov(\boldsymbol{Y}(j), \boldsymbol{X}) - Cov^*(\boldsymbol{Y}, \boldsymbol{X})\}Var(\boldsymbol{X})^{-1}\{Cov(\boldsymbol{X}, \boldsymbol{Y}(j)) - Cov^*(\boldsymbol{X}, \boldsymbol{Y})\}$ *is a constant vector,*

(c") for $j = 0, \ldots, J$, $(1 - 2I\{J = 1\}\pi_0)(\boldsymbol{b}_{Kj} - \boldsymbol{b}_{K0})^\top E[Var(\boldsymbol{X}|S)] = \mathbf{0}$.

Proof. We first derive the necessary and sufficient conditions for $\mathbf{V}^{(\text{ANCOVA})} = \mathbf{V}^{(\text{IMMRM})}$. Recall the derivation, i.e., Equations (18), for showing $\mathbf{V}^{(\text{ANCOVA})} \succeq \mathbf{V}^{(\text{IMMRM})}$ in the proof of Theorem 1. By check the two inequalities and the last row in Equations (18). We have $\mathbf{V}^{(\mathrm{ANCOVA})} = \mathbf{V}^{(\mathrm{IMMRM})}$ if and only if the following three conditions hold:

(i) $\mathbf{L}\ diag\{\pi_j^{-1}\boldsymbol{e}_K^\top(\frac{1}{P^*(M_K=1)}\underline{\boldsymbol{\Sigma}}_j^{(\mathrm{IMMRM})} - E^*[\underline{\mathbf{V}}_{j\boldsymbol{M}}]^{-1})\boldsymbol{e}_K : j = 0,\ldots,J\}\mathbf{L} = \mathbf{0}$,

(ii) $\{1 - P^*(M_K = 1)\}\mathbf{L}diag\{\pi_j^{-1}(\boldsymbol{b}_{Kj} - \boldsymbol{b}_K)^\top Var(\boldsymbol{X})(\boldsymbol{b}_{Kj} - \boldsymbol{b}_K) : j = 0,\ldots,J\}\mathbf{L}^\top = \mathbf{0}$,

(iii) $\mathbf{L}\mathbf{U}^\top\{(diag\{\boldsymbol{\pi}\} - \boldsymbol{\pi}\boldsymbol{\pi}^\top)\otimes\mathbf{I}_p\}\mathbf{U}\mathbf{L}^\top = \mathbf{0}$.

For Condition (i), Lemma 1 (2) and the assumption that $K > 1$ imply that the equation holds if and only if $P^*(M_K = 0, M_t = 1)\boldsymbol{e}_t^\top\underline{\boldsymbol{\Sigma}}_j^{(\mathrm{IMMRM})}\boldsymbol{e}_K = 0$ for $t = 1,\ldots,K-1$ and $j = 0,\ldots,J$. Equation (10) implies that $\boldsymbol{e}_t^\top\underline{\boldsymbol{\Sigma}}_j^{(\mathrm{IMMRM})}\boldsymbol{e}_K = Cov(Y_t(j) - \boldsymbol{b}_{tj}^\top\boldsymbol{X}, Y_K(j) - \boldsymbol{b}_{Kj}^\top\boldsymbol{X})$. The MCAR assumption implies that $P^*(M_K = 0, M_t = 1) = P(M_K(j) = 0, M_t(j) = 1)$ for $j = 0,\ldots,J$. Hence Condition (i) is equivalent to

(i) $P(M_K(j) = 0, M_t(j) = 1)Cov(Y_t(j) - \boldsymbol{b}_{tj}^\top\boldsymbol{X}, Y_K(j) - \boldsymbol{b}_{Kj}^\top\boldsymbol{X})$ for $t = 1,\ldots,K-1$ and $j = 0,\ldots,J$.

Condition (ii) is equivalent to

(ii) $P^*(M_K = 1)$ or $Cov(\boldsymbol{Y}_K(j) - \boldsymbol{Y}_K(0), \boldsymbol{X}) = \mathbf{0}$.

For Condition (iii), since $\mathbf{L}\mathbf{U}^\top\{(diag\{\boldsymbol{\pi}\} - \boldsymbol{\pi}\boldsymbol{\pi}^\top)\otimes\mathbf{I}_p\}\mathbf{U}\mathbf{L}^\top$ is positive semi-definite, then it is $\mathbf{0}$ if and only if all of its diagonal entries are 0. Denoting $\boldsymbol{u}_j = E[Var(\boldsymbol{X}|S)]^{\frac{1}{2}}(\boldsymbol{b}_{Kj} - \boldsymbol{b}_K)$, we get that the (j,j)-th entry of matrix $\mathbf{L}\mathbf{U}_{S=s}(diag\{\boldsymbol{\pi}\} - \boldsymbol{\pi}\boldsymbol{\pi}^\top)\mathbf{U}_{S=s}\mathbf{L}^\top$ is

$$\begin{aligned}
&\pi_j^{-1}\boldsymbol{u}_j^\top\boldsymbol{u}_j + \pi_0^{-1}\boldsymbol{u}_0^\top\boldsymbol{u}_0 - (\boldsymbol{u}_j - \boldsymbol{u}_0)^\top(\boldsymbol{u}_j - \boldsymbol{u}_0) \\
&= \frac{1}{\pi_0\pi_j}\left(\pi_0\boldsymbol{u}_j + \pi_j\boldsymbol{u}_0\right)^\top\left(\pi_0\boldsymbol{u}_j + \pi_j\boldsymbol{u}_0\right) + (1 - \pi_0 - \pi_j)\left(\frac{1}{\pi_j}\boldsymbol{u}_0^\top\boldsymbol{u}_0 + \frac{1}{\pi_0}\boldsymbol{u}_j^\top\boldsymbol{u}_j\right),
\end{aligned}$$

which is equal to 0 if and only if either $\boldsymbol{u}_0 = \boldsymbol{u}_j = \mathbf{0}$, or $\pi_0 + \pi_j = 1$ and $\pi_0 \boldsymbol{u}_j + \pi_j \boldsymbol{u}_0 = \mathbf{0}$. The former case is equivalent to $E[Var(\boldsymbol{X}|S)]^{\frac{1}{2}}(\boldsymbol{b}_{Kj} - \boldsymbol{b}_{K0}) = \mathbf{0}$ for $j = 1, \ldots, K$; and the later case is equivalent to $J = 1$ and $(\pi_1 - \pi_0)E[Var(\boldsymbol{X}|S)]^{\frac{1}{2}}(\boldsymbol{b}_{K1} - \boldsymbol{b}_K)$. Hence Condition (iii) is equivalent to

(iii) $(I\{J = 1\}(\pi_j - \pi_0) + I\{J > 1\})E[Var(\boldsymbol{X}|S)](\boldsymbol{b}_{Kj} - \boldsymbol{b}_K) = \mathbf{0}$ for $j = 1, \ldots, J$.

Combining Conditions (i-iii) together, we observe that, $P^*(M_K = 1) = 1$ in Condition (ii) implies Condition (i), and $Cov(\boldsymbol{Y}_K(j) - \boldsymbol{Y}_K(0), \boldsymbol{X}) = \mathbf{0}$ in Condition (ii) implies Condition (iii). As a result, the three conditions can be summarized into two conditions, which are

(a) for each $j = 1, \ldots, J$, $P(M_K(j) = 1) = 1$ and
$(1 - 2I\{J = 1\}\pi_0)(\boldsymbol{b}_{Kj} - \boldsymbol{b}_{K0})^\top E[Var(\boldsymbol{X}|S)] = \mathbf{0}$;

(b) for each $t = 1, \ldots, K-1$ and $j = 0, \ldots, J$,
$P(M_t(j) = 1, M_K(j) = 0)\ Cov\{Y_t(j) - \boldsymbol{b}_{tj}^\top \boldsymbol{X}, Y_K(j) - \boldsymbol{b}_{Kj}^\top \boldsymbol{X}\} = 0$ and $\boldsymbol{b}_{Kj} = \boldsymbol{b}_{K0}$.

For the MMRM-I estimator, the derivations, i.e., Equations (19), imply that $\mathbf{V}^{(\text{MMRM-I})} = \mathbf{V}^{(\text{IMMRM})}$ if and only if the following conditions hold:

(i') for $j = 0, \ldots, J$ and $\boldsymbol{m} \in \{0,1\}^K \setminus \{\mathbf{0}_K\}$ with $P^*(\boldsymbol{M} = \boldsymbol{m})$, $\boldsymbol{e}_K^\top\{E[\underline{\mathbf{V}}_{j\boldsymbol{M}}]^{-1}\underline{\mathbf{V}}_{j\boldsymbol{m}} - E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1}\underline{\mathbf{V}}_{\boldsymbol{m}}\} = \mathbf{0}$,

(ii') for $j = 0, \ldots, J$ and $\boldsymbol{m} \in \{0,1\}^K \setminus \{\mathbf{0}_K\}$ with $P^*(\boldsymbol{M} = \boldsymbol{m})$, $\boldsymbol{e}_K^\top E[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1}\underline{\mathbf{V}}_{\boldsymbol{m}}\boldsymbol{\Lambda}_j$ is a constant vector,

(iii') $\mathbf{L}\mathbf{Z}^\top\{(diag\{\boldsymbol{\pi}\} - \boldsymbol{\pi}\boldsymbol{\pi}^\top) \otimes \mathbf{I}_p\}\mathbf{Z}\mathbf{L}^\top = \mathbf{0}$.

