ROBUST SCORE TESTS FOR CENSORED OUTCOMES AND INCOMPLETE COVARIATES LEVERAGING HIGH-DIMENSIONAL AUXILIARY VARIABLES

Jiahui Feng[1] and Kin Yau Wong[1,2]

[1] *The Hong Kong Polytechnic University*

[2] *Hong Kong Polytechnic University Shenzhen Research Institute*

S1 Proof of Theorem 1

For simplicity, we temporarily suppress the transformation function index j. We extend the definitions of ξ_i and G_i as functions of $\boldsymbol{\alpha}$ and a general monotone function Λ (instead of a vector of jumps), such that $\xi_i(\boldsymbol{\alpha}, \Lambda) = \int_0^{Y_i} \exp(\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{X}_i)\, \mathrm{d}\Lambda(s)$ and $G_i(\boldsymbol{\alpha}, \Lambda) = G\{\xi_i(\boldsymbol{\alpha}, \Lambda)\}$. The (scaled) score statistic for β can be written as

$$U_\beta(\widehat{\boldsymbol{\alpha}}, \widehat{\Lambda}, \widehat{\boldsymbol{\gamma}}_{\mathcal{K}}) = n^{-1/2} \sum_{i=1}^{n} \mu_{1,i}(\widehat{\boldsymbol{\alpha}}, \widehat{\Lambda}) \big\{ R_i S_i + (1 - R_i) \widehat{\boldsymbol{\gamma}}_{\mathcal{K}}^{\mathrm{T}} \boldsymbol{W}_{\mathcal{K},i} \big\},$$

where

$$\mu_{1,i}(\boldsymbol{\alpha}, \Lambda) = \Delta_i + \Delta_i \psi_i(\boldsymbol{\alpha}, \Lambda) \xi_i(\boldsymbol{\alpha}, \Lambda) - G_i'(\boldsymbol{\alpha}, \Lambda) \xi_i(\boldsymbol{\alpha}, \Lambda).$$

Let μ_1 be $\mu_{1,i}$ for a generic subject. Define the function

$$I_{\beta\Lambda}(\cdot) = -\mathrm{E}\big[\mu_1'(\boldsymbol{\alpha}_0, \Lambda_0) \exp(\boldsymbol{\alpha}_0^{\mathrm{T}} \boldsymbol{X}) \big\{ RS + (1 - R) \boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}} \boldsymbol{W}_{\mathcal{K}} \big\} I(\cdot \le Y) \big],$$

where $\mu_1'(\boldsymbol{\alpha}, \Lambda) = \Delta \eta(\boldsymbol{\alpha}, \Lambda) \xi(\boldsymbol{\alpha}, \Lambda) + \Delta \psi(\boldsymbol{\alpha}, \Lambda) - G''(\boldsymbol{\alpha}, \Lambda) \xi(\boldsymbol{\alpha}, \Lambda) - G'(\boldsymbol{\alpha}, \Lambda)$, and $G'(\boldsymbol{\alpha}, \Lambda)$ and $G''(\boldsymbol{\alpha}, \Lambda)$ denote $G'\{\xi(\boldsymbol{\alpha}, \Lambda)\}$ and $G''\{\xi(\boldsymbol{\alpha}, \Lambda)\}$ for a generic subject. Note that we can also write the true-measure counterparts of $\widehat{\boldsymbol{I}}_{\beta\alpha}$ and $\widehat{\boldsymbol{I}}_{\beta\gamma}$ as

$$\boldsymbol{I}_{\beta\alpha} = -\mathrm{E}\big[\mu_1'(\boldsymbol{\alpha}_0, \Lambda_0)\xi(\boldsymbol{\alpha}_0, \Lambda_0)\boldsymbol{X}\big\{RS + (1-R)\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K}}\big\}\big],$$

$$\boldsymbol{I}_{\beta\gamma} = -\mathrm{E}\big\{\mu_1(\boldsymbol{\alpha}_0, \Lambda_0)(1-R)\boldsymbol{W}_{\mathcal{K}}\big\}.$$

Let $\boldsymbol{I}_{\gamma\gamma}$ be the expectation of $\widehat{\boldsymbol{I}}_{\gamma\gamma}$. Also, let $\widehat{I}_{\beta\Lambda}(\cdot)$ be the empirical counterpart of $I_{\beta\Lambda}(\cdot)$, with the expectation replaced by an empirical mean. For notational convenience, we denote $\mu_1(\boldsymbol{\alpha}_0, \Lambda_0)$ and $\mu_1(\widehat{\boldsymbol{\alpha}}, \widehat{\Lambda})$ by μ_1 and $\widehat{\mu}_1$, respectively.

To prove Theorem 1, we need to first obtain the limiting distribution of $\sqrt{n}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0, \widehat{\Lambda} - \Lambda_0)$. Define the set

$$\mathcal{H} = \{\, (\boldsymbol{h}_\alpha, h_\Lambda) : \boldsymbol{h}_\alpha \in \mathbb{R}^{\|\boldsymbol{\alpha}\|_0}, h_\Lambda \text{ is a function with bounded variation on } [0,\tau];$$
$$\|\boldsymbol{h}_\alpha\| \le 1, \|h_\Lambda\|_V \le 1\}.$$

Let $\ell(\boldsymbol{\alpha}, \Lambda)$ be the log-likelihood for the survival model under H_0 for a generic subject, that is,

$$\ell(\boldsymbol{\alpha}, \Lambda) = \Delta\Big\{\log\lambda(Y) + \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{X} + \log G'(\boldsymbol{\alpha}, \Lambda)\Big\} - G(\boldsymbol{\alpha}, \Lambda),$$

where $G(\boldsymbol{\alpha}, \Lambda)$ denotes $G\{\xi(\boldsymbol{\alpha}, \Lambda)\}$ for a generic subject. We define the derivative of $\ell(\boldsymbol{\alpha}, \Lambda)$ along $\big(\boldsymbol{h}_\alpha, h_\Lambda\big)$ as

$$\begin{aligned}\ell_{\alpha\Lambda}(\boldsymbol{\alpha}, \Lambda)[\boldsymbol{h}_\alpha, h_\Lambda] &= \frac{\mathrm{d}}{\mathrm{d}\epsilon}\ell\Big(\boldsymbol{\alpha} + \epsilon\boldsymbol{h}_\alpha, \Lambda + \epsilon\int h_\Lambda(s)\,\mathrm{d}\Lambda(s)\Big)\Big|_{\epsilon=0}\\ &= \Delta\{\boldsymbol{h}_\alpha^{\mathrm{T}}\boldsymbol{X} + h_\Lambda(Y)\} + \Delta\psi(\boldsymbol{\alpha}, \Lambda)\exp(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{X})\int\{\boldsymbol{h}_\alpha^{\mathrm{T}}\boldsymbol{X} + h_\Lambda(s)\}\,\mathrm{d}\Lambda(s)\\ &\quad - G'(\boldsymbol{\alpha}, \Lambda)\exp(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{X})\int\{\boldsymbol{h}_\alpha^{\mathrm{T}}\boldsymbol{X} + h_\Lambda(s)\}\,\mathrm{d}\Lambda(s).\end{aligned}$$

Clearly, $\mathbb{P}\big\{\ell_{\alpha\Lambda}(\boldsymbol{\alpha}_0, \Lambda_0)[\boldsymbol{h}_\alpha, h_\Lambda]\big\} = 0$ under H_0, where $\mathbb{P}$ is the true probability measure.

By the Taylor series expansion, we have

$$
\begin{aligned}
&\mathbb{P}\Big\{\ell_{\alpha\Lambda}(\widehat{\boldsymbol{\alpha}},\widehat{\Lambda})[\boldsymbol{h}_\alpha,h_\Lambda]-\ell_{\alpha\Lambda}(\boldsymbol{\alpha}_0,\Lambda_0)[\boldsymbol{h}_\alpha,h_\Lambda]\Big\}\\
=&(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)^{\mathrm{T}}\mathbb{P}\Bigg(\boldsymbol{X}\Bigg[\Delta\Big\{\psi(\boldsymbol{\alpha}_0,\Lambda_0)\exp(\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X})+\eta(\boldsymbol{\alpha}_0,\Lambda_0)\exp(2\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X})\Lambda_0(Y)\Big\}\\
&-\Big\{G'(\boldsymbol{\alpha}_0,\Lambda_0)\exp(\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X})+G''(\boldsymbol{\alpha}_0,\Lambda_0)\exp(2\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X})\Lambda_0(Y)\Big\}\Bigg]\int_0^Y\{\boldsymbol{h}_\alpha^{\mathrm{T}}\boldsymbol{X}+h_\Lambda(s)\}\,\mathrm{d}\Lambda_0(s)\Bigg)\\
&+\int_0^\tau\mathbb{P}\Bigg(I(s\le Y)\Bigg[\Big\{\Delta\psi(\boldsymbol{\alpha}_0,\Lambda_0)-G'(\boldsymbol{\alpha}_0,\Lambda_0)\Big\}\exp(\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X})\{\boldsymbol{h}_\alpha^{\mathrm{T}}\boldsymbol{X}+h_\Lambda(s)\}\\
&+\Big\{\Delta\eta(\boldsymbol{\alpha}_0,\Lambda_0)-G''(\boldsymbol{\alpha}_0,\Lambda_0)\Big\}\exp(2\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X})\int_0^Y\{\boldsymbol{h}_\alpha^{\mathrm{T}}\boldsymbol{X}+h_\Lambda(t)\}\,\mathrm{d}\Lambda_0(t)\Bigg]\Bigg)\,\mathrm{d}(\widehat{\Lambda}-\Lambda_0)(s)+o(1)\\
\equiv&(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)^{\mathrm{T}}\boldsymbol{W}_\alpha(\boldsymbol{h}_\alpha,h_\Lambda)+\int_0^\tau W_\Lambda(\boldsymbol{h}_\alpha,h_\Lambda)(s)\,\mathrm{d}(\widehat{\Lambda}-\Lambda_0)(s)+o(1).
\end{aligned}
$$

With the above arguments, we can use Theorem 3.3.1 of van der Vaart and Wellner (1996) to establish the weak convergence of $\sqrt{n}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0,\widehat{\Lambda}-\Lambda_0)$, following the arguments of Zeng, Lin and Lin (2008). We obtain the following lemma, which is analogous to Theorem 2 of Zeng, Lin and Lin (2008).

Lemma S1. *Under H_0 and Conditions (C1), (C7), and (C8), for any $(\boldsymbol{h}_\alpha,h_\Lambda)\in\mathcal{H}$, we have*

$$
\sqrt{n}\Big\{\boldsymbol{h}_\alpha^{\mathrm{T}}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)+\int_0^\tau h_\Lambda(s)\,\mathrm{d}(\widehat{\Lambda}-\Lambda_0)(s)\Big\}=-\sqrt{n}(\mathbb{P}_n-\mathbb{P})\ell_{\alpha\Lambda}(\boldsymbol{\alpha}_0,\Lambda_0)[\widetilde{\boldsymbol{h}}_\alpha,\widetilde{h}_\Lambda]+o_p(1),
$$

where $(\widetilde{\boldsymbol{h}}_\alpha,\widetilde{h}_\Lambda)=(\boldsymbol{W}_\alpha,W_\Lambda)^{-1}(\boldsymbol{h}_\alpha,h_\Lambda)$, and $\mathbb{P}_n$ is the empirical probability measure.

We now turn to the proof of Theorem 1.

Proof of Theorem 1. First we consider the score statistic, $U_\beta^{(j)}(\widehat{\boldsymbol{\alpha}}^{(j)},\widehat{\Lambda}^{(j)},\widehat{\boldsymbol{\gamma}}_{\mathcal{K}})$, under a single transformation model. For simplicity, we suppress the index j. Note that for any fixed $\mathcal{K}$, the Taylor series expansion of $U_\beta(\widehat{\boldsymbol{\alpha}}, \widehat{\Lambda}, \widehat{\boldsymbol{\gamma}}_{\mathcal{K}})$ at $(\boldsymbol{\alpha}_0, \Lambda_0, \boldsymbol{\gamma}_{0\mathcal{K}})$ is

$$
\begin{aligned}
&U_\beta(\widehat{\boldsymbol{\alpha}}, \widehat{\Lambda}, \widehat{\boldsymbol{\gamma}}_{\mathcal{K}})\\
=&\frac{1}{n^{1/2}}\sum_{i=1}^n\Big[\mu_{1,i}\big\{R_iS_i+(1-R_i)\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i}\big\}-\widehat{\boldsymbol{I}}_{\beta\alpha}^{\mathrm{T}}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)\\
&-\int_0^\tau \widehat{I}_{\beta\Lambda}(s)\,\mathrm{d}(\widehat{\Lambda}-\Lambda_0)(s)-\widehat{\boldsymbol{I}}_{\beta\gamma}^{\mathrm{T}}\widehat{\boldsymbol{I}}_{\gamma\gamma}^{-1}\boldsymbol{W}_{\mathcal{K},i}R_i(S_i-\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i})\Big]+o_p(1)\\
=&\frac{1}{n^{1/2}}\sum_{i=1}^n\Big[\mu_{1,i}\{R_iS_i+(1-R_i)\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i}\}-\boldsymbol{I}_{\beta\alpha}^{\mathrm{T}}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)\\
&-\int_0^\tau I_{\beta\Lambda}(s)\,\mathrm{d}(\widehat{\Lambda}-\Lambda_0)(s)-\boldsymbol{I}_{\beta\gamma}^{\mathrm{T}}\boldsymbol{I}_{\gamma\gamma}^{-1}\boldsymbol{W}_{\mathcal{K},i}R_i(S_i-\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i})\Big]+o_p(1)\\
=&\frac{1}{n^{1/2}}\sum_{i=1}^n\Big[\mu_{1,i}\big\{R_iS_i+(1-R_i)\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i}+\widetilde{\boldsymbol{q}}_\alpha^{\mathrm{T}}\boldsymbol{X}_i\big\}+\mu_{2,i}-\boldsymbol{I}_{\beta\gamma}^{\mathrm{T}}\boldsymbol{I}_{\gamma\gamma}^{-1}\boldsymbol{W}_{\mathcal{K},i}R_i(S_i-\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i})\Big]\\
&+o_p(1),
\end{aligned}
\tag{S1}
$$

where

$$
\begin{aligned}
\mu_{2,i}=&-\Delta_i\widetilde{q}_\Lambda(Y_i)-\Delta_i\frac{G_i''(\boldsymbol{\alpha}_0,\Lambda_0)}{G_i'(\boldsymbol{\alpha}_0,\Lambda_0)}\exp(\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X}_i)\int_0^\tau I(s\le Y_i)\widetilde{q}_\Lambda(s)\,\mathrm{d}\Lambda_0(s)\\
&+G_i'(\boldsymbol{\alpha}_0,\Lambda_0)\exp(\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X}_i)\int_0^\tau I(s\le Y_i)\widetilde{q}_\Lambda(s)\,\mathrm{d}\Lambda_0(s),
\end{aligned}
$$

