E-Companion to "Ranking and Selection with Covariates for Personalized Decision Making" by Shen, Hong and Zhang

EC.1. Proof of Lemma 1

The following Lemma EC.1 is a more general version of Lemma 1 in §3. Therefore we only provide the proof of Lemma EC.1 and remark that Lemma 1 is a special case. Also note that Lemma EC.1 is used directly in the proof of Theorem 2.

LEMMA EC.1. *For each* $j=1,\ldots,m$, *let* $Y(\boldsymbol{x}_j)=\boldsymbol{x}_j^\intercal\boldsymbol{\beta}+\epsilon(\boldsymbol{x}_j)$, *where* $\boldsymbol{\beta},\boldsymbol{x}_j\in\mathbb{R}^d$ *and* $\epsilon(\boldsymbol{x}_j)\sim\mathcal{N}(0,\sigma_j^2)$. *Suppose that* $\epsilon(\boldsymbol{x}_1),\ldots,\epsilon(\boldsymbol{x}_m)$ *are independent. Let* $Y_1(\boldsymbol{x}_j),Y_2(\boldsymbol{x}_j),\ldots$ *be independent samples of* $Y(\boldsymbol{x}_j)$. *Let* T *be a set of random variables independent of* $\sum_{\ell=1}^n Y_\ell(\boldsymbol{x}_j)$ *and of* $\{Y_\ell(\boldsymbol{x}_j):\ell\geq n+1\}$, *for all* $j=1,\ldots,m$. *Suppose* $N_j\geq n$ *is an integer-valued function of* T *and no other random variables. Let* $\widehat{Y}_j=N_j^{-1}\sum_{\ell=1}^{N_j}Y_\ell(\boldsymbol{x}_j)$, $\widehat{\boldsymbol{Y}}=(\widehat{Y}_1,\ldots,\widehat{Y}_m)^\intercal$, $\mathcal{X}=(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_m)^\intercal$, $\widehat{\boldsymbol{\beta}}=(\mathcal{X}^\intercal\mathcal{X})^{-1}\mathcal{X}^\intercal\widehat{\boldsymbol{Y}}$, *and* $\Sigma=\mathrm{Diag}(\sigma_1^2/N_1,\ldots,\sigma_m^2/N_m)$. *Then, for any* $\boldsymbol{x}\in\mathbb{R}^d$,

(i) $\boldsymbol{x}^\intercal\widehat{\boldsymbol{\beta}}\big|T\sim\mathcal{N}(\boldsymbol{x}^\intercal\boldsymbol{\beta},\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\mathcal{X}^\intercal\Sigma\mathcal{X}(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x})$;

(ii) $\dfrac{\boldsymbol{x}^\intercal\widehat{\boldsymbol{\beta}}-\boldsymbol{x}^\intercal\boldsymbol{\beta}}{\sqrt{\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\mathcal{X}^\intercal\Sigma\mathcal{X}(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}}}$ *is independent of* T *and has the standard normal distribution.*

Proof. For part (i), by the definition of $\widehat{\boldsymbol{\beta}}$, it suffices to show that $\widehat{\boldsymbol{Y}}\big|T\sim\mathcal{N}(\mathcal{X}\boldsymbol{\beta},\Sigma)$. We first notice that $Y(\boldsymbol{x}_j)\sim\mathcal{N}(\boldsymbol{x}_j^\intercal\boldsymbol{\beta},\sigma_j^2)$. Since T is independent of $\sum_{\ell=1}^n Y_\ell(\boldsymbol{x}_j)$,

$$\sum_{\ell=1}^n Y_\ell(\boldsymbol{x}_j)\Big|T\sim\mathcal{N}(n\boldsymbol{x}_j^\intercal\boldsymbol{\beta},n\sigma_j^2).$$