Similar to the analysis for Condition (iii), Condition (iii') is equivalent to (iii') for $j=1,\ldots,J$, $[\boldsymbol{b}_{Kj}-\underline{\boldsymbol{\beta}}_{\boldsymbol{X}}^{(\text{MMRM-I})}-I\{J=1\}\pi_j(\boldsymbol{b}_{Kj}-\boldsymbol{b}_{K0})]^\top E[Var(\boldsymbol{X}|S)]=\mathbf{0}$,

which is the necessary condition given in Corollary 1.

For the MMRM-II estimator, similarly, we have $\mathbf{V}^{(\text{MMRM-I})}=\mathbf{V}^{(\text{IMMRM})}$ if and only if the following conditions hold:

(i") for $j=0,\ldots,J$ and $\boldsymbol{m}\in\{0,1\}^K\setminus\{\mathbf{0}_K\}$ with $P^*(\boldsymbol{M}=\boldsymbol{m})$, $\boldsymbol{e}_K^\top\{E[\underline{\mathbf{V}}_{j\boldsymbol{M}}]^{-1}\underline{\mathbf{V}}_{j\boldsymbol{m}}-E^*[\overline{\underline{\mathbf{V}}}_{\boldsymbol{M}}]^{-1}\overline{\underline{\mathbf{V}}}_{\boldsymbol{m}}\}=\mathbf{0}$,

(ii") for $j=0,\ldots,J$ and $\boldsymbol{m}\in\{0,1\}^K\setminus\{\mathbf{0}_K\}$ with $P^*(\boldsymbol{M}=\boldsymbol{m})$, $\boldsymbol{e}_K^\top E[\overline{\underline{\mathbf{V}}}_{\boldsymbol{M}}]^{-1}\overline{\underline{\mathbf{V}}}_{\boldsymbol{m}}\{Cov(\boldsymbol{Y}(j),\boldsymbol{X})-Cov^*(\boldsymbol{Y},\boldsymbol{X})\}Var(\boldsymbol{X})^{-1}\{Cov(\boldsymbol{X},\boldsymbol{Y}(j))-Cov^*(\boldsymbol{X},\boldsymbol{Y})\}$ is a constant vector,

(iii") $\mathbf{L}\mathbf{U}^\top\{(diag\{\boldsymbol{\pi}\}-\boldsymbol{\pi}\boldsymbol{\pi}^\top)\otimes\mathbf{I}_p\}\mathbf{U}\mathbf{L}^\top=\mathbf{0}$.

The Condition (iii") is the same as Condition (iii), which is the necessary condition shown in Corollary 1. □

Lemma 5. *Assume the same assumption as in Theorem 1. Then the influence functions of* $\widehat{\boldsymbol{\Delta}}^{(\text{ANCOVA})}$, $\widehat{\boldsymbol{\Delta}}^{(\text{MMRM-I})}$ *and* $\widehat{\boldsymbol{\Delta}}^{(\text{MMRM-II})}$ *are given by Equations (7), (8) and (9), respectively. Under simple randomization, the asymptotic covariance matrix of* $\widehat{\boldsymbol{\Delta}}^{(\text{ANCOVA})}$, $\widehat{\boldsymbol{\Delta}}^{(\text{MMRM-I})}$ *and* $\widehat{\boldsymbol{\Delta}}^{(\text{MMRM-II})}$ *are given by Equations (11), (12) and (13), respectively.*

Proof. We first derive the influence function for the MMRM-I estimator. Theorem 1 of Wang et al. (2023) implies that $IF^{(\text{MMRM-I})}(A,\boldsymbol{X},\boldsymbol{Y},\boldsymbol{M};\boldsymbol{\theta})=\mathbf{B}^{-1}\boldsymbol{\psi}^{(\text{MMRM})}(A,\boldsymbol{X},\boldsymbol{Y},\boldsymbol{M};\boldsymbol{\theta})$, where $\mathbf{B}=E^*\left[\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\psi}^{(\text{MMRM})}(A,\boldsymbol{X},\boldsymbol{Y},\boldsymbol{M};\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\underline{\boldsymbol{\theta}}}\right]$. Using the formula (2) of $\boldsymbol{\psi}^{(\text{MMRM})}$, we can show that

$$\mathbf{B} = \begin{bmatrix} -E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}] & -E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}(\mathbf{I}_K \otimes \boldsymbol{A})^\top] & -E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}\mathbf{u}(\boldsymbol{X})^\top] & \mathbf{0} \\ -E^*[(\mathbf{I}_K \otimes \boldsymbol{A})\underline{\mathbf{V}}_{\boldsymbol{M}}] & -E^*[(\mathbf{I}_K \otimes \boldsymbol{A})\underline{\mathbf{V}}_{\boldsymbol{M}}(\mathbf{I}_K \otimes \boldsymbol{A})^\top] & -E^*[(\mathbf{I}_K \otimes \boldsymbol{A})\underline{\mathbf{V}}_{\boldsymbol{M}}\mathbf{u}(\boldsymbol{X})^\top] & \mathbf{0} \\ -E^*[\mathbf{u}(\boldsymbol{X})\underline{\mathbf{V}}_{\boldsymbol{M}}] & -E^*[\mathbf{u}(\boldsymbol{X})\underline{\mathbf{V}}_{\boldsymbol{M}}(\mathbf{I}_K \otimes \boldsymbol{A})^\top] & -E^*[\mathbf{u}(\boldsymbol{X})\underline{\mathbf{V}}_{\boldsymbol{M}}\mathbf{u}(\boldsymbol{X})^\top] & \mathbf{B}_{34} \\ \mathbf{0} & \mathbf{0} & \mathbf{B}_{34}^\top & \mathbf{B}_{44} \end{bmatrix},$$

where $\mathbf{B}_{34} \in \mathbb{R}^{q\times r}$ and $\mathbf{B}_{44} \in \mathbb{R}^{r\times r}$ are matrices not related to the influence function of $\widehat{\boldsymbol{\Delta}}^{(\text{MMRM-I})}$. The zeros in the above matrix result from the following derivation:

$$\begin{aligned} E^*\left[\underline{\mathbf{V}}_{\boldsymbol{M}}\frac{\partial \boldsymbol{\Sigma}}{\partial \alpha_j}\underline{\mathbf{V}}_{\boldsymbol{M}}(\boldsymbol{Y} - \mathbf{Q}^\top\underline{\boldsymbol{\beta}})\right] &= E^*\left[\underline{\mathbf{V}}_{\boldsymbol{M}}\frac{\partial \boldsymbol{\Sigma}}{\partial \alpha_j}\underline{\mathbf{V}}_{\boldsymbol{M}}\right] E^*[\boldsymbol{Y} - \underline{\boldsymbol{\beta}}_0 - (\mathbf{I}_K \otimes \boldsymbol{A})^\top\underline{\boldsymbol{\beta}}_{\boldsymbol{A}} - \mathbf{u}(\boldsymbol{X})^\top\underline{\boldsymbol{\beta}}_{\mathbf{u}(\boldsymbol{X})}] \\ &= \mathbf{0}, \end{aligned}$$

and, similarly, $E^*\left[(\mathbf{I}_K \otimes \boldsymbol{A})\underline{\mathbf{V}}_{\boldsymbol{M}}\frac{\partial \boldsymbol{\Sigma}}{\partial \alpha_j}\underline{\mathbf{V}}_{\boldsymbol{M}}(\boldsymbol{Y} - \mathbf{Q}^\top\underline{\boldsymbol{\beta}})\right] = \mathbf{0}$. By the regularity conditions, $\mathbf{B}$ is invertible. To compute $\mathbf{B}^{-1}$, we define

$$\mathbf{D} = \begin{pmatrix} \mathbf{I}_K & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -E^*[\mathbf{I}_K \otimes \boldsymbol{A}] & \mathbf{I}_K & \mathbf{0} & \mathbf{0} \\ -E^*[\mathbf{u}(\boldsymbol{X})] & \mathbf{0} & \mathbf{I}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_r \end{pmatrix},$$

and $\mathbf{F} = \mathbf{D}\mathbf{B}\mathbf{D}^\top$. Since $\mathbf{D}$ is a lower triangular matrix and hence invertible, then $\mathbf{F}$ is invertible. Since MCAR implies that $E^*[\mathbf{u}(\boldsymbol{X})\mathbf{V}_{\boldsymbol{M}}] = E^*[\mathbf{u}(\boldsymbol{X})]E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]$ and $E^*[(\mathbf{I}_K \otimes \boldsymbol{A})\underline{\mathbf{V}}_{\boldsymbol{M}}] = E^*[\mathbf{I}_K \otimes \boldsymbol{A}]E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]$, then