and $(\widetilde{\boldsymbol{q}}_\alpha, \widetilde{q}_\Lambda) = (\boldsymbol{W}_\alpha, W_\Lambda)^{-1}(\boldsymbol{I}_{\beta\alpha}, I_{\beta\Lambda})$; the existence of the inverse is guaranteed by Condition (C8). The second equality follows from the convergence of $\widehat{\boldsymbol{\gamma}}_{\mathcal{K}}$, $\widehat{\boldsymbol{I}}_{\beta\alpha}$ and $\widehat{I}_{\beta\Lambda}$ to the true values (by Lemma S2 in Section S3) and the convergence of $n^{-1/2}\sum_{i=1}^n(\widehat{\boldsymbol{I}}_{\beta\gamma}^{\mathrm{T}}\widehat{\boldsymbol{I}}_{\gamma\gamma}^{-1} - \boldsymbol{I}_{\beta\gamma}^{\mathrm{T}}\boldsymbol{I}_{\gamma\gamma}^{-1})\boldsymbol{W}_{\mathcal{K},i}R_i(S_i-\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i})$ to zero (by Lemma S4 in Section S3). The third equality follows from the convergence of $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\Lambda}$ (by Lemma S1). The first term on the right-hand side of (S1) can be written as

$$
\begin{aligned}
&\frac{1}{n^{1/2}}\sum_{i=1}^n\Big[\big\{\mu_{1,i}-\mathrm{E}(\mu_1\mid R_i,\boldsymbol{X}_i)\big\}\big\{R_iS_i+(1-R_i)\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i}+\widetilde{\boldsymbol{q}}_\alpha^{\mathrm{T}}\boldsymbol{X}_i\big\}\\
&+\big\{\mu_{2,i}-\mathrm{E}(\mu_2\mid R_i,\boldsymbol{X}_i)\big\}\Big]
\end{aligned}
$$

$$
\begin{aligned}
&+ \frac{1}{n^{1/2}} \sum_{i=1}^{n} \Big[(\boldsymbol{\gamma}_{0X}^{\mathrm{T}} + \widetilde{\boldsymbol{q}}_{\alpha}^{\mathrm{T}}) \big\{ \mathrm{E}(\mu_1 \mid R_i, \boldsymbol{X}_i) \boldsymbol{X}_i - \mathrm{E}(\mu_1 \boldsymbol{X} \mid R_i) \big\} \\
&+ \big\{ \mathrm{E}(\mu_1 \mid R_i, \boldsymbol{X}_i) - \mathrm{E}(\mu_1 \mid R_i) \big\} \boldsymbol{\gamma}_{0A,\mathcal{K}}^{\mathrm{T}} \boldsymbol{A}_{\mathcal{K},i} \\
&+ \big\{ \mathrm{E}(\mu_1 \mid R_i, \boldsymbol{X}_i) - \boldsymbol{I}_{\beta\gamma}^{\mathrm{T}} \boldsymbol{I}_{\gamma\gamma}^{-1} \boldsymbol{W}_{\mathcal{K},i} \big\} R_i (S_i - \boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}} \boldsymbol{W}_{\mathcal{K},i}) + \big\{ \mathrm{E}(\mu_2 \mid R_i, \boldsymbol{X}_i) - \mathrm{E}(\mu_2 \mid R_i) \big\} \Big] \\
&+ \frac{1}{n^{1/2}} \sum_{i=1}^{n} \Big\{ (\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}} + \widetilde{\boldsymbol{q}}_{\alpha}^{\mathrm{T}}) \mathrm{E}(\mu_1 \boldsymbol{X} \mid R_i) + \mathrm{E}(\mu_1 \mid R_i) \boldsymbol{\gamma}_{0A,\mathcal{K}}^{\mathrm{T}} \boldsymbol{A}_{\mathcal{K},i} + \mathrm{E}(\mu_2 \mid R_i) \Big\} \\
\equiv\ & \frac{1}{n^{1/2}} \sum_{i=1}^{n} U_{1i} + \frac{1}{n^{1/2}} \sum_{i=1}^{n} U_{2i} + \frac{1}{n^{1/2}} \sum_{i=1}^{n} U_{3i},
\end{aligned}
$$

where $\boldsymbol{\gamma}_{0A,\mathcal{K}}$ is the subvector of $\boldsymbol{\gamma}_{0\mathcal{K}}$ that corresponds to the selected components of $\boldsymbol{A}$. Note that U_{1i}, U_{2i} and U_{3i} generally depend on the selected model $\mathcal{K}$.

We now reintroduce the index $j = 1, \ldots, q$ for the transformation function. Let $U_{1i}^{(j)}$, $U_{2i}^{(j)}$ and $U_{3i}^{(j)}$ be U_{1i}, U_{2i} and U_{3i} computed under $G^{(j)}$, respectively. The score statistic under $\beta^{(j)} = 0$ and $G^{(j)}$ can be written as

$$
U_{\beta}^{(j)}(\widehat{\boldsymbol{\alpha}}^{(j)}, \widehat{\Lambda}^{(j)}, \widehat{\boldsymbol{\gamma}}_{\mathcal{K}}) = \frac{1}{n^{1/2}} \sum_{i=1}^{n} U_{1i}^{(j)} + \frac{1}{n^{1/2}} \sum_{i=1}^{n} U_{2i}^{(j)} + \frac{1}{n^{1/2}} \sum_{i=1}^{n} U_{3i}^{(j)} + o_p(1).
$$

Let $\boldsymbol{U}_{ki} = \big(U_{ki}^{(1)}, \ldots, U_{ki}^{(q)}\big)^{\mathrm{T}}$ for $k = 1, 2$ and 3. For $j, l = 1, \ldots, q$ and $k = 1, 2$ and 3, define $\sigma_k^{(jl)}(\mathcal{K})^2 = \mathrm{Cov}\big(U_{k1}^{(j)}, U_{k1}^{(l)}\big)$. Let $\boldsymbol{\Sigma}_k(\mathcal{K}) = \big(\sigma_k^{(jl)}(\mathcal{K})^2\big)_{j,l=1,\ldots,q}$ for $k = 1, 2$ and 3, and $\boldsymbol{\Sigma}(\mathcal{K}) = \sum_{k=1}^{3} \boldsymbol{\Sigma}_k(\mathcal{K})$.

By the Cramer–Wold device, it suffices to show that $\boldsymbol{t}^{\mathrm{T}} \boldsymbol{\Sigma}(\mathcal{K}^*)^{-1/2} \widehat{\boldsymbol{U}}_{\beta}(\mathcal{K}^*)$ converges to a standard normal distribution for any vector $\boldsymbol{t} \in \mathbb{R}^q$ with $\|\boldsymbol{t}\| = 1$. By a version of the portmanteau theorem (Pollard, 2002, p.177), it suffices to show that for any $g \in \mathcal{C}_{\mathrm{B}}^3$,

$$
\mathrm{E}\Big[g\Big\{ \boldsymbol{t}^{\mathrm{T}} \boldsymbol{\Sigma}(\mathcal{K}^*)^{-1/2} \widehat{\boldsymbol{U}}_{\beta}(\mathcal{K}^*) \Big\} \Big] \to \mathrm{E}\{g(Z)\}, \tag{S2}
$$

where Z is a standard normal random variable. Based on the above results and the mean-value theorem,

$$\begin{aligned}
&\mathrm{E}\Big[g\Big\{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K}^*)^{-1/2}\widehat{\boldsymbol{U}}_{\beta}(\mathcal{K}^*)\Big\}\Big] \\
&= \int \mathrm{E}\Big[g\Big\{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}\widehat{\boldsymbol{U}}_{\beta}(\mathcal{K})\Big\} \mid \mathcal{K}^* = \mathcal{K}\Big]\,\mathrm{d}\mathbb{P}_{\mathcal{K}^*}(\mathcal{K}) \\
&= \int_{\mathcal{K}\in\Omega_n} \mathrm{E}\Big[g\Big\{\frac{1}{n^{1/2}}\sum_{i=1}^{n}\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}(\boldsymbol{U}_{1i}+\boldsymbol{U}_{2i}+\boldsymbol{U}_{3i})\Big\} \mid \mathcal{K}^* = \mathcal{K}\Big]\,\mathrm{d}\mathbb{P}_{\mathcal{K}^*}(\mathcal{K}) + o(1), \quad \text{(S3)}
\end{aligned}$$

where $\mathbb{P}_{\mathcal{K}^*}$ is the probability measure of $\mathcal{K}^*$.

For $i = 1, \ldots, n$, let

$$\widetilde{\boldsymbol{U}}_{1i} = \mathrm{Var}\big(\boldsymbol{U}_1 \mid R_i, \boldsymbol{X}_i, S_i, \boldsymbol{A}_i\big)^{1/2}\boldsymbol{Z}_{1i},$$

where $\boldsymbol{Z}_{11}, \ldots, \boldsymbol{Z}_{1n}$ are i.i.d. standard multivariate normal variables that are independent of the observed data. Let $\boldsymbol{V}_{1i} = \widetilde{\boldsymbol{U}}_{11} + \cdots + \widetilde{\boldsymbol{U}}_{1,i-1} + \boldsymbol{U}_{1,i+1} + \cdots + \boldsymbol{U}_{1n}$ for $i = 1, \ldots, n$. Note that

$$\begin{aligned}
&\mathrm{E}\Big[g\Big\{\frac{1}{n^{1/2}}\sum_{i=1}^{n}\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}(\boldsymbol{U}_{1i}+\boldsymbol{U}_{2i}+\boldsymbol{U}_{3i})\Big\} \\
&\qquad - g\Big\{\frac{1}{n^{1/2}}\sum_{i=1}^{n}\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}(\widetilde{\boldsymbol{U}}_{1i}+\boldsymbol{U}_{2i}+\boldsymbol{U}_{3i})\Big\} \mid \mathcal{K}^* = \mathcal{K}\Big] \\
&= \sum_{i=1}^{n}\mathrm{E}\Big[g\Big\{\frac{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}}{n^{1/2}}\Big(\boldsymbol{V}_{1i} + \sum_j \boldsymbol{U}_{2j} + \sum_j \boldsymbol{U}_{3j} + \boldsymbol{U}_{1i}\Big)\Big\} \\
&\qquad\qquad - g\Big\{\frac{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}}{n^{1/2}}\Big(\boldsymbol{V}_{1i} + \sum_j \boldsymbol{U}_{2j} + \sum_j \boldsymbol{U}_{3j} + \widetilde{\boldsymbol{U}}_{1i}\Big)\Big\} \mid \mathcal{K}^* = \mathcal{K}\Big] \\
&= \frac{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}}{n^{1/2}}\sum_{i=1}^{n}\mathrm{E}\Big[g'\Big\{\frac{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}}{n^{1/2}}\Big(\boldsymbol{V}_{1i} + \sum_j \boldsymbol{U}_{2j} + \sum_j \boldsymbol{U}_{3j}\Big)\Big\}(\boldsymbol{U}_{1i} - \widetilde{\boldsymbol{U}}_{1i}) \mid \mathcal{K}^* = \mathcal{K}\Big] \\
&\quad + \frac{1}{2n}\sum_{i=1}^{n}\mathrm{E}\Big(g''\Big\{\frac{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}}{n^{1/2}}\Big(\boldsymbol{V}_{1i} + \sum_j \boldsymbol{U}_{2j} + \sum_j \boldsymbol{U}_{3j}\Big)\Big\} \\
&\qquad\qquad\qquad \times \big[\{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}\boldsymbol{U}_{1i}\}^2 - \{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}\widetilde{\boldsymbol{U}}_{1i}\}^2\big] \mid \mathcal{K}^* = \mathcal{K}\Big) \\
&\quad + \frac{1}{6n^{3/2}}\sum_{i=1}^{n}\mathrm{E}\Big[g'''(a)\{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}\boldsymbol{U}_{1i}\}^3 - g'''(\widetilde{a})\{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}\widetilde{\boldsymbol{U}}_{1i}\}^3 \mid \mathcal{K}^* = \mathcal{K}\Big] \qquad \text{(S4)}
\end{aligned}$$

for some variables a and $\widetilde{a}$. By construction, $\boldsymbol{U}_{1i}$ and $\widetilde{\boldsymbol{U}}_{1i}$ are independent of $\boldsymbol{V}_{1i}$ and $\sum_j(\boldsymbol{U}_{2j}+\boldsymbol{U}_{3j})$ given $\mathcal{O}_1 \equiv (R_i, S_i, \boldsymbol{X}_i, \boldsymbol{A}_i)_{i=1,\ldots,n}$. The expectation in the first term on the right-hand side of (S4) is

$$\mathrm{E}\left(\mathrm{E}\left[g'\left\{\frac{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}}{n^{1/2}}\left(\boldsymbol{V}_{1i}+\sum_j \boldsymbol{U}_{2j}+\sum_j \boldsymbol{U}_{3j}\right)\right\} \mid \mathcal{O}_1, \mathcal{K}^*=\mathcal{K}\right]\right.$$
$$\left.\times\, \mathrm{E}(\boldsymbol{U}_{1i}-\widetilde{\boldsymbol{U}}_{1i} \mid \mathcal{O}_1, \mathcal{K}^*=\mathcal{K}) \mid \mathcal{K}^*=\mathcal{K}\right) = \boldsymbol{0},$$

because $\mathrm{E}(\boldsymbol{U}_{1i}-\widetilde{\boldsymbol{U}}_{1i} \mid \mathcal{O}_1, \mathcal{K}^*=\mathcal{K}) = \mathrm{E}(\boldsymbol{U}_{1i}-\widetilde{\boldsymbol{U}}_{1i} \mid \mathcal{O}_1) = \boldsymbol{0}$. Likewise, the second term on the right-hand side of (S4) is 0, because the conditional second moments of $\boldsymbol{U}_{1i}$ and $\widetilde{\boldsymbol{U}}_{1i}$ given $\mathcal{O}_1$ match $(i=1,\ldots,n)$. For $\mathcal{K}\in\Omega_n$, the right-hand side of (S4) is bounded above by

$$\sum_{j=1}^q \zeta_{1n}^{(j)} \equiv M\sum_{j=1}^q n^{-3/2}\sum_{i=1}^n \mathrm{E}\left\{\sup_{\mathcal{K}\in\Omega_n}\left(|U_{1i}^{(j)}|^3+|\widetilde{U}_{1i}^{(j)}|^3\right) \mid \mathcal{K}^*=\mathcal{K}\right\}$$

for some positive constant M. By Lemma S5 in Section S3, $\int_{\Omega_n}\zeta_{1n}^{(j)}\,\mathrm{d}\mathbb{P}_{\mathcal{K}^*} \to 0$ for $j=1,\ldots,q$. Therefore, the $\boldsymbol{U}_{1i}$'s in (S3) can be replaced by $\widetilde{\boldsymbol{U}}_{1i}$'s. Furthermore, note that $n^{-1/2}\sum_{i=1}^n \widetilde{\boldsymbol{U}}_{1i}$ can be written as $\widehat{\boldsymbol{\Sigma}}_1(\mathcal{K})^{1/2}\boldsymbol{Z}_1$, where $\boldsymbol{Z}_1$ is a standard multivariate normal random variable independent of the observed data, $\widehat{\boldsymbol{\Sigma}}_1(\mathcal{K}) = \left(\widehat{\sigma}_1^{(jl)}(\mathcal{K})^2\right)_{j,l=1,\ldots,q}$, and $\widehat{\sigma}_1^{(jl)}(\mathcal{K})^2$ is the empirical counterpart of $\sigma_1^{(jl)}(\mathcal{K})^2$; the precise definition of $\widehat{\sigma}_1^{(jl)}(\mathcal{K})^2$ is given in Section S3. By linear expansion, we conclude that the right-hand side of (S3) is equal to

$$\int_{\mathcal{K}\in\Omega_n}\mathrm{E}\left[g\left\{\frac{1}{n^{1/2}}\sum_{i=1}^n \boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}(n^{1/2}\boldsymbol{\Sigma}_1(\mathcal{K})^{1/2}\boldsymbol{Z}_1+\boldsymbol{U}_{2i}+\boldsymbol{U}_{3i})\right\} \mid \mathcal{K}^*=\mathcal{K}\right]\mathrm{d}\mathbb{P}_{\mathcal{K}^*}(\mathcal{K})+o(1), \tag{S5}$$

up to an additive term bounded above by

$$\sup\|\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}g'\|_2\mathrm{E}\left[\sup_{\mathcal{K}\in\Omega_n}\left\|\left\{\widehat{\boldsymbol{\Sigma}}_1(\mathcal{K})^{1/2}-\boldsymbol{\Sigma}_1(\mathcal{K})^{1/2}\right\}\boldsymbol{Z}_1\right\|_2\right],$$

which tends to 0 by Lemma S3 in Section S3.