On the other hand, since T is independent of $\{Y_\ell(\boldsymbol{x}_j):\ell\geq n+1\}$ and N_j is a function only of T,

$$\sum_{\ell=n+1}^{N_j} Y_\ell(\boldsymbol{x}_j)\Big|T\sim\mathcal{N}((N_j-n)\boldsymbol{x}_j^\intercal\boldsymbol{\beta},(N_j-n)\sigma_j^2).$$

Since $\sum_{\ell=1}^n Y_\ell(\boldsymbol{x}_j)$ and $\sum_{\ell=n+1}^{N_j}Y_\ell(\boldsymbol{x}_j)$ are independent,

$$\widehat{Y}_j\Big|T=\frac{1}{N_j}\left(\sum_{\ell=1}^n Y_\ell(\boldsymbol{x}_j)+\sum_{\ell=n+1}^{N_j}Y_\ell(\boldsymbol{x}_j)\right)\Bigg|T\sim\mathcal{N}\left(\boldsymbol{x}_j^\intercal\boldsymbol{\beta},\sigma_j^2/N_j\right).$$

Notice that $\widehat{Y}_1,\ldots,\widehat{Y}_m$ are independent conditionally on T, so $\widehat{\boldsymbol{Y}}\big|T\sim N(\mathcal{X}\boldsymbol{\beta},\Sigma)$.

For part (ii), let

$$V=\frac{\boldsymbol{x}^\intercal\widehat{\boldsymbol{\beta}}-\boldsymbol{x}^\intercal\boldsymbol{\beta}}{\sqrt{\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\mathcal{X}^\intercal\Sigma\mathcal{X}(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}}},$$

then $V|T\sim\mathcal{N}(0,1)$ by part (i). Notice that $\mathbb{P}(V<v|T)=\Phi(v)$ is not a function of T for any v, so V is independent of T. □

REMARK EC.1. It is easy to see that Lemma 1 in §3 is a special case of Lemma EC.1 with $\sigma_1=\cdots=\sigma_m=\sigma$ and $N_1=\cdots=N_m=N$.

EC.2. Computing h in High Dimensions

If $\boldsymbol{X}$ is high-dimensional, the numerical integration in (4) for computing h suffers from the curse of dimensionality. For instance, the error in the trapezoidal rule for d-dimensional numerical integration is $\mathcal{O}(n^{-2/d})$ in general. One solution is the Monte Carlo method. Let

$$f(\boldsymbol{x},h) := \int_0^\infty \left[\int_0^\infty \Phi\left(\frac{h}{\sqrt{(n_0m-d)(t^{-1}+s^{-1})\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}}}\right)\eta(s)\mathrm{d}s\right]^{k-1}\eta(t)\mathrm{d}t,$$

and generate n i.i.d. samples of $\boldsymbol{X}$, $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$. Then, $\mathbb{E}[f(\boldsymbol{X},h)]$, the left-hand side of (4), can be approximated by $n^{-1}\sum_{i=1}^n f(\boldsymbol{x}_i,h)$ with error $\mathcal{O}(n^{-1/2})$. So the Monte Carlo method is more efficient when $d>4$. We can then solve $n^{-1}\sum_{i=1}^n f(\boldsymbol{x}_i,h)=1-\alpha$ for h by using the MATLAB built-in root finding function `fzero`.

Another approach to computing h in high dimensions is the stochastic approximation method (Robbins and Monro 1951). Given an initial value $h_0\geq 0$, define $h_{n+1}=\Pi\{h_n-a_n(f(\boldsymbol{x}_n,h_n)-(1-\alpha))\}$, where $\Pi\{\cdot\}$ denotes a projection that maps a point outside $[0,\infty)$ to $[0,\infty)$ (e.g., $\Pi\{\cdot\}=\|\cdot\|$ or $\Pi\{\cdot\}=\max\{0,\cdot\}$), $\boldsymbol{x}_n$ is an independent realization of $\boldsymbol{X}$, and $\{a_n\}$ is a sequence of constants satisfying $\sum_{n=0}^\infty a_n=\infty$ and $\sum_{n=0}^\infty a_n^2<\infty$. A common choice of $\{a_n\}$ is $a_n=a/n$, for some $a>0$. It can be shown that h_n converges to h at a rate of $\mathcal{O}(n^{-1/2})$.