$$\mathbf{F} = \begin{pmatrix} -E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}] & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}] \otimes Var^*(\boldsymbol{A}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -Var\{\mathbf{u}(\boldsymbol{X})E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{\frac{1}{2}}\} & \mathbf{B}_{34} \\ \mathbf{0} & \mathbf{0} & \mathbf{B}_{34}^\top & \mathbf{B}_{44} \end{pmatrix},$$

where $Var^*(\boldsymbol{A}) = E^*[\boldsymbol{A}\boldsymbol{A}^\top] - E^*[\boldsymbol{A}]E^*[\boldsymbol{A}]^\top$, $\mathbf{F}_{33} \in \mathbb{R}^{q\times q}$ is a matrix not related to the influence function of $\widehat{\boldsymbol{\beta}}_A$. Then

$$
\begin{aligned}
\mathbf{B}^{-1} &= \mathbf{D}^\top \mathbf{F}^{-1}\mathbf{D} \\
&= \mathbf{D}^\top \begin{bmatrix} -E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1} \otimes Var^*(\boldsymbol{A})^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \tilde{\mathbf{B}}_{33} & \tilde{\mathbf{B}}_{34} \\ \mathbf{0} & \mathbf{0} & \tilde{\mathbf{B}}_{34}^\top & \tilde{\mathbf{B}}_{44} \end{bmatrix} \mathbf{D} \\
&= \begin{bmatrix} \tilde{\mathbf{B}}_{11} & \mathbf{C}^\top & -E^*[\mathbf{u}(\boldsymbol{X})^\top]\tilde{\mathbf{B}}_{33} & -E^*[\mathbf{u}(\boldsymbol{X})^\top]\tilde{\mathbf{B}}_{34} \\ \mathbf{C} & -E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1} \otimes Var^*(\boldsymbol{A})^{-1} & \mathbf{0} & \mathbf{0} \\ -\tilde{\mathbf{B}}_{33}E^*[\mathbf{u}(\boldsymbol{X})] & \mathbf{0} & \tilde{\mathbf{B}}_{33} & \tilde{\mathbf{B}}_{34} \\ -\tilde{\mathbf{B}}_{34}^\top E^*[\mathbf{u}(\boldsymbol{X})] & \mathbf{0} & \tilde{\mathbf{B}}_{34}^\top & \tilde{\mathbf{B}}_{44} \end{bmatrix},
\end{aligned}
$$

where $\mathbf{C} = \{E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1} \otimes Var^*(\boldsymbol{A})^{-1}\}E^*[\mathbf{I}_K \otimes \boldsymbol{A}]$ and $\tilde{\mathbf{B}}_{11} \in \mathbb{R}^{K\times K}$, $\tilde{\mathbf{B}}_{33} \in \mathbb{R}^{q\times q}$, $\tilde{\mathbf{B}}_{34} \in \mathbb{R}^{q\times r}$ and $\tilde{\mathbf{B}}_{44} \in \mathbb{R}^{r\times r}$ are matrices that are not related to the influence function of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{A}}$ (as shown below). Since $\boldsymbol{\beta}_{\boldsymbol{A}}$ are the $(K+1)$-th, $\ldots$, $((J+1)K)$-th entries in $\boldsymbol{\theta}$, we need the $(K+1)$-th, $\ldots$, $((J+1)K)$-th rows of $\mathbf{B}^{-1}$ to derive the influence function for $\widehat{\boldsymbol{\beta}}_{\boldsymbol{A}}$, which are

$$
\begin{bmatrix} \mathbf{C} & -\{E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1} \otimes Var^*(\boldsymbol{A})^{-1}\} & \mathbf{0} & \mathbf{0} \end{bmatrix}.
$$

Then the influence function for $\widehat{\boldsymbol{\beta}}_{\boldsymbol{A}}$ is

$$
\{E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1} \otimes Var^*(\boldsymbol{A})^{-1}\}\{\mathbf{I}_K \otimes (\boldsymbol{A} - E^*[\boldsymbol{A}])\}\underline{\mathbf{V}}_{\boldsymbol{M}}(\boldsymbol{Y} - \mathbf{Q}^\top \underline{\boldsymbol{\beta}}),
$$

which implies that the influence function for $\widehat{\boldsymbol{\Delta}}^{(\text{MMRM-I})}$ is

$$
IF^{(\text{MMRM-I})} = Var^*(\boldsymbol{A})^{-1}(\boldsymbol{A} - E^*[\boldsymbol{A}])\boldsymbol{e}_K^\top E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1}\underline{\mathbf{V}}_{\boldsymbol{M}}(\boldsymbol{Y} - \mathbf{Q}^\top \underline{\boldsymbol{\beta}}).
$$

Since $Var^*(\boldsymbol{A})^{-1}(\boldsymbol{A} - E^*[\boldsymbol{A}]) = \mathbf{L}(\frac{I\{A=0\}}{\pi_0}, \ldots, \frac{I\{A=J\}}{\pi_J})^\top$, we get the desired formula of $IF^{(\text{MMRM-I})}$.

We next compute $\boldsymbol{Y} - \mathbf{Q}^\top \underline{\boldsymbol{\beta}}$. By $E^*[\boldsymbol{\psi}^{(\text{MMRM})}(A, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{M}; \boldsymbol{\theta})] = \mathbf{0}$, we have $\underline{\boldsymbol{\beta}} = E^*[\mathbf{Q}\underline{\mathbf{V}}_{\boldsymbol{M}}\mathbf{Q}]^{-1}E^*[\mathbf{Q}\underline{\mathbf{V}}_{\boldsymbol{M}}\boldsymbol{Y}]$. Recalling $\mathbf{u}(\boldsymbol{X}) = \boldsymbol{X}\mathbf{1}_K^\top$ for the MMRM-I estimator and following a similar procedure for calculating $\mathbf{B}^{-1}$, we have

$$\underline{\boldsymbol{\beta}} = \begin{pmatrix} \mathbf{I}_K & \mathbf{0} & \mathbf{0} \\ -E^*[\mathbf{I}_K \otimes \boldsymbol{A}] & \mathbf{I}_K & \mathbf{0} \\ -E^*[\mathbf{u}(\boldsymbol{X})] & \mathbf{0} & \mathbf{I}_q \end{pmatrix}^\top \begin{pmatrix} E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1} \otimes Var^*(\boldsymbol{A})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & Var\{\mathbf{u}(\boldsymbol{X})E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{\frac{1}{2}}\}^{-1} \end{pmatrix}$$
$$\begin{pmatrix} \mathbf{I}_K & \mathbf{0} & \mathbf{0} \\ -E^*[\mathbf{I}_K \otimes \boldsymbol{A}] & \mathbf{I}_K & \mathbf{0} \\ -E^*[\mathbf{u}(\boldsymbol{X})] & \mathbf{0} & \mathbf{I}_q \end{pmatrix} \begin{pmatrix} E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}\boldsymbol{Y}] \\ E^*[(\underline{\mathbf{V}}_{\boldsymbol{M}}\boldsymbol{Y}) \otimes \boldsymbol{A}] \\ E^*[\mathbf{u}(\boldsymbol{X})\underline{\mathbf{V}}_{\boldsymbol{M}}\boldsymbol{Y}] \end{pmatrix},$$

which implies $\underline{\boldsymbol{\beta}}_{\boldsymbol{A}} = E^*[\widetilde{\boldsymbol{Y}} \otimes \widetilde{\boldsymbol{A}}]$ and $\underline{\boldsymbol{\beta}}_{\boldsymbol{X}}^\top = \frac{\mathbf{1}_K^\top E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]}{\mathbf{1}_K^\top E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]\mathbf{1}_K} Cov^*(\boldsymbol{Y}, \boldsymbol{X})Var(\boldsymbol{X})^{-1}$. Since $\boldsymbol{\beta}_0$ satisfies $E^*[\boldsymbol{Y} - \underline{\boldsymbol{\beta}}_0 - (\mathbf{I}_K \otimes \boldsymbol{A})^\top \underline{\boldsymbol{\beta}}_{\boldsymbol{A}} - \mathbf{u}(\boldsymbol{X})^\top \underline{\boldsymbol{\beta}}_{\mathbf{u}(\boldsymbol{X})}] = \mathbf{0}$, we get $\boldsymbol{Y} - \mathbf{Q}^\top \underline{\boldsymbol{\beta}} = \widetilde{\boldsymbol{Y}} - \underline{\boldsymbol{\beta}}_{\boldsymbol{A}}^\top (\mathbf{I}_K \otimes \widetilde{\boldsymbol{A}}) - \mathbf{1}_K \beta_{\boldsymbol{X}}^\top \widetilde{\boldsymbol{X}}$. Then direct calculation gives the desired formula of $\boldsymbol{Y} - \mathbf{Q}^\top \underline{\boldsymbol{\beta}}$.