Next, we show that $\boldsymbol{U}_{2i}$'s in (S5) can be similarly replaced by normal random variables. Define

$$\widetilde{\boldsymbol{U}}_{2i} = \mathrm{Var}(\boldsymbol{U}_2 \mid R_i, \boldsymbol{A}_i, S_i - \boldsymbol{\gamma}_{0X}^{\mathrm{T}} \boldsymbol{X}_i)^{1/2} \boldsymbol{Z}_{2i}$$

for $i = 1, \ldots, n$, where $\boldsymbol{Z}_{21}, \ldots, \boldsymbol{Z}_{2n}$ are i.i.d. standard multivariate normal random variables that are independent of the observed data and $\boldsymbol{Z}_{11}, \ldots, \boldsymbol{Z}_{1n}$. Note that the above conditional variance is taken with respect to $\boldsymbol{X}$. We wish to show that

$$\begin{aligned}\int_{\Omega_n} \mathrm{E}\Big[g\Big\{ \sum_{i=1}^{n} \frac{\boldsymbol{t}^{\mathrm{T}} \boldsymbol{\Sigma}(\mathcal{K})^{-1/2}}{n^{1/2}} (n^{1/2} \boldsymbol{\Sigma}_1(\mathcal{K})^{1/2} \boldsymbol{Z}_1 + \boldsymbol{U}_{2i} + \boldsymbol{U}_{3i}) \Big\} \\ - g\Big\{ \sum_{i=1}^{n} \frac{\boldsymbol{t}^{\mathrm{T}} \boldsymbol{\Sigma}(\mathcal{K})^{-1/2}}{n^{1/2}} (n^{1/2} \boldsymbol{\Sigma}_1(\mathcal{K})^{1/2} \boldsymbol{Z}_1 + \widetilde{\boldsymbol{U}}_{2i} + \boldsymbol{U}_{3i}) \Big\} \mid \mathcal{K}^* = \mathcal{K} \Big] \, \mathrm{d}\mathbb{P}_{\mathcal{K}^*}(\mathcal{K}) = o(1).\end{aligned} \tag{S6}$$

Following the arguments in the proof of Theorem 1 in Wong and Feng (2023), the event $\{\mathcal{K}^* = \mathcal{K}\}$ in the conditional expectation in (S6) can be replaced by $\{\mathcal{K}_0^* = \mathcal{K}\}$. Let $\boldsymbol{V}_{2i} = \widetilde{\boldsymbol{U}}_{21} + \cdots + \widetilde{\boldsymbol{U}}_{2,i-1} + \boldsymbol{U}_{2,i+1} + \cdots + \boldsymbol{U}_{2n}$ for $i = 1, \ldots, n$. The term inside the integration of the left-hand side of (S6) is up to a vanishing term equal to

$$\begin{aligned}&\sum_{i=1}^{n} \mathrm{E}\Big[g\Big\{ \frac{\boldsymbol{t}^{\mathrm{T}} \boldsymbol{\Sigma}(\mathcal{K})^{-1/2}}{n^{1/2}} \Big(\boldsymbol{V}_{2i} + n^{1/2} \boldsymbol{\Sigma}_1(\mathcal{K})^{1/2} \boldsymbol{Z}_1 + \boldsymbol{U}_{2i} + \sum_j \boldsymbol{U}_{3j} \Big) \Big\} \\ &\qquad - g\Big\{ \frac{\boldsymbol{t}^{\mathrm{T}} \boldsymbol{\Sigma}(\mathcal{K})^{-1/2}}{n^{1/2}} \Big(\boldsymbol{V}_{2i} + n^{1/2} \boldsymbol{\Sigma}_1(\mathcal{K})^{1/2} \boldsymbol{Z}_1 + \widetilde{\boldsymbol{U}}_{2i} + \sum_j \boldsymbol{U}_{3j} \Big) \Big\} \mid \mathcal{K}_0^* = \mathcal{K} \Big] \\ &= \frac{\boldsymbol{t}^{\mathrm{T}} \boldsymbol{\Sigma}(\mathcal{K})^{-1/2}}{n^{1/2}} \sum_{i=1}^{n} \mathrm{E}\Big[g'\Big\{ \frac{\boldsymbol{t}^{\mathrm{T}} \boldsymbol{\Sigma}(\mathcal{K})^{-1/2}}{n^{1/2}} \Big(\boldsymbol{V}_{2i} + n^{1/2} \boldsymbol{\Sigma}_1(\mathcal{K})^{1/2} \boldsymbol{Z}_1 + \sum_j \boldsymbol{U}_{3j} \Big) \Big\} (\boldsymbol{U}_{2i} - \widetilde{\boldsymbol{U}}_{2i}) \mid \mathcal{K}_0^* = \mathcal{K} \Big] \\ &\quad + \frac{1}{2n} \sum_{i=1}^{n} \mathrm{E}\Big(g''\Big\{ \frac{\boldsymbol{t}^{\mathrm{T}} \boldsymbol{\Sigma}(\mathcal{K})^{-1/2}}{n^{1/2}} \Big(\boldsymbol{V}_{2i} + n^{1/2} \boldsymbol{\Sigma}_1(\mathcal{K})^{1/2} \boldsymbol{Z}_1 + \sum_j \boldsymbol{U}_{3j} \Big) \Big\} \\ &\qquad\qquad \times \big[\{ \boldsymbol{t}^{\mathrm{T}} \boldsymbol{\Sigma}(\mathcal{K})^{-1/2} \boldsymbol{U}_{2i} \}^2 - \{ \boldsymbol{t}^{\mathrm{T}} \boldsymbol{\Sigma}(\mathcal{K})^{-1/2} \widetilde{\boldsymbol{U}}_{2i} \}^2 \big] \mid \mathcal{K}_0^* = \mathcal{K} \Big)\end{aligned}$$

$$+\frac{1}{6n^{3/2}}\sum_{i=1}^{n}\mathrm{E}\Big[g'''(b)\{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}\boldsymbol{U}_{2i}\}^{3}-g'''(\widetilde{b})\{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}\widetilde{\boldsymbol{U}}_{2i}\}^{3}\mid\mathcal{K}_0^*=\mathcal{K}\Big] \tag{S7}$$

for some variables b and $\widetilde{b}$. Let $\mathcal{O}_2=(R_i,\boldsymbol{A}_i,S_i-\boldsymbol{\gamma}_{0X}^{\mathrm{T}}\boldsymbol{X}_i)_{i=1,\dots,n}$. Since the event $\{\mathcal{K}_0^*=\mathcal{K}\}$ is implied by $\mathcal{O}_2$, we have

$$\begin{aligned}
&\mathrm{E}\Big[g'\Big\{\frac{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}}{n^{1/2}}\Big(\boldsymbol{V}_{2i}+n^{1/2}\boldsymbol{\Sigma}_1(\mathcal{K})^{1/2}\boldsymbol{Z}_1+\sum_j\boldsymbol{U}_{3j}\Big)\Big\}(\boldsymbol{U}_{2i}-\widetilde{\boldsymbol{U}}_{2i})\mid\mathcal{K}_0^*=\mathcal{K}\Big]\\
&=\mathrm{E}\Big(\mathrm{E}\Big[g'\Big\{\frac{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}}{n^{1/2}}\Big(\boldsymbol{V}_{2i}+n^{1/2}\boldsymbol{\Sigma}_1(\mathcal{K})^{1/2}\boldsymbol{Z}_1+\sum_j\boldsymbol{U}_{3j}\Big)\Big\}\mid\mathcal{O}_2,\mathcal{K}_0^*=\mathcal{K}\Big]\\
&\qquad\times\mathrm{E}(\boldsymbol{U}_{2i}-\widetilde{\boldsymbol{U}}_{2i}\mid\mathcal{O}_2)\mid\mathcal{K}_0^*=\mathcal{K}\Big)=\boldsymbol{0}.
\end{aligned}$$

Likewise, the second term on the right-hand side of (S7) is 0 because the conditional second moments of $\boldsymbol{U}_{2i}$ and $\widetilde{\boldsymbol{U}}_{2i}$ match. By Lemma S5 in Section S3, the third term on the right-hand side of (S7) is bounded by a variable $\sum_{j=1}^{q}\zeta_{2n}^{(j)}$ such that $\int_{\Omega_n}\zeta_{2n}^{(j)}\,\mathrm{d}\mathbb{P}_{\mathcal{K}^*}\to 0$, so (S6) holds. Furthermore, we can show that $\widetilde{\boldsymbol{U}}_{2i}$ in (S6) can be replaced by $n^{1/2}\boldsymbol{\Sigma}_2(\mathcal{K})^{1/2}\boldsymbol{Z}_2$ for a standard multivariate normal random vector independent of the observed data and $\boldsymbol{Z}_1$.

Let $\widetilde{\boldsymbol{U}}_{3i}=\mathrm{Var}(\boldsymbol{U}_3\mid\boldsymbol{A}_{\mathcal{K},i})^{1/2}\boldsymbol{Z}_{3i}$ for $i=1,\dots,n$, where $\boldsymbol{Z}_{31},\dots,\boldsymbol{Z}_{3n}$ are i.i.d. standard multivariate normal variables that are independent of the observed data and $(\boldsymbol{Z}_{1i},\boldsymbol{Z}_{2i})_{i=1,\dots,n}$. By arguments similar to the proof of Theorem 1 in Wong and Feng (2023), we can show that

$$\begin{aligned}
&\int_{\Omega_n}\mathrm{E}\Big[g\Big\{\sum_{i=1}^{n}\frac{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}}{n^{1/2}}(n^{1/2}\boldsymbol{\Sigma}_1(\mathcal{K})^{1/2}\boldsymbol{Z}_1+n^{1/2}\boldsymbol{\Sigma}_2(\mathcal{K})^{1/2}\boldsymbol{Z}_2+\boldsymbol{U}_{3i})\Big\}\\
&\quad-g\Big\{\sum_{i=1}^{n}\frac{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}}{n^{1/2}}(n^{1/2}\boldsymbol{\Sigma}_1(\mathcal{K})^{1/2}\boldsymbol{Z}_1+n^{1/2}\boldsymbol{\Sigma}_2(\mathcal{K})^{1/2}\boldsymbol{Z}_2+\widetilde{\boldsymbol{U}}_{3i})\Big\}\mid\mathcal{K}^*=\mathcal{K}\Big]\,\mathrm{d}\mathbb{P}_{\mathcal{K}^*}(\mathcal{K})
\end{aligned}$$

tends to 0. Combining the above results and applying Lemma S3 again, we have

$$\mathrm{E}\Big[g\Big\{\sum_{i=1}^{n}\frac{\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}}{n^{1/2}}(\boldsymbol{U}_{1i}+\boldsymbol{U}_{2i}+\boldsymbol{U}_{3i})\Big\}\mid\mathcal{K}^*=\mathcal{K}\Big]$$

$$= \mathrm{E}\Big(g\Big[\boldsymbol{t}^{\mathrm{T}}\boldsymbol{\Sigma}(\mathcal{K})^{-1/2}\big\{\boldsymbol{\Sigma}_1(\mathcal{K})^{1/2}\boldsymbol{Z}_1 + \boldsymbol{\Sigma}_2(\mathcal{K})^{1/2}\boldsymbol{Z}_2 + \boldsymbol{\Sigma}_3(\mathcal{K})^{1/2}\boldsymbol{Z}_3\big\}\Big]\Big) + o(1)$$

uniformly over $\mathcal{K} \in \Omega_n$, where $\boldsymbol{Z}_3$ is standard multivariate normal random vector independent of the observed data and $(\boldsymbol{Z}_1, \boldsymbol{Z}_2)$. Because $\boldsymbol{\Sigma}_1(\mathcal{K})^{1/2}\boldsymbol{Z}_1 + \boldsymbol{\Sigma}_2(\mathcal{K})^{1/2}\boldsymbol{Z}_2 + \boldsymbol{\Sigma}_3(\mathcal{K})^{1/2}\boldsymbol{Z}_3$ is multivariate normal with mean $\mathbf{0}$ and variance $\boldsymbol{\Sigma}(\mathcal{K})$, the desired convergence (S2) follows. □

S2 Proof of Theorem 2

Proof of Theorem 2. We establish the consistency of the proposed variance estimator under the case with a single transformation function G; the case with multiple transformation functions is analogous. Following the arguments in the proof of Theorem 1, for any fixed $\mathcal{K} \in \Omega_n$, the Taylor series expansion of $U_\beta(\widehat{\boldsymbol{\alpha}}, \widehat{\Lambda}, \widehat{\boldsymbol{\gamma}}_{\mathcal{K}})$ at $(\boldsymbol{\alpha}_0, \Lambda_0, \boldsymbol{\gamma}_{0\mathcal{K}})$ yields

$$\begin{aligned} U_\beta(\widehat{\boldsymbol{\alpha}}, \widehat{\Lambda}, \widehat{\boldsymbol{\gamma}}_{\mathcal{K}}) &= \frac{1}{n^{1/2}}\sum_{i=1}^n \Big[\mu_{1,i}(\boldsymbol{\alpha}_0, \Lambda_0)\big\{R_i S_i + (1-R_i)\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i}\big\} - \ell_{\alpha\Lambda,i}(\boldsymbol{\alpha}_0, \Lambda_0)[\widetilde{\boldsymbol{q}}_\alpha, \widetilde{q}_\Lambda] \\ &\qquad - \boldsymbol{I}_{\beta\gamma}^{\mathrm{T}}\boldsymbol{I}_{\gamma\gamma}^{-1}\boldsymbol{W}_{\mathcal{K},i}R_i(S_i - \boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i})\Big] + o_p(1) \\ &\equiv \frac{1}{n^{1/2}}\sum_{i=1}^n \widehat{\sigma}_{0i}(\mathcal{K}) + o_p(1). \end{aligned}$$