We next calculate $\widetilde{\mathbf{V}}^{(\text{MMRM-I})}$. Since $\boldsymbol{\Sigma}$ is unstructured, the second set of estimating equations $\boldsymbol{\psi}^{(\text{MMRM})}$ implies that, for each $r, s = 1, \ldots, K$ and $j = 0, \ldots, J$, we have

$$\begin{aligned} 0 &= E^*\left[-\text{tr}(\underline{\mathbf{V}}_{\boldsymbol{M}}(\boldsymbol{e}_r\boldsymbol{e}_s^\top + \boldsymbol{e}_s\boldsymbol{e}_r^\top)) + (\boldsymbol{Y} - \mathbf{Q}^\top \underline{\boldsymbol{\beta}})^\top \underline{\mathbf{V}}_{\boldsymbol{M}}(\boldsymbol{e}_r\boldsymbol{e}_s^\top + \boldsymbol{e}_s\boldsymbol{e}_r^\top))\underline{\mathbf{V}}_{\boldsymbol{M}}(\boldsymbol{Y} - \mathbf{Q}^\top \underline{\boldsymbol{\beta}})\right] \\ &= E^*\left[2\boldsymbol{e}_r^\top\{-\underline{\mathbf{V}}_{\boldsymbol{M}} + \underline{\mathbf{V}}_{\boldsymbol{M}}(\boldsymbol{Y} - \mathbf{Q}^\top \underline{\boldsymbol{\beta}})(\boldsymbol{Y} - \mathbf{Q}^\top \underline{\boldsymbol{\beta}})^\top \underline{\mathbf{V}}_{\boldsymbol{M}}\}\boldsymbol{e}_s\right], \end{aligned}$$

which implies that, for $j = 0, \ldots, J$, $E^*[-\underline{\mathbf{V}}_{\boldsymbol{M}}] + E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}(\boldsymbol{Y} - \mathbf{Q}^\top \underline{\boldsymbol{\beta}})(\boldsymbol{Y} - \mathbf{Q}^\top \underline{\boldsymbol{\beta}})^\top \underline{\mathbf{V}}_{\boldsymbol{M}}] = \mathbf{0}$. Since $E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}] = E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}\underline{\boldsymbol{\Sigma}}^{(\text{MMRM-I})}\underline{\mathbf{V}}_{\boldsymbol{M}}]$, the regularity condition (3) implies that

$\underline{\boldsymbol{\Sigma}}^{(\text{MMRM-I})} = E^*[(\boldsymbol{Y} - \mathbf{Q}^\top \underline{\boldsymbol{\beta}})(\boldsymbol{Y} - \mathbf{Q}^\top \underline{\boldsymbol{\beta}})^\top]$. Thus,

$$
\begin{aligned}
&\widetilde{\mathbf{V}}^{(\text{MMRM-I})} \\
&= E[IF^{(\text{MMRM-I})} IF^{(\text{MMRM-I})\top}] \\
&= \mathbf{L}\ diag\Bigg\{ \boldsymbol{e}_K^\top E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1} E^* \left[\frac{I\{A=j\}}{\pi_j^2} \underline{\mathbf{V}}_{\boldsymbol{M}} (\boldsymbol{Y} - \mathbf{Q}^\top \underline{\boldsymbol{\beta}})(\boldsymbol{Y} - \mathbf{Q}^\top \underline{\boldsymbol{\beta}})^\top \underline{\mathbf{V}}_{\boldsymbol{M}} \right] \\
&\qquad E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1} \boldsymbol{e}_K : j = 0, \dots, J \Bigg\} \mathbf{L}^\top \\
&= \mathbf{L}\ diag\Bigg\{ \boldsymbol{e}_K^\top E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1} E^* \left[\pi_j^{-1} \underline{\mathbf{V}}_{\boldsymbol{M}} \underline{\boldsymbol{\Sigma}}_j^{(\text{MMRM-I})} \underline{\mathbf{V}}_{\boldsymbol{M}} \right] E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1} \boldsymbol{e}_K : j = 0, \dots, J \Bigg\} \mathbf{L}^\top.
\end{aligned}
$$

For the MMRM-II estimator, we can follow a similar proof for the MMRM-I estimator and get the desired influence function and asymptotic covariance matrix. The only difference comes from $\mathbf{u}(\boldsymbol{X}) = \mathbf{I}_K \otimes \boldsymbol{X}$.

For the ANCOVA estimator, we observed that it is a special case of the MMRM-I estimator setting $K = 1$. Then we have $\underline{\mathbf{V}}_{\boldsymbol{M}} = I\{M_K = 1\}(\boldsymbol{e}_K^\top \underline{\boldsymbol{\Sigma}}^{(\text{ANCOVA})} \boldsymbol{e}_K)^{-1}$, which naturally implies the desired influence function and asymptotic covariance matrix. $\square$

Lemma 6. *Assume the same assumption as in Theorem 1 and assume $\boldsymbol{\Sigma}_j, j = 0, \dots, J$ are unstructured. Then the influence function of $\widehat{\boldsymbol{\Delta}}^{(\text{IMMRM})}$ is given by Equation (10) and the asymptotic covariance matrix of $\widehat{\boldsymbol{\Delta}}^{(\text{IMMRM})}$ is given by Equation (14).*

Proof. Following a similar procedure as in Lemma 5, we get that the influence function for $\widehat{\boldsymbol{\Delta}}^{(\text{IMMRM})}$ is $IF^{(\text{IMMRM})} = \boldsymbol{U}_1 + \boldsymbol{U}_2$, where

$$
\begin{aligned}
\boldsymbol{U}_1 &= (\boldsymbol{e}_K^\top \otimes \mathbf{I}_J) \mathbf{H}^{-1} (\mathbf{I}_K \otimes \boldsymbol{A} - E^*[\mathbf{V}_{\boldsymbol{AM}} \otimes \boldsymbol{A}] E^*[\mathbf{V}_{\boldsymbol{AM}}]^{-1}) \mathbf{V}_{\boldsymbol{AM}} (\boldsymbol{Y} - \mathbf{R}^\top \underline{\boldsymbol{\gamma}}) \\
\boldsymbol{U}_2 &= (\boldsymbol{e}_K \otimes \mathbf{I}_J)^\top \underline{\boldsymbol{\beta}}_{\boldsymbol{A}} + (\boldsymbol{e}_K \otimes \boldsymbol{X} \otimes \mathbf{I}_J)^\top \underline{\boldsymbol{\beta}}_{\boldsymbol{AX}} - \boldsymbol{\Delta}^*.
\end{aligned}
$$

where $\mathbf{H} = E^*[\mathbf{V}_{\boldsymbol{AM}} \otimes \boldsymbol{A}\boldsymbol{A}^\top] - E^*[\mathbf{V}_{\boldsymbol{AM}} \otimes \boldsymbol{A}] E^*[\mathbf{V}_{\boldsymbol{AM}}]^{-1} E^*[\mathbf{V}_{\boldsymbol{AM}} \otimes \boldsymbol{A}^\top]$.

We next compute $\boldsymbol{U}_1$. Define $\boldsymbol{\delta}_j \in \mathbb{R}^J$ has the J-th entry 1 and the rest 0. We have

$$E^*[\mathbf{V}_{A\boldsymbol{M}} \otimes \boldsymbol{A}\boldsymbol{A}^\top]^{-1} = \left(\sum_{j=1}^{J} E[\pi_j \underline{\mathbf{V}}_{j\boldsymbol{M}}] \otimes \boldsymbol{\delta}_j \boldsymbol{\delta}_j^\top \right)^{-1} = \sum_{j=1}^{J} E[\pi_j \underline{\mathbf{V}}_{j\boldsymbol{M}}]^{-1} \otimes \boldsymbol{\delta}_j \boldsymbol{\delta}_j^\top$$

and hence

$$\begin{aligned}
&- E^*[\mathbf{V}_{A\boldsymbol{M}}] + E^*[\mathbf{V}_{A\boldsymbol{M}} \otimes \boldsymbol{A}^\top] E^*[\mathbf{V}_{A\boldsymbol{M}} \otimes \boldsymbol{A}\boldsymbol{A}^\top]^{-1} E^*[\mathbf{V}_{A\boldsymbol{M}} \otimes \boldsymbol{A}] \\
&= -E[\mathbf{V}_{A\boldsymbol{M}}] + \left(\sum_{j'=1}^{J} E[\pi_{j'} \underline{\mathbf{V}}_{j'\boldsymbol{M}}] \otimes \boldsymbol{\delta}_{j'}^\top \right) \left(\sum_{j=1}^{J} E[\pi_j \underline{\mathbf{V}}_{j\boldsymbol{M}}]^{-1} \otimes \boldsymbol{\delta}_j \boldsymbol{\delta}_j^\top \right) \left(\sum_{j''=1}^{J} E[\pi_{j''} \underline{\mathbf{V}}_{j''\boldsymbol{M}}] \otimes \boldsymbol{\delta}_{j''} \right) \\
&= -E^*[\mathbf{V}_{A\boldsymbol{M}}] + \sum_{j=1}^{J} E[\pi_j \underline{\mathbf{V}}_{j\boldsymbol{M}}] \otimes \boldsymbol{\delta}_j^\top \boldsymbol{\delta}_j \\
&= -E[\pi_0 \underline{\mathbf{V}}_{0\boldsymbol{M}}].
\end{aligned}$$