Note that the proposed variance estimator for the score statistic, $n^{-1}\sum_{i=1}^n \big\{\widehat{\sigma}_i(\mathcal{K}) - \overline{\sigma}(\mathcal{K})\big\}^2$, is equal to

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^n \big\{\widehat{\sigma}_i(\mathcal{K}) - \widehat{\sigma}_{0i}(\mathcal{K}) + \widehat{\sigma}_{0i}(\mathcal{K})\big\}^2 - \overline{\sigma}(\mathcal{K})^2 \\ &= \frac{1}{n}\sum_{i=1}^n \widehat{\sigma}_{0i}(\mathcal{K})^2 + \frac{1}{n}\sum_{i=1}^n \big\{\widehat{\sigma}_i(\mathcal{K}) - \widehat{\sigma}_{0i}(\mathcal{K})\big\}^2 + \frac{2}{n}\sum_{i=1}^n \big\{\widehat{\sigma}_i(\mathcal{K}) - \widehat{\sigma}_{0i}(\mathcal{K})\big\}\widehat{\sigma}_{0i}(\mathcal{K}) - \overline{\sigma}(\mathcal{K})^2. \end{aligned} \tag{S8}$$

To prove the convergence of the proposed variance estimator, we first show that $n^{-1}\sum_{i=1}^{n}\big\{\widehat{\sigma}_i(\mathcal{K})-\widehat{\sigma}_{0i}(\mathcal{K})\big\}^2 = o_p(1)$ uniformly over $\mathcal{K}\in\Omega_n$. Note that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\big\{\widehat{\sigma}_i(\mathcal{K})-\widehat{\sigma}_{0i}(\mathcal{K})\big\}^2\\
\lesssim&\frac{1}{n}\sum_{i=1}^{n}\big[\mu_{1,i}(\widehat{\boldsymbol{\zeta}})\big\{R_iS_i+(1-R_i)\widehat{\boldsymbol{\gamma}}_{\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K}}\big\}-\mu_{1,i}(\boldsymbol{\alpha}_0,\Lambda_0)\big\{R_iS_i+(1-R_i)\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i}\big\}\big]^2\\
&+\frac{1}{n}\sum_{i=1}^{n}\big\{\widehat{\boldsymbol{I}}_{\beta\gamma}^{\mathrm{T}}\widehat{\boldsymbol{I}}_{\gamma\gamma}^{-1}\boldsymbol{W}_{\mathcal{K},i}R_i(S_i-\widehat{\boldsymbol{\gamma}}_{\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i})-\boldsymbol{I}_{\beta\gamma}^{\mathrm{T}}\boldsymbol{I}_{\gamma\gamma}^{-1}\boldsymbol{W}_{\mathcal{K},i}R_i(S_i-\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i})\big\}^2\\
&+\frac{1}{n}\sum_{i=1}^{n}\big\{\widehat{\boldsymbol{I}}_{\beta\zeta}^{\mathrm{T}}\widehat{\boldsymbol{I}}_{\zeta\zeta}^{-1}\widehat{\boldsymbol{U}}_{\zeta,i}-\ell_{\alpha\Lambda,i}(\boldsymbol{\alpha}_0,\Lambda_0)[\widetilde{\boldsymbol{q}}_\alpha,\widetilde{q}_\Lambda]\big\}^2\\
\equiv& A_1+A_2+A_3. \qquad\qquad \text{(S9)}
\end{aligned}$$

Using the arguments in the proof of Theorem 2 of Wong and Feng (2023), we can show that A_1 and A_2 converge in mean to zero uniformly over $\mathcal{K}\in\Omega_n$. It remains to show that A_3 converges in mean to zero. Note that A_3 consists of the difference between the derivative of the log-likelihood with respect to $(\boldsymbol{\alpha},\lambda_1,\ldots,\lambda_m)$ and the derivative of the log-likelihood with respect to $(\boldsymbol{\alpha},\Lambda)$ (along some direction). Let $\boldsymbol{\nu}_\alpha$ and $\boldsymbol{\nu}_\lambda$ be the first $\|\boldsymbol{\alpha}\|_0$ and the last m components of $\widehat{\boldsymbol{I}}_{\zeta\zeta}^{-1}\widehat{\boldsymbol{I}}_{\beta\zeta}$, respectively, and partition $\widehat{\boldsymbol{U}}_{\zeta,i}=(\widehat{\boldsymbol{U}}_{\alpha,i}^{\mathrm{T}},\widehat{\boldsymbol{U}}_{\lambda,i}^{\mathrm{T}})^{\mathrm{T}}$ correspondingly. Note that

$$\boldsymbol{\nu}_\alpha^{\mathrm{T}}\widehat{\boldsymbol{U}}_{\alpha,i}=\big\{\Delta_i\psi_i(\widehat{\boldsymbol{\zeta}})\xi_i(\widehat{\boldsymbol{\zeta}})-G_i'(\widehat{\boldsymbol{\zeta}})\xi_i(\widehat{\boldsymbol{\zeta}})\big\}\boldsymbol{\nu}_\alpha^{\mathrm{T}}\boldsymbol{X}_i,$$

and

$$\boldsymbol{\nu}_\lambda^{\mathrm{T}}\widehat{\boldsymbol{U}}_{\lambda,i}=\frac{\nu_{\lambda,k(i)}\Delta_i}{\widehat{\lambda}_{k(i)}}+\sum_{k=1}^{m}I(Y_i\geq t_k)\nu_{\lambda,k}\big\{\Delta_i\psi_i(\widehat{\boldsymbol{\zeta}})-G_i'(\widehat{\boldsymbol{\zeta}})\big\}\exp(\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\boldsymbol{X}_i),$$

where $\nu_{\lambda,j}$ is the jth term of $\boldsymbol{\nu}_\lambda$. Define a step function a_Λ that jumps at $t_1,\ldots,t_m$, with $a_\Lambda(t_j)=\nu_{\lambda,j}/\widehat{\lambda}_j$ for $j=1,\ldots,m$. We have that

$$\widehat{\boldsymbol{I}}_{\beta\zeta}^{\mathrm{T}}\widehat{\boldsymbol{I}}_{\zeta\zeta}^{-1}\widehat{\boldsymbol{U}}_{\zeta,i}=\boldsymbol{\nu}_\alpha^{\mathrm{T}}\widehat{\boldsymbol{U}}_{\alpha,i}+\boldsymbol{\nu}_\lambda^{\mathrm{T}}\widehat{\boldsymbol{U}}_{\lambda,i}=\ell_{\alpha\Lambda,i}(\widehat{\boldsymbol{\alpha}},\widehat{\Lambda})[\boldsymbol{\nu}_\alpha,a_\Lambda].$$

Recall that $(\boldsymbol{W}_\alpha, W_\Lambda)^{-1}$ is a continuous linear map from $\mathcal{H}$ to $\mathcal{H}$. Thus, A_3 can be written as

$$
\begin{aligned}
&\mathbb{P}_n\big\{\ell_{\alpha\Lambda}(\widehat{\boldsymbol{\alpha}},\widehat{\Lambda})[\widetilde{\boldsymbol{q}}_\alpha,\widetilde{q}_\Lambda] - \ell_{\alpha\Lambda}(\widehat{\boldsymbol{\alpha}},\widehat{\Lambda})[\boldsymbol{\nu}_\alpha, a_\Lambda]\big\}^2 \\
&= \mathbb{P}_n\Big\{\ell_{\alpha\Lambda}(\widehat{\boldsymbol{\alpha}},\widehat{\Lambda})\big[(\boldsymbol{W}_\alpha, W_\Lambda)^{-1}\big\{\boldsymbol{W}_\alpha(\widetilde{\boldsymbol{q}}_\alpha,\widetilde{q}_\Lambda), W_\Lambda(\widetilde{\boldsymbol{q}}_\alpha,\widetilde{q}_\Lambda)\big\}\big] \\
&\qquad - \ell_{\alpha\Lambda}(\widehat{\boldsymbol{\alpha}},\widehat{\Lambda})\big[(\boldsymbol{W}_\alpha, W_\Lambda)^{-1}\big\{\boldsymbol{W}_\alpha(\boldsymbol{\nu}_\alpha,a_\Lambda), W_\Lambda(\boldsymbol{\nu}_\alpha,a_\Lambda)\big\}\big]\Big\}^2 \\
&= \mathbb{P}_n\Big[\ell_{\alpha\Lambda}(\widehat{\boldsymbol{\alpha}},\widehat{\Lambda})\big[(\boldsymbol{W}_\alpha, W_\Lambda)^{-1}\big\{(\boldsymbol{I}_{\beta\alpha}, I_{\beta\Lambda}) - (\boldsymbol{W}_\alpha(\boldsymbol{\nu}_\alpha,a_\Lambda), W_\Lambda(\boldsymbol{\nu}_\alpha,a_\Lambda))\big\}\big]\Big]^2. \qquad \text{(S10)}
\end{aligned}
$$

Consider the difference term on the right-hand side of the above:

$$
(\boldsymbol{I}_{\beta\alpha}, I_{\beta\Lambda}) - (\boldsymbol{W}_\alpha(\boldsymbol{\nu}_\alpha,a_\Lambda), W_\Lambda(\boldsymbol{\nu}_\alpha,a_\Lambda)). \qquad \text{(S11)}
$$

The first component of (S11) can be written as

$$
\boldsymbol{I}_{\beta\alpha} - \boldsymbol{W}_\alpha(\boldsymbol{\nu}_\alpha,a_\Lambda) = \boldsymbol{I}_{\beta\alpha} - \widehat{\boldsymbol{I}}_{\beta\alpha} + \widehat{\boldsymbol{I}}_{\alpha\alpha}\boldsymbol{\nu}_\alpha + \widehat{\boldsymbol{I}}_{\alpha\lambda}\boldsymbol{\nu}_\lambda - \boldsymbol{W}_\alpha(\boldsymbol{\nu}_\alpha,a_\Lambda).
$$

Since $\|\boldsymbol{I}_{\beta\alpha} - \widehat{\boldsymbol{I}}_{\beta\alpha}\|^2 = o_p(1)$ and $\big\|\widehat{\boldsymbol{I}}_{\alpha\alpha}\boldsymbol{\nu}_\alpha + \widehat{\boldsymbol{I}}_{\alpha\lambda}\boldsymbol{\nu}_\lambda - \boldsymbol{W}_\alpha(\boldsymbol{\nu}_\alpha,a_\Lambda)\big\|_2 = o_p(1)$, the L_2-norm of the left-hand side of the above display is $o_p(1)$.

The second component of (S11) can be written as

$$
I_{\beta\Lambda} - (\widehat{\boldsymbol{I}}_{\beta\lambda})(\cdot) + (\widehat{\boldsymbol{I}}_{\alpha\lambda}^{\mathrm{T}}\boldsymbol{\nu}_\alpha)(\cdot) + (\widehat{\boldsymbol{I}}_{\lambda\lambda}\boldsymbol{\nu}_\lambda)(\cdot) - W_\Lambda(\boldsymbol{\nu}_\alpha,a_\Lambda),
$$

where for an m-vector $\boldsymbol{a}$, $(\boldsymbol{a})(\cdot)$ denotes a step function that jumps at $t_1,\ldots,t_m$ with $(\boldsymbol{a})(t_j)$ equals the jth component of $\boldsymbol{a}$. By Lemma S2, $\|I_{\beta\Lambda} - (\widehat{\boldsymbol{I}}_{\beta\lambda})(\cdot)\|_2^2 \equiv \int_0^\tau\{I_{\beta\Lambda}(s) - (\widehat{\boldsymbol{I}}_{\beta\lambda})(s)\}^2\,\mathrm{d}s = o_p(1)$. Because the derivatives of the log-likelihood evaluated at $(\widehat{\boldsymbol{\alpha}},\widehat{\lambda}_1,\ldots,\widehat{\lambda}_m)$ equal 0, we have

$$
-\frac{1}{n\widehat{\lambda}_j} = \frac{1}{n}\sum_{i=1}^n I(Y_i \ge t_j)\big\{\Delta_i\psi_i(\widehat{\boldsymbol{\alpha}},\widehat{\Lambda}) - G_i'(\widehat{\boldsymbol{\alpha}},\widehat{\Lambda})\big\}\exp(\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\boldsymbol{X}_i) \qquad \text{for } j=1,\ldots,m.
$$

Therefore,

$$
\begin{aligned}
&(\widehat{\boldsymbol{I}}_{\lambda\lambda}\boldsymbol{\nu}_\lambda)(s) \\
=& -\sum_{j=1}^{m} I(t_{j-1} < s \le t_j)\Big[a_\Lambda(t_j)\frac{1}{n}\sum_{i=1}^{n} I(Y_i \ge t_j)\big\{\Delta_i\psi_i(\boldsymbol{\alpha}_0,\Lambda_0) - G_i'(\boldsymbol{\alpha}_0,\Lambda_0)\big\}\exp(\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X}_i) \\
&+ \frac{1}{n}\sum_{i=1}^{n}\big\{\Delta_i\eta_i(\boldsymbol{\alpha}_0,\Lambda_0) - G_i''(\boldsymbol{\alpha}_0,\Lambda_0)\big\}\exp(2\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X}_i)\int I\big(Y_i \ge \max(t,t_j)\big)a_\Lambda(t)\,\mathrm{d}\Lambda_0(t)\Big] \\
&+ o_p(1).
\end{aligned}
$$

Also,

$$
\begin{aligned}
&(\widehat{\boldsymbol{I}}_{\alpha\lambda}^{\mathrm{T}}\boldsymbol{\nu}_\alpha)(s) \\
=& -\sum_{j=1}^{m} I(t_{j-1} < s \le t_j)\Big[\frac{1}{n}\sum_{i=1}^{n} I(Y_i \ge t_j)\big\{\Delta_i\psi_i(\boldsymbol{\alpha}_0,\Lambda_0) - G_i'(\boldsymbol{\alpha}_0,\Lambda_0)\big\}\exp(\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X}_i)\boldsymbol{\nu}_\alpha^{\mathrm{T}}\boldsymbol{X}_i \\
&+ \frac{1}{n}\sum_{i=1}^{n} I(Y_i \ge t_j)\big\{\Delta_i\eta_i(\boldsymbol{\alpha}_0,\Lambda_0) - G_i''(\boldsymbol{\alpha}_0,\Lambda_0)\big\}\Lambda(Y_i)\exp(2\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X}_i)\boldsymbol{\nu}_\alpha^{\mathrm{T}}\boldsymbol{X}_i\Big] + o_p(1).
\end{aligned}
$$

We can see that $(\widehat{\boldsymbol{I}}_{\alpha\lambda}^{\mathrm{T}}\boldsymbol{\nu}_\alpha + \widehat{\boldsymbol{I}}_{\lambda\lambda}\boldsymbol{\nu}_\lambda)(\cdot)$ is (up to an $o_p(1)$ term) the empirical counterpart of $W_\Lambda(\boldsymbol{\nu}_\alpha, a_\Lambda)$, so by the Glivenko–Cantelli properties of the functions involved, $\|(\widehat{\boldsymbol{I}}_{\alpha\lambda}^{\mathrm{T}}\boldsymbol{\nu}_\alpha + \widehat{\boldsymbol{I}}_{\lambda\lambda}\boldsymbol{\nu}_\lambda)(\cdot) - W_\Lambda(\boldsymbol{\nu}_\alpha, a_\Lambda)\|_2 = o_p(1)$. Because the class of functions $\{\ell_{\alpha\Lambda}(\boldsymbol{\alpha},\Lambda)[\boldsymbol{h}_\alpha, h_\Lambda]^2\}$ is Glivenko–Cantelli and $(\boldsymbol{W}_\alpha, W_\Lambda)^{-1}$ is a bounded linear map, we conclude that the right-hand side of (S10) converges to 0 in probability.