Using the Woodbury matrix identity, we get

$$\begin{aligned}
\mathbf{H}^{-1} &= E^*[\mathbf{V}_{A\boldsymbol{M}} \otimes \boldsymbol{A}\boldsymbol{A}^\top]^{-1} - E^*[\mathbf{V}_{A\boldsymbol{M}} \otimes \boldsymbol{A}\boldsymbol{A}^\top]^{-1} E^*[\mathbf{V}_{A\boldsymbol{M}} \otimes \boldsymbol{A}] \\
&\qquad (-E^*[\mathbf{V}_{A\boldsymbol{M}}] + E^*[\mathbf{V}_{A\boldsymbol{M}} \otimes \boldsymbol{A}^\top] E^*[\mathbf{V}_{A\boldsymbol{M}} \otimes \boldsymbol{A}\boldsymbol{A}^\top]^{-1} E^*[\mathbf{V}_{A\boldsymbol{M}} \otimes \boldsymbol{A}])^{-1} \\
&\qquad\quad E^*[\mathbf{V}_{A\boldsymbol{M}} \otimes \boldsymbol{A}^\top] E^*[\mathbf{V}_{A\boldsymbol{M}} \otimes \boldsymbol{A}\boldsymbol{A}^\top]^{-1} \\
&= \sum_{j=1}^{J} E[\pi_j \underline{\mathbf{V}}_{j\boldsymbol{M}}]^{-1} \otimes \boldsymbol{\delta}_j \boldsymbol{\delta}_j^\top - (\mathbf{I}_K \otimes \mathbf{1}_J)(-E[\pi_0 \underline{\mathbf{V}}_{0\boldsymbol{M}}])^{-1} (\mathbf{I}_K \otimes \mathbf{1}_J^\top) \\
&= \sum_{j=1}^{J} E[\pi_j \underline{\mathbf{V}}_{j\boldsymbol{M}}]^{-1} \otimes \boldsymbol{\delta}_j \boldsymbol{\delta}_j^\top + E[\pi_0 \underline{\mathbf{V}}_{0\boldsymbol{M}}]^{-1} \otimes \mathbf{1}_J \mathbf{1}_J^\top
\end{aligned}$$

Then we get

$$
\begin{aligned}
&\mathbf{H}^{-1}(\mathbf{I}_K \otimes \boldsymbol{A} - E^*[\mathbf{V}_{A\boldsymbol{M}} \otimes \boldsymbol{A}]E^*[\mathbf{V}_{A\boldsymbol{M}}]^{-1}) \\
&= \left(\sum_{j=1}^J E[\pi_j \underline{\mathbf{V}}_{j\boldsymbol{M}}]^{-1} \otimes \boldsymbol{\delta}_j\boldsymbol{\delta}_j^\top + E[\pi_0\underline{\mathbf{V}}_{0\boldsymbol{M}}]^{-1} \otimes \mathbf{1}_J\mathbf{1}_J^\top\right)\left(\mathbf{I}_K \otimes \boldsymbol{A} - \sum_{j=1}^J E[\pi_j\underline{\mathbf{V}}_{j\boldsymbol{M}}]E^*[\mathbf{V}_{A\boldsymbol{M}}]^{-1} \otimes \boldsymbol{\delta}_j\right) \\
&= \sum_{j=1}^J E[\pi_j \underline{\mathbf{V}}_{j\boldsymbol{M}}]^{-1} \otimes \boldsymbol{\delta}_j I\{A=j\} + E[\pi_0\underline{\mathbf{V}}_{0\boldsymbol{M}}]^{-1} \otimes \mathbf{1}_J(1 - I\{A=0\}) - \sum_{j=1}^J E^*[\mathbf{V}_{A\boldsymbol{M}}]^{-1} \otimes \boldsymbol{\delta}_j \\
&\quad - E[\pi_0\underline{\mathbf{V}}_{0\boldsymbol{M}}]^{-1}(E^*[\mathbf{V}_{A\boldsymbol{M}}] - E[\pi_0\underline{\mathbf{V}}_{0\boldsymbol{M}}])E^*[\mathbf{V}_{A\boldsymbol{M}}]^{-1} \otimes \mathbf{1}_J \\
&= \sum_{j=1}^J E[\pi_j \underline{\mathbf{V}}_{j\boldsymbol{M}}]^{-1} \otimes \boldsymbol{\delta}_j I\{A=j\} - E[\pi_0\underline{\mathbf{V}}_{0\boldsymbol{M}}]^{-1} \otimes \mathbf{1}_J I\{A=0\},
\end{aligned}
$$

which implies that $\boldsymbol{U}_1 = \mathbf{L}\mathbf{T}^{(\mathrm{IMMRM})}\mathbf{V}_{A\boldsymbol{M}}(\boldsymbol{Y} - \mathbf{R}^\top\underline{\boldsymbol{\gamma}})$. By $E^*[\boldsymbol{\psi}^{(\mathrm{IMMRM})}(A, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{M}; \boldsymbol{\theta})] = \mathbf{0}$, we have $\underline{\boldsymbol{\gamma}} = E^*[\mathbf{R}\underline{\mathbf{V}}_{A\boldsymbol{M}}\mathbf{R}]^{-1}E^*[\mathbf{R}\underline{\mathbf{V}}_{A\boldsymbol{M}}\boldsymbol{Y}]$. Following a similar procedure for calculating $\underline{\boldsymbol{\beta}}$ in Lemma 5, we get

$$
\boldsymbol{Y} - \mathbf{R}^\top\underline{\boldsymbol{\gamma}} = \sum_{j=0}^J I\{A=j\}\left\{\boldsymbol{Y}(j) - E[\boldsymbol{Y}(j)] - Cov(\boldsymbol{Y}(j), \boldsymbol{X})Var(\boldsymbol{X})^{-1}(\boldsymbol{X} - E[\boldsymbol{X}])\right\}.
$$

We next compute $\boldsymbol{U}_2$. For each $j = 1, \dots, J$, we have the j-th entry of $\boldsymbol{U}_2$ is

$$
\boldsymbol{\delta}_j^\top\boldsymbol{U}_2 = \underline{\beta}_{AjK} + \underline{\boldsymbol{\beta}}_{A\boldsymbol{X}jK}^\top\boldsymbol{X} - \Delta_j^*.
$$

Since we have shown the model-robustness of $\widehat{\boldsymbol{\Delta}}^{(\mathrm{IMMRM})}$ in the proof of Theorem 1, then $\Delta_j^* = \underline{\beta}_{AjK} + \underline{\boldsymbol{\beta}}_{A\boldsymbol{X}jK}^\top E[\boldsymbol{X}]$ and hence $\boldsymbol{\delta}_j^\top\boldsymbol{U}_2 = \underline{\boldsymbol{\beta}}_{A\boldsymbol{X}jK}^\top(\boldsymbol{X} - E[\boldsymbol{X}])$. By the formula of $\boldsymbol{Y} - \mathbf{R}^\top\underline{\boldsymbol{\gamma}}$, we have

$$
\begin{aligned}
\underline{\boldsymbol{\beta}}_{A\boldsymbol{X}jK}^\top(\boldsymbol{X} - E[\boldsymbol{X}]) &= (I\{A=j\} - I\{A=0\})\boldsymbol{e}_K^\top(\boldsymbol{Y} - \mathbf{R}^\top\underline{\boldsymbol{\gamma}}) \\
&= Cov(Y_K(j) - Y_K(0), \boldsymbol{X})Var(\boldsymbol{X})^{-1}(\boldsymbol{X} - E[\boldsymbol{X}]) \\
&= (\boldsymbol{b}_{Kj} - \boldsymbol{b}_{K0})^\top(\boldsymbol{X} - E[\boldsymbol{X}]),
\end{aligned}
$$

which implies $\boldsymbol{U}_2 = \mathbf{L}\mathbf{r}^\top(\boldsymbol{X} - E[\boldsymbol{X}])$.

We next compute $\widetilde{\mathbf{V}}^{(\mathrm{IMMRM})}$, which is

$$\widetilde{\mathbf{V}}^{(\mathrm{IMMRM})} = E^*[IF^{(\mathrm{IMMRM})} IF^{(\mathrm{IMMRM})\top}] = E^*[\boldsymbol{U}_1\boldsymbol{U}_1^\top] + E^*[\boldsymbol{U}_1\boldsymbol{U}_2^\top] + E^*[\boldsymbol{U}_2\boldsymbol{U}_1^\top] + E^*[\boldsymbol{U}_2\boldsymbol{U}_2^\top].$$

For $E^*[\boldsymbol{U}_1\boldsymbol{U}_2^\top]$, since $E^*[\boldsymbol{\psi}^{(\mathrm{IMMRM})}(A, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{M})] = \mathbf{0}$ imply that $E^*[\mathbf{V}_{A\boldsymbol{M}}(\boldsymbol{Y} - \mathbf{R}^\top \underline{\boldsymbol{\gamma}})|\boldsymbol{X}] = \mathbf{0}$ and $E^*[(\mathbf{I}_K \otimes \boldsymbol{A})\mathbf{V}_{A\boldsymbol{M}}(\boldsymbol{Y} - \mathbf{R}^\top\underline{\boldsymbol{\gamma}})|\boldsymbol{X}] = \mathbf{0}$, then $E^*[\boldsymbol{U}_1|\boldsymbol{X}] = \mathbf{0}$ and hence $E^*[\boldsymbol{U}_1\boldsymbol{U}_2^\top] = E^*[E^*[\boldsymbol{U}_1|\boldsymbol{X}]\boldsymbol{U}_2^\top] = \mathbf{0}$. Thus, $\widetilde{\mathbf{V}}^{(\mathrm{IMMRM})} = E^*[\boldsymbol{U}_1\boldsymbol{U}_1^\top] + E^*[\boldsymbol{U}_2\boldsymbol{U}_2^\top]$. For $E^*[\boldsymbol{U}_1\boldsymbol{U}_1^\top]$, since $\boldsymbol{\Sigma}_j$ is unstructured for each $j = 0, \ldots, J$, the second set of estimating equations $\boldsymbol{\psi}^{(\mathrm{IMMRM})}$ implies that, for each $r, s = 1, \ldots, K$ and $j = 0, \ldots, J$, we have

$$\begin{aligned}
\mathbf{0} = E^* &\left[-\mathrm{tr}(\mathbf{V}_{A\boldsymbol{M}} I\{A = j\}(\boldsymbol{e}_r\boldsymbol{e}_s^\top + \boldsymbol{e}_s\boldsymbol{e}_r^\top))\right.\\
&\quad \left.+(\boldsymbol{Y} - \mathbf{R}^\top\underline{\boldsymbol{\gamma}})^\top \mathbf{V}_{A\boldsymbol{M}} I\{A = j\}(\boldsymbol{e}_r\boldsymbol{e}_s^\top + \boldsymbol{e}_s\boldsymbol{e}_r^\top))\mathbf{V}_{A\boldsymbol{M}}(\boldsymbol{Y} - \mathbf{R}^\top\underline{\boldsymbol{\gamma}})\right]\\
= E^* &\left[2I\{A = j\}\boldsymbol{e}_r^\top\{-\mathbf{V}_{A\boldsymbol{M}} + \mathbf{V}_{A\boldsymbol{M}}(\boldsymbol{Y} - \mathbf{R}^\top\underline{\boldsymbol{\gamma}})(\boldsymbol{Y} - \mathbf{R}^\top\underline{\boldsymbol{\gamma}})^\top\mathbf{V}_{A\boldsymbol{M}}\}\boldsymbol{e}_s\right],
\end{aligned}$$

which implies that, for $j = 0, \ldots, J$,

$$E[-\pi_j \underline{\mathbf{V}}_{j\boldsymbol{M}}] + E^*[\underline{\mathbf{V}}_{j\boldsymbol{M}} I\{A = j\}(\boldsymbol{Y} - \mathbf{R}^\top\underline{\boldsymbol{\gamma}})(\boldsymbol{Y} - \mathbf{R}^\top\underline{\boldsymbol{\gamma}})^\top\underline{\mathbf{V}}_{j\boldsymbol{M}}] = \mathbf{0}. \tag{20}$$

Hence

$$\begin{aligned}
E^*[\boldsymbol{U}_1\boldsymbol{U}_1^\top] &= E^*[\mathbf{L}\mathbf{T}^{(\mathrm{IMMRM})}\mathbf{V}_{A\boldsymbol{M}}(\boldsymbol{Y} - \mathbf{R}^\top\underline{\boldsymbol{\gamma}})(\boldsymbol{Y} - \mathbf{R}^\top\underline{\boldsymbol{\gamma}})^\top\mathbf{V}_{A\boldsymbol{M}}\mathbf{T}^{(\mathrm{IMMRM})}\mathbf{L}^\top]\\
&= \mathbf{L}\ diag\left\{\frac{1}{\pi_j^2}\boldsymbol{e}_K^\top E[\underline{\mathbf{V}}_{j\boldsymbol{M}}]^{-1}E^*[\mathbf{V}_{A\boldsymbol{M}} I\{A = j\}(\boldsymbol{Y} - \mathbf{R}^\top\underline{\boldsymbol{\gamma}})\right.\\
&\qquad \left.(\boldsymbol{Y} - \mathbf{R}^\top\underline{\boldsymbol{\gamma}})^\top\mathbf{V}_{A\boldsymbol{M}}]E[\underline{\mathbf{V}}_{j\boldsymbol{M}}]^{-1}\boldsymbol{e}_K : j = 0, \ldots, J\right\}\mathbf{L}^\top\\
&= \mathbf{L}\ diag\{\boldsymbol{e}_K^\top E[\pi_j\underline{\mathbf{V}}_{j\boldsymbol{M}}]^{-1}\boldsymbol{e}_K : j = 0, \ldots, J\}\mathbf{L}^\top,
\end{aligned}$$

where the last equation results from Equation (20). Since $E^*[\boldsymbol{U}_2\boldsymbol{U}_2^\top] = \mathbf{L}\mathbf{r}^\top Var(\boldsymbol{X})\mathbf{r}\mathbf{L}^\top$, we get the desired formula of $\widetilde{\mathbf{V}}^{(\mathrm{IMMRM})}$. □

E.2 Proof of Corollary 1

Proof. The ANHECOVA estimator is a special case of the IMMRM model setting $K = 1$. Hence the consistency and asymptotic normality under simple or stratified randomization are implied by Theorem 1. Furthermore, by Equation (14), we have

$$\widetilde{\mathbf{V}}^{(\text{ANHECOVA})} = \mathbf{L}\left(diag\{P^*(M_K = 1)^{-1}\pi_j^{-1}\boldsymbol{e}_K^\top \underline{\boldsymbol{\Sigma}}^{(\text{IMMRM})}\boldsymbol{e}_K : j = 0, \ldots, J\} + \mathbf{r}^\top Var(\boldsymbol{X})\mathbf{r}\right)\mathbf{L}.$$

Then, by Lemma 1 (2),

$$\begin{aligned}
&\widetilde{\mathbf{V}}^{(\text{ANHECOVA})} - \widetilde{\mathbf{V}}^{(\text{IMMRM})}\\
&= \mathbf{L}diag\{\pi_j^{-1}\boldsymbol{e}_K^\top(P^*(M_K = 1)^{-1}\underline{\boldsymbol{\Sigma}}^{(\text{IMMRM})} - E[\underline{\mathbf{V}}_{jM}]^{-1})\boldsymbol{e}_K : j = 0, \ldots, J\}\mathbf{L}\\
&\succeq \mathbf{0},
\end{aligned}$$

with equality holds if and only if $P(M_t(j) = 1, M_K(j) = 0)\ Cov\{Y_t(j) - \boldsymbol{b}_{tj}^\top\boldsymbol{X}, Y_K(j) - \boldsymbol{b}_{Kj}^\top\boldsymbol{X}\} = 0$ for each $t = 1, \ldots, K-1$ and $j = 0, \ldots, J$. The final result comes from the fact that $Cov\{\boldsymbol{b}_{tj}^\top\boldsymbol{X}, Y_K(j) - \boldsymbol{b}_{Kj}^\top\boldsymbol{X}\} = \mathbf{0}$. □

E.3 Proof of Corollary 2

Proof. By Equations (16) and (17), we have

$$\begin{aligned}
&\widetilde{\mathbf{V}}^{(\text{ANCOVA})} - \mathbf{V}^{(\text{ANCOVA})}\\
&= \widetilde{\mathbf{V}}^{(\text{MMRM-II})} - \mathbf{V}^{(\text{MMRM-II})}\\
&= \mathbf{L}[diag\{\pi_j^{-1}(\boldsymbol{b}_{Kj} - \boldsymbol{b}_K)^\top Var(E[\boldsymbol{X}|S])(\boldsymbol{b}_{Kj} - \boldsymbol{b}_K) : j = 0, \ldots, J\} - \mathbf{z}^\top Var(E[\boldsymbol{X}|S])\mathbf{z}]\mathbf{L}^\top.
\end{aligned}$$

If $J = 1$ and $\pi_1 = \pi_0 = 0.5$, then $\mathbf{L} = (-1, 1)$, $\boldsymbol{b}_K = 0.5(\boldsymbol{b}_{K1} + \boldsymbol{b}_{K0})$ and $\mathbf{z}\mathbf{L}^\top = \boldsymbol{b}_{K1} - \boldsymbol{b}_{K0}$. Hence

$$
\begin{aligned}
&\widetilde{V}^{(\text{ANCOVA})} - V^{(\text{ANCOVA})} \\
&= \sum_{j=0}^{1} 2\{\boldsymbol{b}_{K1} - 0.5(\boldsymbol{b}_{K1} + \boldsymbol{b}_{K0})\}^\top Var(E[\boldsymbol{X}|S])\{\boldsymbol{b}_{K1} - 0.5(\boldsymbol{b}_{K1} + \boldsymbol{b}_{K0})\} \\
&\quad - (\boldsymbol{b}_{K1} - \boldsymbol{b}_{K0})^\top Var(E[\boldsymbol{X}|S])(\boldsymbol{b}_{K1} - \boldsymbol{b}_{K0}) \\
&= 0.
\end{aligned}
$$