Note that the fourth term on the right-hand side of (S8) converges to zero in mean uniformly over $\mathcal{K} \in \Omega_n$ following the proof of Theorem 2 in Wong and Feng (2023). Combining the above results, we have

$$
\frac{1}{n}\sum_{i=1}^{n}\big\{\widehat{\sigma}_i(\mathcal{K}) - \overline{\sigma}(\mathcal{K})\big\}^2 = \frac{1}{n}\sum_{i=1}^{n}\widehat{\sigma}_{0i}(\mathcal{K})^2 + o_p(1).
$$

Note that $\widehat{\sigma}_{0i}(\mathcal{K}) = \sum_{k=1}^{3} U_{ki}$. By arguments similar to the proof of Lemma S3, we have

$$\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{k=1}^{3} U_{ki}\right)^2 = \mathrm{Var}\left(\sum_{k=1}^{3} U_{k1}\right) + o_p(1) = \mathrm{Var}(U_{11}) + \mathrm{Var}(U_{21}) + \mathrm{Var}(U_{31}) + o_p(1),$$

where the $o_p(1)$ term converges in mean to 0 uniformly over $\mathcal{K} \in \Omega_n$. The desired result follows.

□

S3 Additional theoretical results

Lemma S2. *Under Conditions (C1)–(C3), and (C7), there exist positive constants C_1 and C_2 such that the inequalities*

$$\sup_{\mathcal{K}\in\Omega_n}\left\|\widehat{\boldsymbol{\gamma}}_{\mathcal{K}} - \boldsymbol{\gamma}_{0\mathcal{K}}\right\| > C_1\left\{\left(\frac{t + \log r_n + q_n}{n}\right)^{1/2} + \frac{q_n(\log n)^{2/\xi}(t + \log r_n + q_n)^{2/\xi}}{n}\right\},$$

$$\sup_{\mathcal{K}\in\Omega_n}\left\|(\widehat{\boldsymbol{I}}_{\gamma\gamma}^{(j)})^{-1}\widehat{\boldsymbol{I}}_{\beta\gamma}^{(j)} - (\boldsymbol{I}_{\gamma\gamma}^{(j)})^{-1}\boldsymbol{I}_{\beta\gamma}^{(j)}\right\| > C_1\left\{\left(\frac{t + \log r_n + q_n}{n}\right)^{1/2} + \frac{q_n(\log n)^{2/\xi}(t + \log r_n + q_n)^{2/\xi}}{n}\right\},$$

$$\sup_{\mathcal{K}\in\Omega_n}\left\|\widehat{\boldsymbol{I}}_{\beta\alpha}^{(j)} - \boldsymbol{I}_{\beta\alpha}^{(j)}\right\| > C_1\left\{\left(\frac{t + \log r_n}{n}\right)^{1/2} + \frac{q_n^{1/2}(\log n)^{1/\xi}(t + \log r_n)^{1/\min(1,\xi)}}{n}\right\}, \text{ and}$$

$$\sup_{\mathcal{K}\in\Omega_n}\sup_{0\le s\le\tau}\left|\widehat{I}_{\beta\Lambda}^{(j)}(s) - I_{\beta\Lambda}^{(j)}(s)\right| > C_1\left[\frac{(tq_n)^{1/2}}{n^{1/4}} + \left\{\frac{t + \log n + \log(2r_n)}{n}\right\}^{1/2} + \frac{q_n^{1/2}\{\log(2n)\}^{1/\xi}\{t + \log n + \log(2r_n)\}^{1/\min(1,\xi)}}{n}\right]$$

hold with probability at most $C_2\exp(-t)$ for $j = 1, \ldots, q$ and large enough n and t.

For $j, l = 1, \ldots, q$, define

$$\widehat{\sigma}_1^{(jl)}(\mathcal{K})^2 = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\big(U_{1i}^{(j)}U_{1i}^{(l)} \mid R_i, \boldsymbol{X}_i, S_i, \boldsymbol{A}_i\big)$$

$$\widehat{\sigma}_2^{(jl)}(\mathcal{K})^2 = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\big(U_{2i}^{(j)}U_{2i}^{(l)} \mid R_i, \boldsymbol{A}_i, S_i - \boldsymbol{\gamma}_{0X}^{\mathrm{T}}\boldsymbol{X}_i\big)$$

$$\widehat{\sigma}_3^{(jl)}(\mathcal{K})^2 = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\big(U_{3i}^{(j)}U_{3i}^{(l)} \mid \boldsymbol{A}_{\mathcal{K},i}\big).$$

Lemma S3. *Under Conditions (C1)–(C4), for $j, l = 1, \ldots, q$ and large enough n and t, there exist positive constants C_1 and C_2 such that*

$$P\left[\sup_{\mathcal{K}\in\Omega_n}\sum_{k=1}^{3}\left|\widehat{\sigma}_k^{(jl)}(\mathcal{K})^2-\sigma_k^{(jl)}(\mathcal{K})^2\right|>C_1\left\{\left(\frac{t+\log r_n}{n}\right)^{1/2}+\frac{q_n(\log n)^{2/\xi}(t+\log r_n)^{2/\xi}}{n}\right\}\right]$$

$\leq C_2\exp(-t)$.

Lemma S4. *Under Conditions (C1)–(C4), for $j = 1, \ldots, q$,*

$$\mathrm{E}\left[\sup_{\mathcal{K}\in\Omega_n}\left|\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left\{(\widehat{\boldsymbol{I}}_{\beta\gamma}^{(j)})^{\mathrm{T}}(\widehat{\boldsymbol{I}}_{\gamma\gamma}^{(j)})^{-1}-(\boldsymbol{I}_{\beta\gamma}^{(j)})^{\mathrm{T}}(\boldsymbol{I}_{\gamma\gamma}^{(j)})^{-1}\right\}\boldsymbol{W}_{\mathcal{K},i}R_i(S_i-\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i})\right|\right]=o(1).$$

Lemma S5. *Assume that Conditions (C1)–(C3) hold. For $U_{ki}^{(j)}$ and $\widetilde{U}_{ki}^{(j)}$ ($k = 1, 2, 3$; $i = 1, \ldots, n$; $j = 1, \ldots, q$) defined in the proof of Theorem 1, there exist positive constants C_1 and C_2 such that*

$$P\left[\sum_{k=1}^{3}\sup_{\mathcal{K}\in\Omega_n}\frac{1}{n^{3/2}}\sum_{i=1}^{n}\left(|U_{ki}^{(j)}|^3+|\widetilde{U}_{ki}^{(j)}|^3\right)>C_1\left\{\frac{(t+\log r_n)^{1/2}}{n}+\frac{q_n^{3/2}(\log n)^{3/\xi}(t+\log r_n)^{3/\xi}}{n^{3/2}}\right\}\right]$$

is smaller than $C_2\exp(-t)$ for large enough n and t.

We omit the proofs of Lemmas S3 and S4, which are analogous to the proofs of Lemmas S2 and S3 in Wong and Feng (2023).

Proof of Lemma S2. We refer the proofs of the first, second, and third results to the proof of Lemma S1 in Wong and Feng (2023). For the fourth result, we consider a single transformation function, and the same arguments can be extended to multiple transformation functions. Let $m \equiv \lfloor n^{1/2} \rfloor$ be the integer part of $n^{1/2}$ and $\zeta_m = \{0, \frac{\tau}{m}, \frac{2\tau}{m}, \ldots, \tau\}$. We have

$$\sup_{\mathcal{K}\in\Omega_n}\sup_{0\leq s\leq\tau}\left|\widehat{I}_{\beta\Lambda}(s)-I_{\beta\Lambda}(s)\right|\leq\sup_{\mathcal{K}\in\Omega_n}\sup_{s\in\zeta_m}\left|\widehat{I}_{\beta\Lambda}(s)-I_{\beta\Lambda}(s)\right|+\sup_{\mathcal{K}\in\Omega_n}\sup_{s,s':|s-s'|\leq\tau/m}\left|\widehat{I}_{\beta\Lambda}(s)-\widehat{I}_{\beta\Lambda}(s')\right|$$
$$+\sup_{\mathcal{K}\in\Omega_n}\sup_{s,s':|s-s'|\leq\tau/m}\left|I_{\beta\Lambda}(s)-I_{\beta\Lambda}(s')\right|. \quad \text{(S12)}$$

The second term on the right-hand side of (S12) can be written as

$$
\begin{aligned}
&\sup_{\mathcal{K}\in\Omega_n}\sup_{s,s':|s-s'|\le\tau/m}\Big|\frac{1}{n}\sum_{i=1}^{n}\widehat{I}_{\beta\Lambda,i}\big\{I(s\le Y_i)-I(s'\le Y_i)\big\}\Big|\\
\le&\sup_{\mathcal{K}\in\Omega_n}\sup_{s,s':|s-s'|\le\tau/m}\Big|\Big(\frac{1}{n}\sum_{i=1}^{n}\widehat{I}^2_{\beta\Lambda,i}\Big)^{1/2}\Big[\frac{1}{n}\sum_{i=1}^{n}\big\{I(s\le Y_i)-I(s'\le Y_i)\big\}^2\Big]^{1/2}\Big|\\
=&\sup_{\mathcal{K}\in\Omega_n}\Big(\frac{1}{n}\sum_{i=1}^{n}\widehat{I}^2_{\beta\Lambda,i}\Big)^{1/2}\times\sup_{s,s':|s-s'|\le\tau/m}\Big[\frac{1}{n}\sum_{i=1}^{n}\big\{I(s\le Y_i)-I(s'\le Y_i)\big\}^2\Big]^{1/2},
\end{aligned}
$$

where $\widehat{I}_{\beta\Lambda,i}=\mu'_{1i}\exp(\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X}_i)\{R_iS_i+(1-R_i)\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i}\}$, and the inequality follows from the Cauchy–Schwarz inequality. Note that

$$
\sup_{\mathcal{K}\in\Omega_n}\frac{1}{n}\sum_{i=1}^{n}\widehat{I}^2_{\beta\Lambda,i}>M_1\Big[q_n+\Big\{\frac{t+\log(2r_n)}{n}\Big\}^{1/2}+\frac{q_n\{\log(2n)\}^{1/\xi}\{t+\log(2r_n)\}^{2/\xi}}{n}\Big]
$$

with probability at most $3\exp(-t)$ for any $t>0$ and some positive constant M_1, following the arguments in the proof of Lemma S1 in Wong and Feng (2023). Let I_k denote the interval $\big[\frac{(k-2)\tau}{m},\frac{k\tau}{m}\big)$ for $k=2,\ldots,m$. Since $|I(s\le Y_i)-I(s'\le Y_i)|$ is equal to 1 if Y_i is between s and s' and is equal to 0 if otherwise, we have

$$
\begin{aligned}
&P\Big(\sup_{s,s':|s-s'|\le\tau/m}\Big[\frac{1}{n}\sum_{i=1}^{n}\big\{I(s\le Y_i)-I(s'\le Y_i)\big\}^2\Big]^{1/2}>\Big(\frac{t}{m}\Big)^{1/2}\Big)\\
=&P\Big[\sup_{s,s':|s-s'|\le\tau/m}\frac{1}{n}\sum_{i=1}^{n}I\big(\min(s,s')\le Y_i<\max(s,s')\big)>\frac{t}{m}\Big]\\
\le&P\Big\{\sup_{k=1,\ldots,m}\frac{1}{n}\sum_{i=1}^{n}I(Y_i\in I_k)>\frac{t}{m}\Big\}\\
\le&m\sup_{k=1,\ldots,m}P\Big\{\frac{1}{n}\sum_{i=1}^{n}I(Y_i\in I_k)>\frac{t}{m}\Big\},
\end{aligned}
$$

for any positive t. Let $p_k=P(Y_i\in I_k)$. By Bernstein's inequality, we have

$$
P\Big\{\frac{1}{n}\sum_{i=1}^{n}I(Y_i\in I_k)>\frac{t}{m}\Big\}\le\exp\Big\{-\frac{n^2(\frac{t}{m}-p_k)^2/2}{np_k+n(\frac{t}{m}-p_k)/3}\Big\}.
$$

By the mean-value theorem, $p_k=m^{-1}f_Y(s_k^*)$ for some $s_k^*\in I_k$, where f_Y is the marginal density of Y. Then

$$P\Big(\sup_{s,s':|s-s'|\le\tau/m}\Big[\frac{1}{n}\sum_{i=1}^{n}\big\{I(s\le Y_i)-I(s'\le Y_i)\big\}^2\Big]^{1/2}>\Big(\frac{t}{m}\Big)^{1/2}\Big)$$
$$\le m\sup_k\exp\Big[-\frac{n}{2m}\frac{\{t-f_Y(s_k^*)\}^2}{\{t/2+2f_Y(s_k^*)/3\}}\Big]$$

which is bounded by $\exp(-t)$ for large enough t and n. Thus,

$$\sup_{\mathcal{K}\in\Omega_n}\sup_{s,s':|s-s'|\le\tau/m}\Big|\frac{1}{n}\sum_{i=1}^{n}\widehat{I}_{\beta\Lambda,i}\big\{I(s\le Y_i)-I(s'\le Y_i)\big\}\Big|$$
$$>M_2t^{1/2}\Big[\frac{q_n}{n^{1/2}}+\Big\{\frac{t+\log(2r_n)}{n^2}\Big\}^{1/2}+\frac{q_n\{\log(2n)\}^{1/\xi}\{t+\log(2r_n)\}^{2/\xi}}{n^{3/2}}\Big]^{1/2}$$

with probability at most $4\exp(-t)$ for any $t>0$ and some positive constant M_2. Clearly, because the derivative of $I_{\beta\Lambda}$ is bounded, $\sup_{\mathcal{K}\in\Omega_n}\sup_{s,s':|s-s'|\le\tau/m}\Big|I_{\beta\Lambda}(s)-I_{\beta\Lambda}(s')\Big|=O(m^{-1})$. Using Theorem A.1 of Kuchibhotla et al. (2021), we have

$$\sup_{\mathcal{K}\in\Omega_n}\sup_{s\in\zeta_m}\Big|\widehat{I}_{\beta\Lambda}(s)-I_{\beta\Lambda}(s)\Big|$$
$$>M_3\Big[\Big\{\frac{t+\log n+\log(2r_n)}{n}\Big\}^{1/2}+\frac{q_n^{1/2}\{\log(2n)\}^{1/\xi}\{t+\log n+\log(2r_n)\}^{1/\min(1,\xi)}}{n}\Big]$$

with probability at most $4\exp(-t)$. Combining the above results yields the desired result.