We next compare $\mathbf{V}^{(\text{ANCOVA})}$ and $\mathbf{V}^{(\text{MMRM-II})}$. By the definition of $\underline{\boldsymbol{\Sigma}}^{(\text{ANCOVA})}$ and $\underline{\boldsymbol{\Sigma}}_j^{(\text{ANCOVA})}$, we have $2\underline{\boldsymbol{\Sigma}}^{(\text{ANCOVA})} = \underline{\boldsymbol{\Sigma}}_0^{(\text{ANCOVA})} + \underline{\boldsymbol{\Sigma}}_1^{(\text{ANCOVA})}$. Then Equation (11) implies that $\widetilde{V}^{(\text{ANCOVA})} = 4P^*(M_K = 1)^{-1}\boldsymbol{e}_K^\top \underline{\boldsymbol{\Sigma}}^{(\text{ANCOVA})}\boldsymbol{e}_K$. Similarly, we have $\widetilde{V}^{(\text{MMRM-II})} = 4\boldsymbol{e}_K^\top E^*[\underline{\overline{\mathbf{V}}_{\boldsymbol{M}}}]^{-1}\boldsymbol{e}_K$. Since $\underline{\boldsymbol{\Sigma}}^{(\text{ANCOVA})} = \underline{\boldsymbol{\Sigma}}^{(\text{MMRM-II})}$, then Lemma 1 (2) implies that $\widetilde{V}^{(\text{ANCOVA})} - \widetilde{V}^{(\text{MMRM-II})} \geq \boldsymbol{e}_K^\top(\underline{\boldsymbol{\Sigma}}^{(\text{ANCOVA})} - \underline{\boldsymbol{\Sigma}}^{(\text{MMRM-II})})\boldsymbol{e}_K = 0$.

Finally, we show $V^{(\text{MMRM-I})} \geq V^{(\text{MMRM-II})}$. Under two-armed equal randomization, Equation (15) implies that

$$
\widetilde{V}^{(\text{MMRM-I})} - V^{(\text{MMRM-I})} = 4\{\boldsymbol{b}_K - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}\}^\top Var(E[\boldsymbol{X}|S])\{\boldsymbol{b}_K - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}\}.
$$

In addition, since $2\underline{\boldsymbol{\Sigma}}^{(\text{MMRM-I})} = \underline{\boldsymbol{\Sigma}}_0^{(\text{MMRM-I})} + \underline{\boldsymbol{\Sigma}}_1^{(\text{MMRM-I})}$, Equation (12) implies that $\widetilde{\mathbf{V}}^{(\text{MMRM-I})} = 4\boldsymbol{e}_K^\top E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1}\boldsymbol{e}_K$. By the definition of $\underline{\boldsymbol{\Sigma}}^{(\text{MMRM-I})}$ and $\underline{\boldsymbol{\Sigma}}^{(\text{MMRM-II})}$, we have

$$
\begin{aligned}
&\boldsymbol{e}_K^\top(\underline{\boldsymbol{\Sigma}}^{(\text{MMRM-I})} - \underline{\boldsymbol{\Sigma}}^{(\text{MMRM-II})})\boldsymbol{e}_K \\
&= Var^*(Y_K) + \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}^\top Var(\boldsymbol{X})\underline{\boldsymbol{\beta}}_{\boldsymbol{X}} - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}^\top Cov^*(\boldsymbol{X}, Y_K) - Cov^*(Y_K, \boldsymbol{X})\underline{\boldsymbol{\beta}}_{\boldsymbol{X}} \\
&\quad - Var^*(Y_K) - \boldsymbol{b}_K^\top Var(\boldsymbol{X})\boldsymbol{b}_K \\
&= \{\boldsymbol{b}_K - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}\}^\top Var(\boldsymbol{X})\{\boldsymbol{b}_K - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}\}.
\end{aligned}
$$

Hence

$$
\begin{aligned}
&V^{(\text{MMRM-I})} - V^{(\text{MMRM-II})} \\
&= V^{(\text{MMRM-I})} - \widetilde{V}^{(\text{MMRM-I})} + \widetilde{V}^{(\text{MMRM-I})} - V^{(\text{MMRM-II})} \\
&= -4\{\boldsymbol{b}_K - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}\}^\top Var(E[\boldsymbol{X}|S])\{\boldsymbol{b}_K - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}\} + 4\boldsymbol{e}_K^\top(E^*[\underline{\mathbf{V}}_{\boldsymbol{M}}]^{-1} - E^*[\overline{\underline{\mathbf{V}}}_{\boldsymbol{M}}]^{-1})\boldsymbol{e}_K \\
&\geq -4\{\boldsymbol{b}_K - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}\}^\top Var(E[\boldsymbol{X}|S])\{\boldsymbol{b}_K - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}\} + 4\boldsymbol{e}_K^\top(\underline{\boldsymbol{\Sigma}}^{(\text{MMRM-I})} - \underline{\boldsymbol{\Sigma}}^{(\text{MMRM-II})})\boldsymbol{e}_K \\
&= 4\{\boldsymbol{b}_K - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}\}^\top E[Var(\boldsymbol{X}|S)]\{\boldsymbol{b}_K - \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}\} \\
&\geq 0,
\end{aligned}
$$

where the inequality comes from Lemma 1 (5). □

F A counterexample showing MMRM-II is less precise than ANCOVA

We assume Assumption 1, simple randomization, $\pi_1 = \frac{1}{3}$, $\pi_0 = \frac{2}{3}$, and

$$
\begin{aligned}
E[\boldsymbol{Y}(1)] &= E[\boldsymbol{Y}(0)] = \mathbf{0}, \\
Var(\boldsymbol{Y}(1)) &= \begin{pmatrix} 4 & -3 \\ -3 & 4 \end{pmatrix}, \quad Var(\boldsymbol{Y}(0)) = \begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}, \\
p_{(1,0)} = p_{(0,1)} = p_{(1,1)} &= \frac{1}{3},
\end{aligned}
$$

where $p_{\boldsymbol{m}} = P^*(\boldsymbol{M} = \boldsymbol{m})$ for $\boldsymbol{m} \in \{0,1\}^2$. Then we have $\underline{\boldsymbol{\Sigma}}_j^{(\text{ANCOVA})} = \underline{\boldsymbol{\Sigma}}_j^{(\text{MMRM-II})} = Var(\boldsymbol{Y}(j))$ for $j = 0, 1$. Furthermore,

$$
\underline{\boldsymbol{\Sigma}}^{(\text{MMRM-II})} = \pi_0 \underline{\boldsymbol{\Sigma}}_0^{(\text{MMRM-II})} + \pi_1 \underline{\boldsymbol{\Sigma}}_1^{(\text{MMRM-II})} = \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix}
$$

We define

$$\mathbf{C} = \pi_0^{-1}\underline{\boldsymbol{\Sigma}}_0^{(\text{ANCOVA})} + \pi_1^{-1}\underline{\boldsymbol{\Sigma}}_1^{(\text{ANCOVA})} = \frac{9}{2}\begin{pmatrix} 4 & -1 \\ -1 & 4 \end{pmatrix}.$$

Then, by Equation (11),

$$Var(\widehat{\Delta}^{(\text{ANCOVA})}) = \frac{1}{P(M_K = 1)}\boldsymbol{e}_K^\top \mathbf{C}\boldsymbol{e}_K = 27.$$

To compute $Var(\widehat{\Delta}^{(\text{MMRM-II})})$, recall that $\underline{\overline{\mathbf{V}}}_{\boldsymbol{M}} = \mathbf{V}_{\boldsymbol{M}}(\underline{\boldsymbol{\Sigma}}^{(\text{MMRM-II})})$. Then

$$\underline{\overline{\mathbf{V}}}_{(1,0)} = \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & 0 \end{pmatrix}, \quad \underline{\overline{\mathbf{V}}}_{(0,1)} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{4} \end{pmatrix}, \quad \underline{\overline{\mathbf{V}}}_{(1,1)} = \begin{pmatrix} \frac{4}{15} & -\frac{1}{15} \\ -\frac{1}{15} & \frac{4}{15} \end{pmatrix},$$

which implies

$$E^*[\underline{\overline{\mathbf{V}}}_{\boldsymbol{M}}] = \sum_{\boldsymbol{m}} p_{\boldsymbol{m}}\underline{\overline{\mathbf{V}}}_{\boldsymbol{m}} = \frac{1}{180}\begin{pmatrix} 31 & -4 \\ -4 & 31 \end{pmatrix}, \quad E^*[\underline{\overline{\mathbf{V}}}_{\boldsymbol{M}}]^{-1} = \frac{4}{21}\begin{pmatrix} 31 & 4 \\ 4 & 31 \end{pmatrix}.$$