□

Proof of Lemma S5. We consider a single transformation function G. The arguments for $G^{(j)}$, $j=1,\ldots,q$ are essentially the same. Recall that $U_{1i}=\big\{\mu_{1,i}-\mathrm{E}(\mu_1\mid R_i,\boldsymbol{X}_i)\big\}\big\{R_iS_i+(1-R_i)\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i}+\widetilde{\boldsymbol{q}}_{\alpha}^{\mathrm{T}}\boldsymbol{X}_i\big\}+\big\{\mu_{2,i}-\mathrm{E}(\mu_2\mid R_i,\boldsymbol{X}_i)\big\}$. Note that

$$\big\|U_{1i}^3\big\|_{\psi_{\xi/3}}=O\big(1+\big\|R_iS_i+(1-R_i)\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i}+\widetilde{\boldsymbol{q}}_{\alpha}^{\mathrm{T}}\boldsymbol{X}_i\big\|_{\psi_\xi}^3\big)\lesssim q_n^{3/2}.$$

By Theorem A.1 of Kuchibhotla et al. (2021),

$$\sup_{\mathcal{K}\in\Omega_n}\Big\{\frac{1}{n}\sum_{i=1}^{n}|U_{1i}|^3-\mathrm{E}(|U_{11}|^3)\Big\}$$
$$>M_1\Big[\Big\{\frac{t+\log(2r_n)}{n}\Big\}^{1/2}+\frac{q_n^{3/2}\{\log(2n)\}^{3/\xi}\{t+\log(2r_n)\}^{3/\xi}}{n}\Big]$$

with probability at most $3\exp(-t)$ for any $t>0$ and some positive constant M_1. Because $\mathrm{E}(|U_{11}|^3)$ is uniformly bounded over $\mathcal{K}\in\Omega_n$,

$$\frac{1}{n^{3/2}}\sup_{\mathcal{K}\in\Omega_n}\sum_{i=1}^{n}|U_{1i}|^3>M_2\Big[\frac{\{t+\log(2r_n)\}^{1/2}}{n}+\frac{q_n^{3/2}\{\log(2n)\}^{3/\xi}\{t+\log(2r_n)\}^{3/\xi}}{n^{3/2}}\Big]$$

with probability at most $M_3\exp(-t)$ for any $t>0$ and some positive constants M_2 and M_3.

Recall that U_{2i} is equal to

$$(\boldsymbol{\gamma}_{0X}^{\mathrm{T}}+\widetilde{\boldsymbol{q}}_{\alpha}^{\mathrm{T}})\big\{\mathrm{E}(\mu_1\mid R_i,\boldsymbol{X}_i)\boldsymbol{X}_i-\mathrm{E}(\mu_1\boldsymbol{X}\mid R_i)\big\}+\big\{\mathrm{E}(\mu_1\mid R_i,\boldsymbol{X}_i)-\mathrm{E}(\mu_1\mid R_i)\big\}\boldsymbol{\gamma}_{0A,\mathcal{K}}^{\mathrm{T}}\boldsymbol{A}_{\mathcal{K},i}$$
$$+\big\{\mathrm{E}(\mu_1\mid R_i,\boldsymbol{X}_i)+\boldsymbol{I}_{\beta\gamma}^{\mathrm{T}}\boldsymbol{I}_{\gamma\gamma}^{-1}\boldsymbol{W}_{\mathcal{K},i}\big\}R_i(S_i-\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i})+\big\{\mathrm{E}(\mu_2\mid R_i,\boldsymbol{X}_i)-\mathrm{E}(\mu_2\mid R_i)\big\}.$$

By the independence of $\boldsymbol{X}$ and $\boldsymbol{A}_{\mathcal{K}}$, $\|\boldsymbol{I}_{\beta\gamma}^{\mathrm{T}}\boldsymbol{I}_{\gamma\gamma}^{-1}\boldsymbol{W}_{\mathcal{K}}\|$ involves only $\boldsymbol{X}$ and thus is bounded, so

$$\big\|U_{2i}^3\big\|_{\psi_{\xi/3}}=O\big(1+\big\|\boldsymbol{\gamma}_{0A,\mathcal{K}}^{\mathrm{T}}\boldsymbol{A}_{\mathcal{K},i}\big\|_{\psi_\xi}^3+\big\|S_i-\boldsymbol{\gamma}_{0\mathcal{K}}^{\mathrm{T}}\boldsymbol{W}_{\mathcal{K},i}\big\|_{\psi_\xi}^3\big)\lesssim q_n^{3/2}.$$

By Theorem A.1 of Kuchibhotla et al. (2021),

$$\sup_{\mathcal{K}\in\Omega_n}\Big\{\frac{1}{n}\sum_{i=1}^{n}|U_{2i}|^3-\mathrm{E}(|U_{21}|^3)\Big\}$$
$$>M_4\Big[\Big\{\frac{t+\log(2r_n)}{n}\Big\}^{1/2}+\frac{q_n^{3/2}\{\log(2n)\}^{3/\xi}\{t+\log(2r_n)\}^{3/\xi}}{n}\Big]$$

with probability at most $3\exp(-t)$ for any $t>0$ and some positive constant M_4. Because

$\mathrm{E}(|U_{21}|^3)$ is uniformly bounded over $\mathcal{K} \in \Omega_n$,

$$\frac{1}{n^{3/2}} \sup_{\mathcal{K}\in\Omega_n} \sum_{i=1}^{n} |U_{2i}|^3 > M_5 \Big[\frac{\{t+\log(2r_n)\}^{1/2}}{n} + \frac{q_n^{3/2}\{\log(2n)\}^{3/\xi}\{t+\log(2r_n)\}^{3/\xi}}{n^{3/2}}\Big]$$

with probability at most $M_6 \exp(-t)$ for any $t > 0$ and some positive constants M_5 and M_6. Similar arguments show that the same bound applies to the terms involving $\widetilde{U}_{1i}$, $\widetilde{U}_{2i}$, U_{3i}, and $\widetilde{U}_{3i}$. □

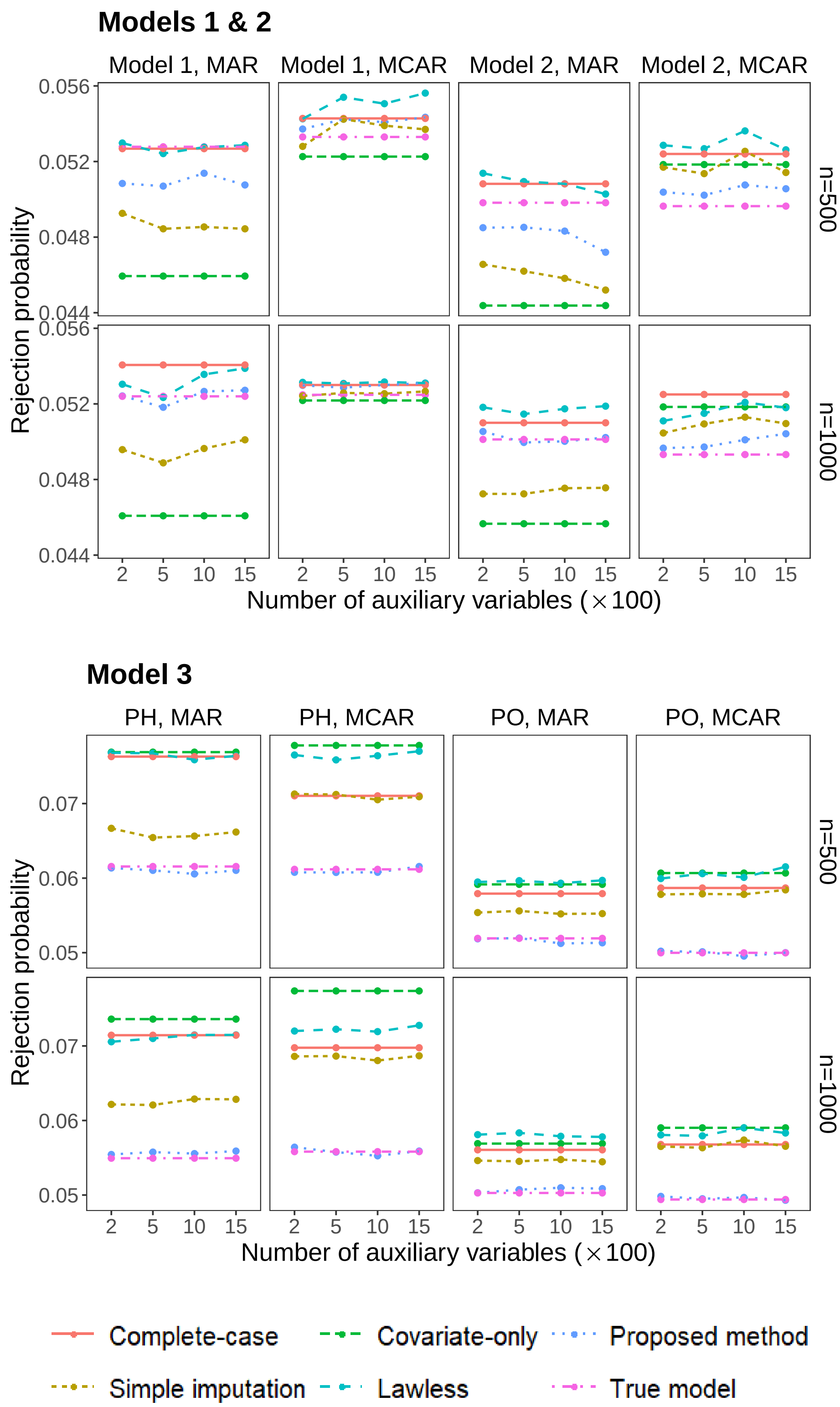

Figure S1: Study 1 — Rejection probabilities under a missing proportion of 30% and the null hypothesis.

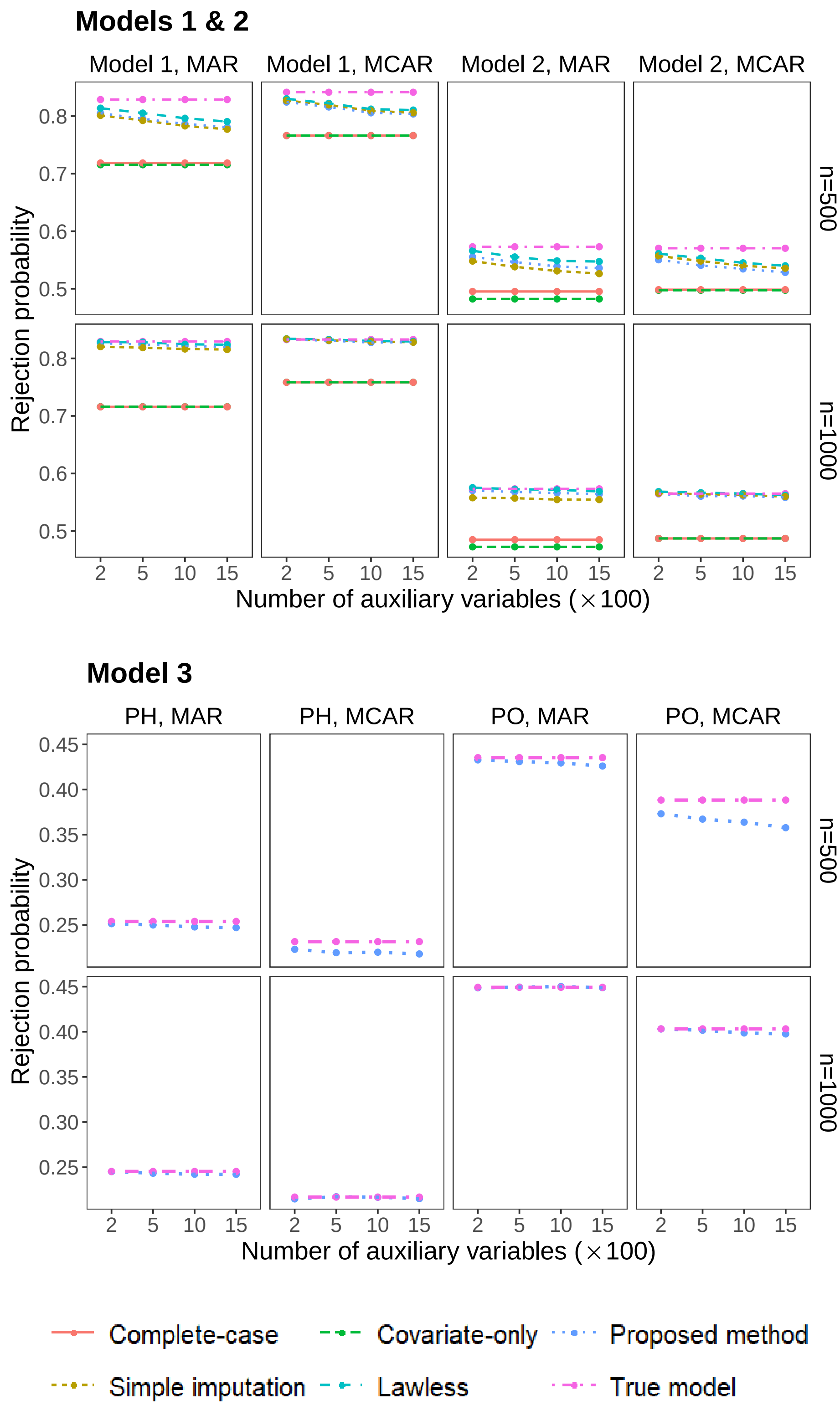

Figure S2: Study 1 — Rejection probabilities under a missing proportion of 30% and the alternative hypothesis.

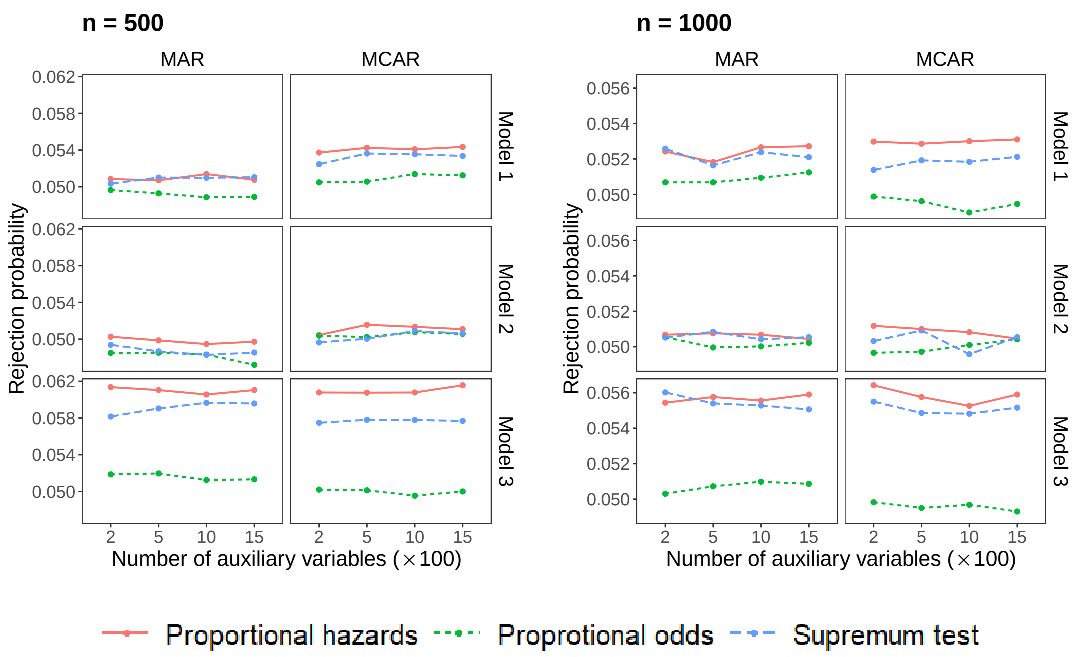

Figure S3: Study 2 — Rejection probabilities under a missing proportion of 30% and the null hypothesis.

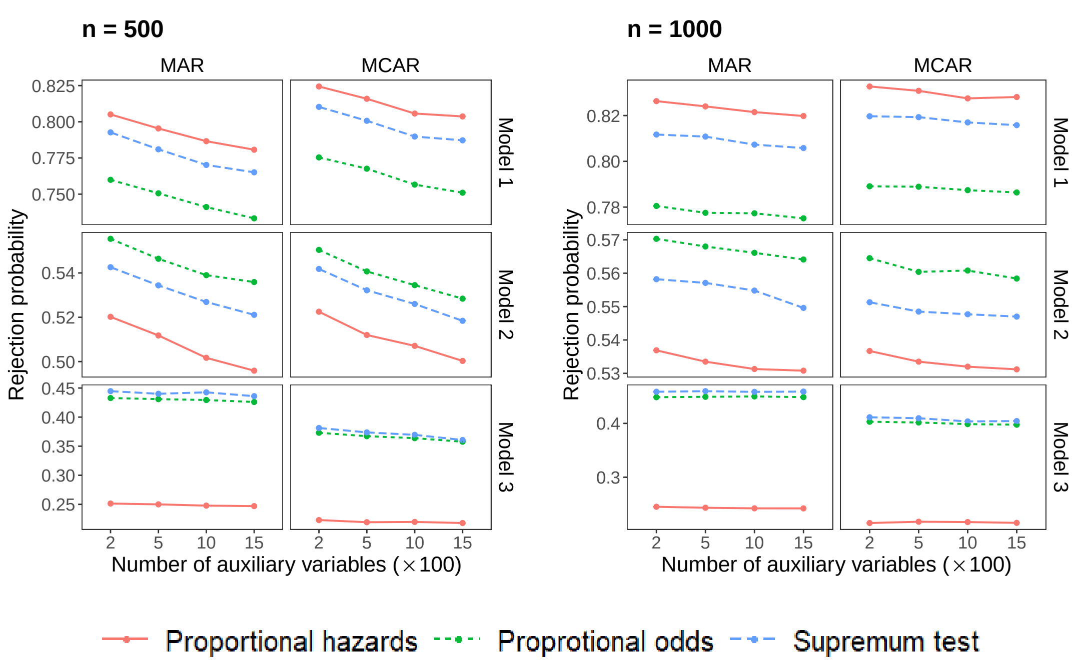

Figure S4: Study 2 — Rejection probabilities under a missing proportion of 30% and the alternative hypothesis.

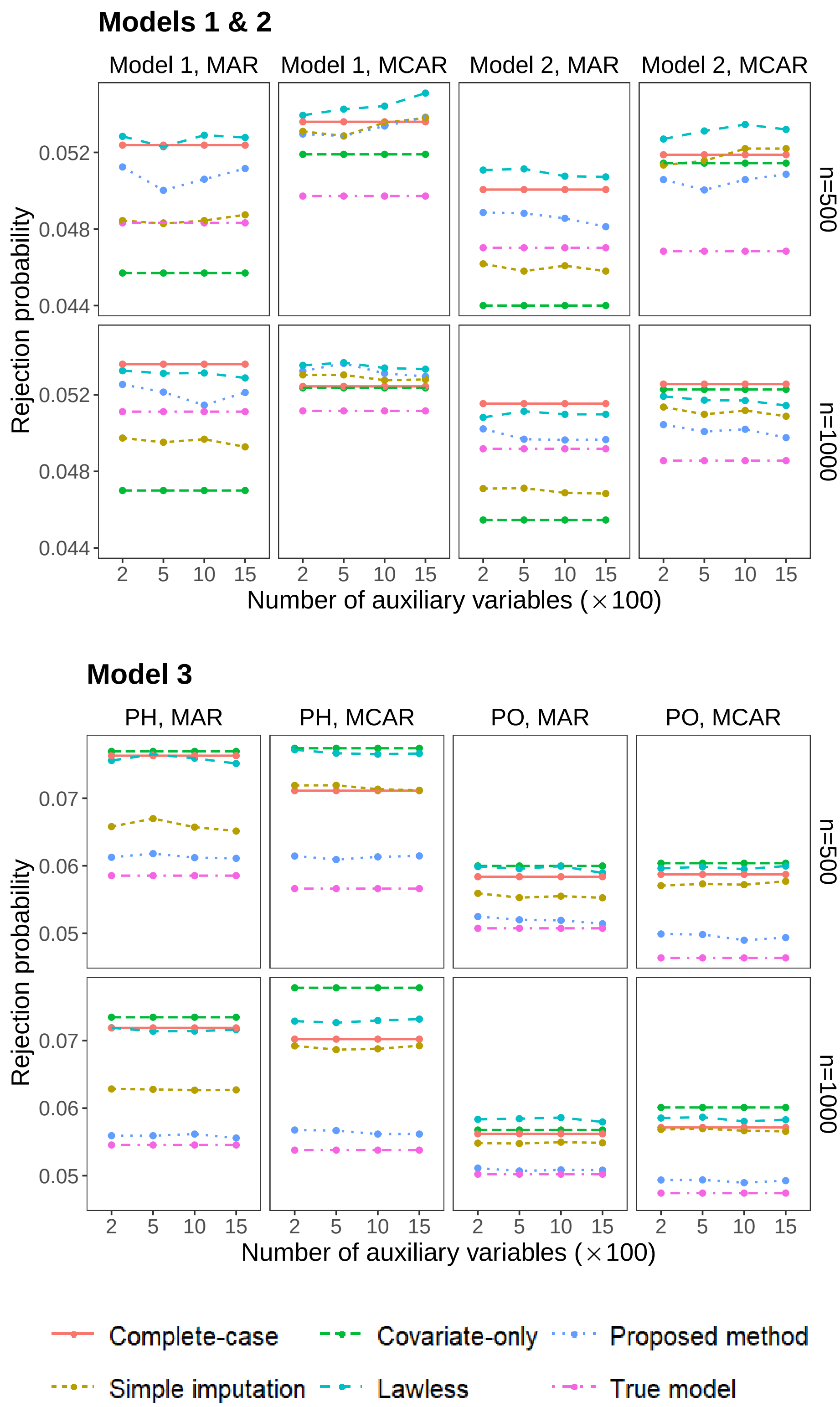

Figure S5: Study 3 — Rejection probabilities under a missing proportion of 30% and the null hypothesis.

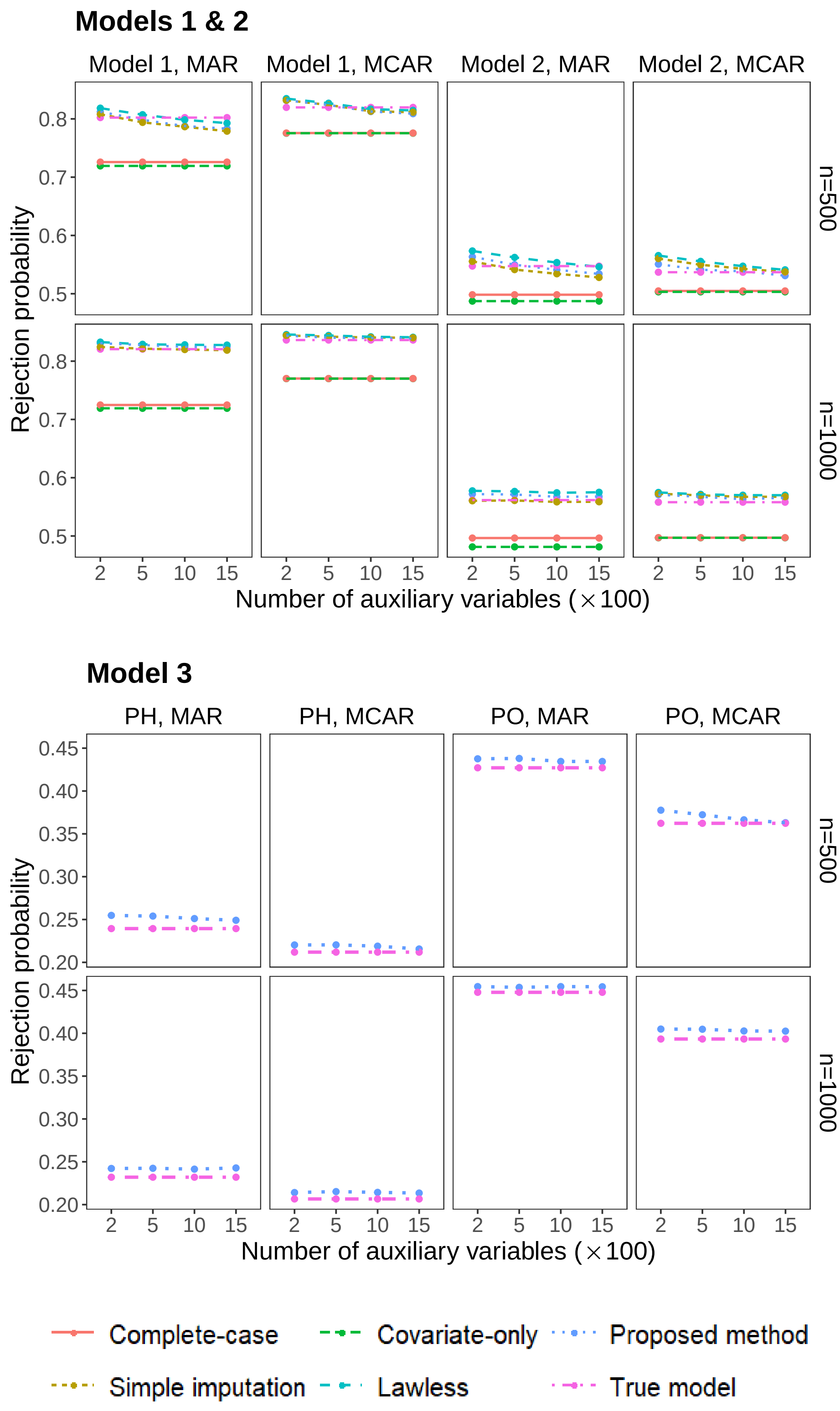

Figure S6: Study 3 — Rejection probabilities under a missing proportion of 30% and the alternative hypothesis.

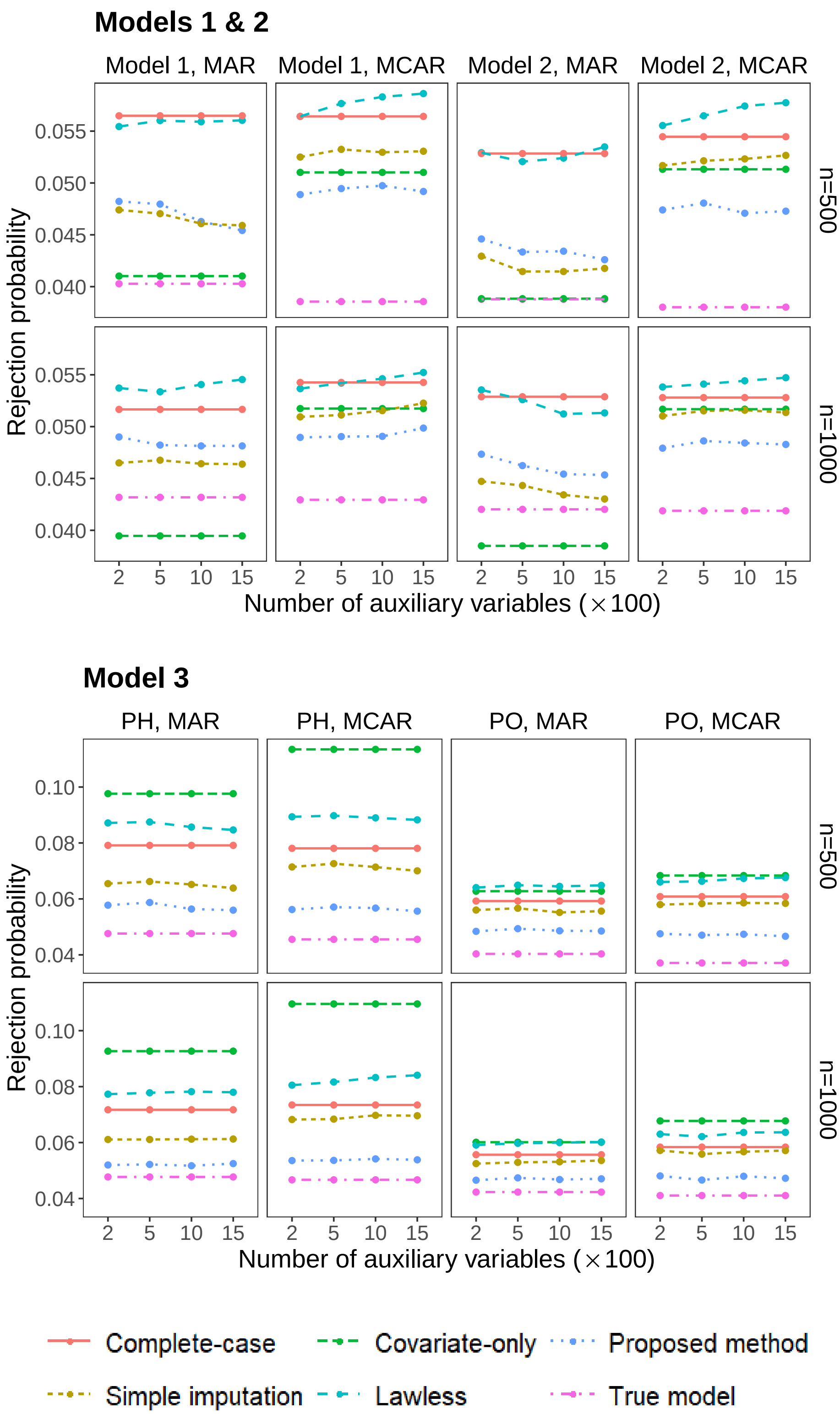

Figure S7: Study 3 — Rejection probabilities under a missing proportion of 60% and the null hypothesis.