In addition, we have

$$\underline{\overline{\mathbf{V}}}_{(1,0)}\mathbf{C}\underline{\overline{\mathbf{V}}}_{(1,0)} = \begin{pmatrix} \frac{9}{8} & 0 \\ 0 & 0 \end{pmatrix}, \; \underline{\overline{\mathbf{V}}}_{(0,1)}\mathbf{C}\underline{\overline{\mathbf{V}}}_{(0,1)} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{9}{8} \end{pmatrix}, \; \underline{\overline{\mathbf{V}}}_{(1,1)}\mathbf{C}\underline{\overline{\mathbf{V}}}_{(1,1)} = \frac{1}{50}\begin{pmatrix} 76 & -49 \\ -49 & 76 \end{pmatrix},$$

which implies

$$E^*[\underline{\overline{\mathbf{V}}}_{\boldsymbol{M}}\mathbf{C}\underline{\overline{\mathbf{V}}}_{\boldsymbol{M}}] = \sum_{\boldsymbol{m}} p_{\boldsymbol{m}}\underline{\overline{\mathbf{V}}}_{\boldsymbol{m}}\mathbf{C}\underline{\overline{\mathbf{V}}}_{\boldsymbol{m}} = \frac{1}{600}\begin{pmatrix} 529 & -196 \\ -196 & 529 \end{pmatrix}.$$

Then, by $\underline{\boldsymbol{\Sigma}}_j^{(\text{MMRM-II})} = \underline{\boldsymbol{\Sigma}}_j^{(\text{ANCOVA})}$ and Equation (13), we have

$$\begin{aligned}
Var(\widehat{\Delta}^{(\text{MMRM-II})}) &= \boldsymbol{e}_K^\top E^*[\underline{\overline{\mathbf{V}}}_{\boldsymbol{M}}]^{-1} E^*[\underline{\overline{\mathbf{V}}}_{\boldsymbol{M}}\mathbf{C}\underline{\overline{\mathbf{V}}}_{\boldsymbol{M}}] E^*[\underline{\overline{\mathbf{V}}}_{\boldsymbol{M}}]^{-1}\boldsymbol{e}_K \\
&= \frac{4}{21}(4\ 31)\frac{1}{600}\begin{pmatrix} 529 & -196 \\ -196 & 529 \end{pmatrix}\frac{4}{21}\begin{pmatrix} 4 \\ 31 \end{pmatrix} \\
&= \frac{12486}{21^2} \approx 28.31.
\end{aligned}$$

Since $27 < 28.31$, we have $Var(\widehat{\Delta}^{(\text{ANCOVA})}) < Var(\widehat{\Delta}^{(\text{MMRM-II})})$.

G Missing data mechanism in the simulation study

Given $(\boldsymbol{Y}_i(0), \boldsymbol{Y}_i(1), \boldsymbol{Y}_i(2), A_i, X_{i1})$ defined in Section 6.1 of the main paper, we define $R_{it}(j)$ as the residual of $Y_{it}(j)$ regressing on X_{i1} by a simple linear regression. We then define the censoring time C_i by the following sequential conditional model

1. $P(C_i = 1|X_{i1}, A_i) = 1 - expit[logit(0.99) - 0.14I\{A_i < 2\}X_{i1} - 0.12I\{A_i = 2\}X_{i1}]$,

2. $P(C_i = 2|X_{i1}, A_i, C_i > 1, Y_{i1}) = 1 - expit[logit(0.969) - 0.7I\{A_i < 2\}R_{i1}(A_i) - 0.5I\{A_i = 2\}R_{i1}(A_i)]$,

3. $P(C_i = 3|X_{i1}, A_i, C_i > 2, Y_{i1}, Y_{i2}) = 1 - expit[logit(0.958) - 0.72I\{A_i < 2\}R_{i2}(A_i) - 0.51I\{A_i = 2\}R_{i2}(A_i)]$,

4. $P(C_i = 4|X_{i1}, A_i, C_i > 3, Y_{i1}, Y_{i2}, Y_{i3}) = 1 - expit[logit(0.967) - 0.74I\{A_i < 2\}R_{i3}(A_i) - 0.52I\{A_i = 2\}R_{i3}(A_i)]$,

5. $P(C_i = 5|X_{i1}, A_i, C_i > 4, Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4}) = 1 - expit[logit(0.977) - 0.76I\{A_i < 2\}R_{i4}(A_i) - 0.53I\{A_i = 2\}R_{i4}(A_i)]$.

Once we have C_i, then $\boldsymbol{M}_i$ is defined as $M_{it} = 1$ if $t \leq C_i$ and 0 otherwise.

H Additional simulations under homogeneity and homoscedasticity

References

Bugni, F. A., Canay, I. A., and Shaikh, A. M. (2019). Inference under covariate-adaptive randomization with multiple treatments. *Quantitative Economics*, 10(4):1747–1785.

Table 1: Simulation results comparing candidate estimators with 50 samples per arm under MCAR and MAR with no heterogeneity or heteroscedasticity. For each estimator, we estimate the average treatment effect of TRT1 and TRT2, both comparing the control group. The following measures are used: bias, empirical standard error (ESE), averaged standard error (ASE), coverage probability (CP), probability of rejecting the null (PoR), relative mean squared error compared to IMMRM (RMSE). For RMSE, a number bigger than 1 indicates a larger mean squared error than IMMRM.

		Group	Bias	ESE	ASE	CP(%)	PoR(%)	RMSE
MCAR	ANCOVA	TRT1	0.005	0.197	0.190	94.6	5.4	1.016
		TRT2	0.004	0.196	0.190	94.6	99.9	1.032
	MMRM-I	TRT1	0.004	0.198	0.205	95.9	4.1	1.026
		TRT2	0.004	0.197	0.206	95.9	99.8	1.048
	MMRM-II	TRT1	0.006	0.193	0.186	94.3	5.7	0.976
		TRT2	0.005	0.191	0.186	94.2	99.9	0.981
	IMMRM	TRT1	0.006	0.195	0.183	93.8	6.2	-
		TRT2	0.005	0.193	0.183	93.5	99.9	-
MAR	ANCOVA	TRT1	-0.001	0.197	0.191	94.9	5.2	0.99
		TRT2	0.005	0.197	0.191	94.3	99.7	1.010
	MMRM-I	TRT1	0	0.197	0.208	96.6	3.4	0.995
		TRT2	-0.002	0.196	0.207	96	99.7	1.003
	MMRM-II	TRT1	-0.001	0.195	0.187	94.9	5.2	0.969
		TRT2	0	0.193	0.186	93.9	99.9	0.974
	IMMRM	TRT1	-0.002	0.198	0.185	93.7	6.3	-
		TRT2	0	0.196	0.184	93.4	99.8	-

Table 2: Simulation results comparing candidate estimators with 200 samples per arm under MCAR and MAR with no heterogeneity or heteroscedasticity. For each estimator, we estimate the average treatment effect of TRT1 and TRT2, both comparing the control group. The following measures are used: bias, empirical standard error (ESE), averaged standard error (ASE), coverage probability (CP), probability of rejecting the null (PoR), relative mean squared error compared to IMMRM (RMSE). For RMSE, a number bigger than 1 indicates a larger mean squared error than IMMRM.

		Group	Bias	ESE	ASE	CP(%)	PoR(%)	RMSE
MCAR	ANCOVA	TRT1	0.002	0.098	0.097	94.8	5.2	1.032
		TRT2	0	0.098	0.097	95.4	100	1.032
	MMRM-I	TRT1	0.002	0.099	0.104	96.3	3.7	1.054
		TRT2	0.001	0.099	0.104	96.1	100	1.054
	MMRM-II	TRT1	0.002	0.096	0.095	94.4	5.6	1.000
		TRT2	0.001	0.096	0.095	95.2	100	0.989
	IMMRM	TRT1	0.002	0.096	0.095	94.4	5.6	-
		TRT2	0.001	0.096	0.095	94.9	100	-
MAR	ANCOVA	TRT1	-0.001	0.099	0.098	94.7	5.4	1.032
		TRT2	0.002	0.098	0.098	94.6	100	1.043
	MMRM-I	TRT1	0	0.099	0.105	96.2	3.8	1.032
		TRT2	-0.002	0.099	0.105	96	100	1.054
	MMRM-II	TRT1	0	0.097	0.096	94.8	5.2	0.989
		TRT2	-0.001	0.096	0.095	94.7	100	0.989
	IMMRM	TRT1	0	0.097	0.096	94.4	5.6	-
		TRT2	-0.001	0.096	0.095	94.5	100	-

Dwyer, P. S. (1967). Some applications of matrix derivatives in multivariate analysis. *Journal of the American Statistical Association*, 62(318):607–625.

MacRae, E. C. et al. (1974). Matrix derivatives with an application to an adaptive linear decision problem. *The Annals of Statistics*, 2(2):337–346.

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wang, B., Susukida, R., Mojtabai, R., Amin-Esmaeili, M., and Rosenblum, M. (2023). Model-robust inference for clinical trials that improve precision by stratified randomization and covariate adjustment. *Journal of the American Statistical Association*, 118(542):1152–1163.