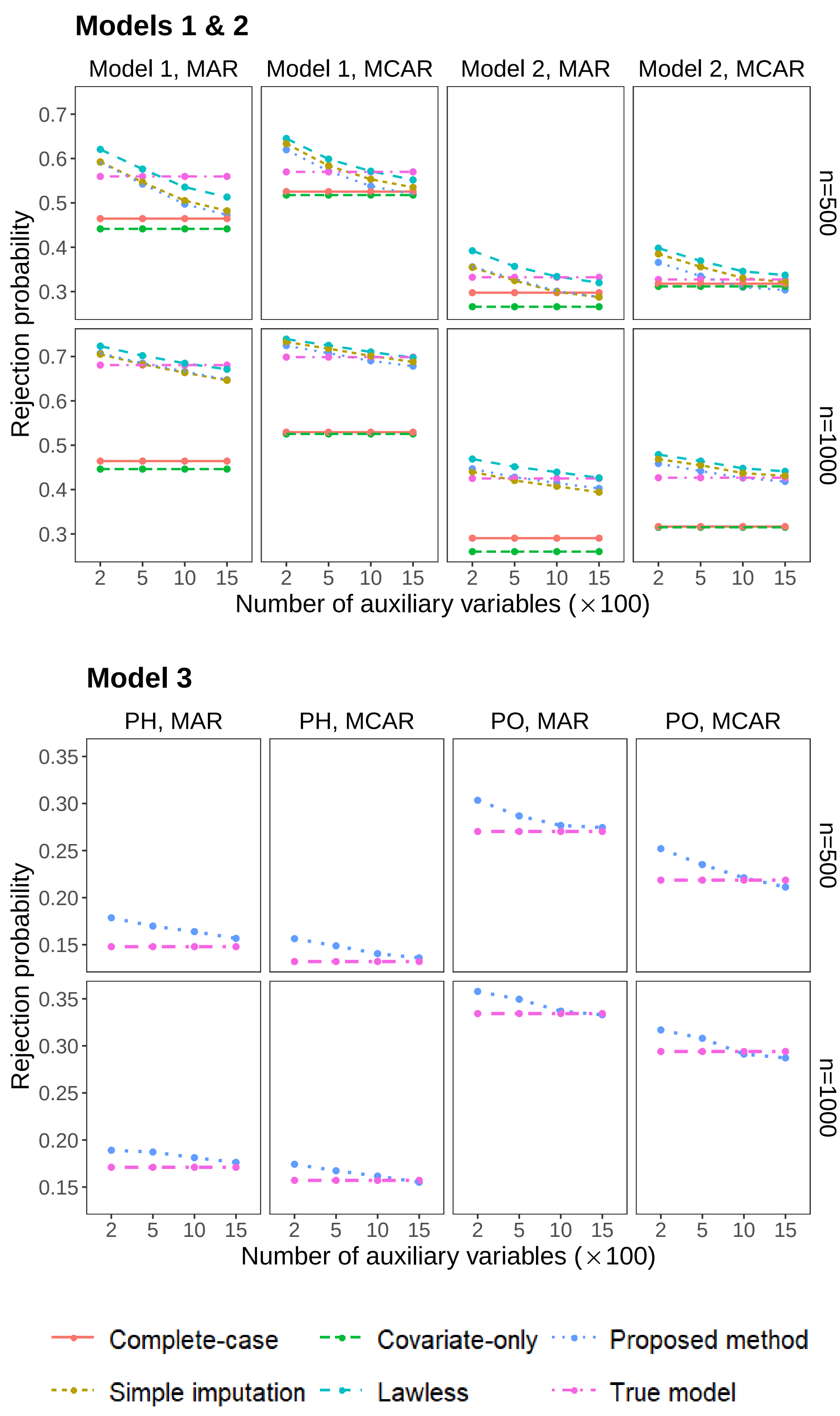

Figure S8: Study 3 — Rejection probabilities under a missing proportion of 60% and the alternative hypothesis.

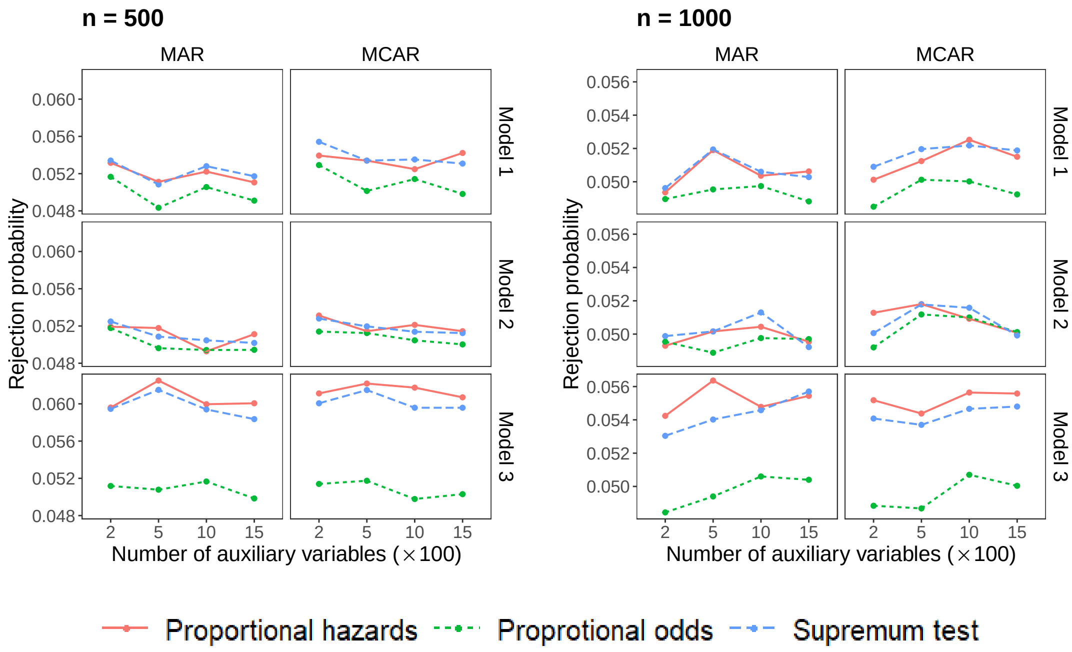

Figure S9: Study 4 — Rejection probabilities under a missing proportion of 30% and the null hypothesis.

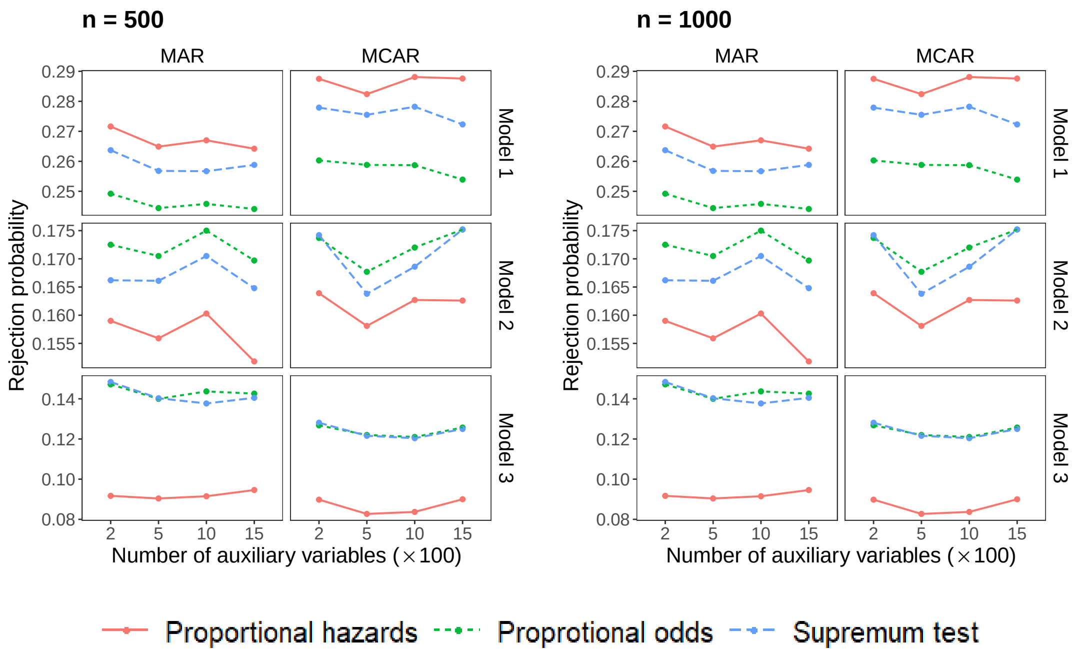

Figure S10: Study 4 — Rejection probabilities under a missing proportion of 30% and the alternative hypothesis.

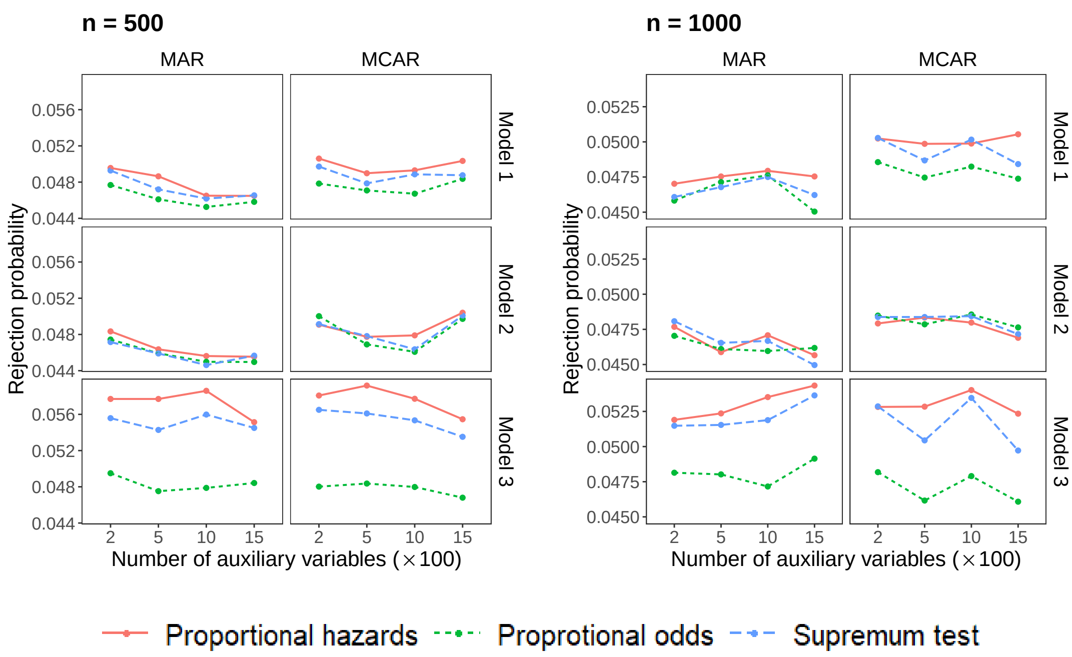

Figure S11: Study 4 — Rejection probabilities under a missing proportion of 60% and the null hypothesis.

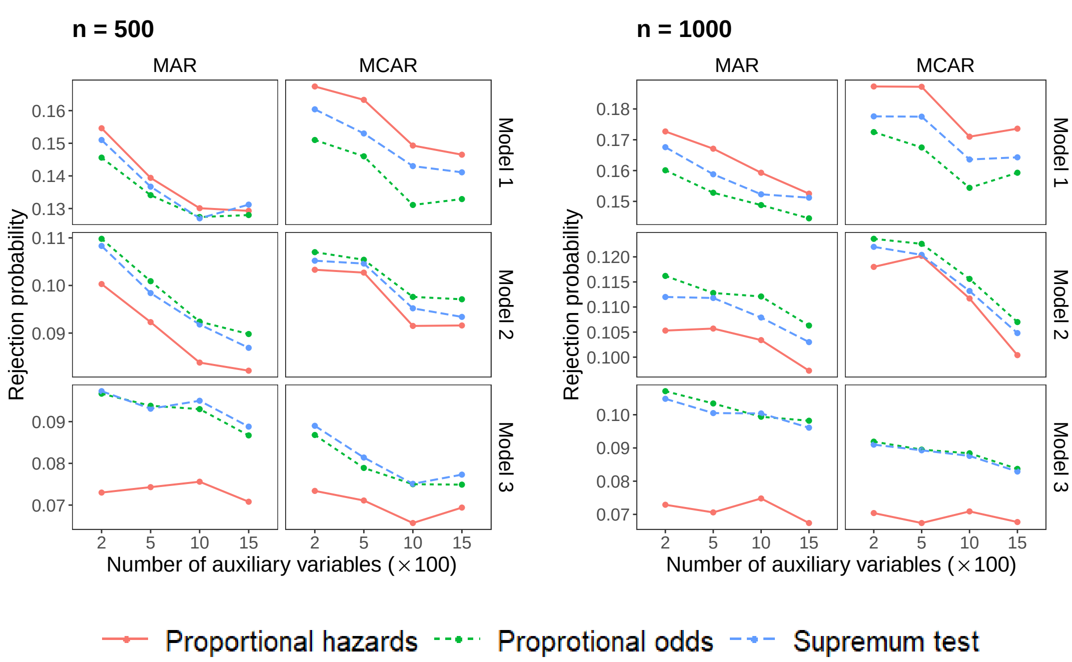

Figure S12: Study 4 — Rejection probabilities under a missing proportion of 60% and the alternative hypothesis.

References

Kuchibhotla, A. K., Brown, L. D., Buja, A., George, E. I. and Zhao, L. (2021). Uniform-in-submodel bounds for linear regression in a model-free framework. *Econometric Theory* **39**, 1202–1248.

Pollard, D. (2002). *A User's Guide to Measure Theoretic Probability* (No. 8). Cambridge University Press.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

Wong, K. Y. and Feng, J. (2023). Score tests with incomplete covariates and high-dimensional auxiliary variables. *Statistica Sinica* **33**, 1483–1505.

Zeng, D., Lin, D. Y. and Lin, X. (2008). Semiparametric transformation models with random effects for clustered failure time data. *Statistica Sinica* **18**, 355–377.