EC.3. Proof of Theorem 1

The proof of Theorem 1 critically relies on the extended Stein's lemma (Lemma 1). It also needs the following lemma, often known as Slepian's Inequality (Slepian 1962).

LEMMA EC.2 **(Slepian's Inequality)**. *Suppose that $(Z_1,\ldots,Z_k)^\intercal$ has a multivariate normal distribution. If $\mathrm{Cov}(Z_i,Z_j)\geq 0$ for all $1\leq i,j\leq k$, then, for any constants c_i, $i=1,\ldots,k$,*

$$\mathbb{P}\left(\bigcap_{i=1}^k\{Z_i\geq c_i\}\right)\geq\prod_{i=1}^k\mathbb{P}(Z_i\geq c_i).$$

Proof of Theorem 1. Notice that N_i is an integer-valued function only of S_i^2, which is the OLS estimator of σ^2. Under Assumption 1, by Lemma 1 and Remark 7,

$$\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i\Big|\left(\boldsymbol{X},S_i^2\right)\sim\mathcal{N}\left(\boldsymbol{X}^\intercal\boldsymbol{\beta}_i,\frac{\sigma_i^2}{N_i}\boldsymbol{X}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{X}\right),\quad i=1,\ldots,k. \tag{EC.1}$$

Moreover, let $\xi_i=(n_0m-d)S_i^2/\sigma_i^2$ for all $i=1,\ldots,k$. Then, ξ_i has the chi-square distribution with (n_0m-d) degrees of freedom, for $i=1,\ldots,k$ (see Remark 7).

For notational simplicity, we let $V(\boldsymbol{X}):=\boldsymbol{X}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{X}$ and temporarily write $i^*=i^*(\boldsymbol{X})$ to suppress the dependence on $\boldsymbol{X}$. Let $\Omega(\boldsymbol{x}):=\{i:\boldsymbol{X}^\intercal\boldsymbol{\beta}_{i^*}-\boldsymbol{X}^\intercal\boldsymbol{\beta}_i\geq\delta|\boldsymbol{X}=\boldsymbol{x}\}$ be the set of alternatives outside the IZ given $\boldsymbol{X}=\boldsymbol{x}$. For each $i\in\Omega(\boldsymbol{X})$, $\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*}$ is independent of $\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i$ given $\boldsymbol{X}$. It then follows from (EC.1) that

$$\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*}-\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i\Big|\left(\boldsymbol{X},S_{i^*}^2,S_i^2\right)\sim\mathcal{N}\left(\boldsymbol{X}^\intercal\boldsymbol{\beta}_{i^*}-\boldsymbol{X}^\intercal\boldsymbol{\beta}_i,(\sigma_{i^*}^2/N_{i^*}+\sigma_i^2/N_i)V(\boldsymbol{X})\right). \tag{EC.2}$$

Hence, letting Z denote a standard normal random variable, for each $i\in\Omega(\boldsymbol{X})$, we have

$$\begin{aligned}\mathbb{P}\left(\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*}-\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i>0\Big|\boldsymbol{X},S_{i^*}^2,S_i^2\right)&=\mathbb{P}\left(Z>\frac{-(\boldsymbol{X}^\intercal\boldsymbol{\beta}_{i^*}-\boldsymbol{X}^\intercal\boldsymbol{\beta}_i)}{\sqrt{(\sigma_{i^*}^2/N_{i^*}+\sigma_i^2/N_i)V(\boldsymbol{X})}}\Bigg|\boldsymbol{X},S_{i^*}^2,S_i^2\right)\\&\geq\mathbb{P}\left(Z>\frac{-\delta}{\sqrt{[\sigma_{i^*}^2\delta^2/(h^2S_{i^*}^2)+\sigma_i^2\delta^2/(h^2S_i^2)]V(\boldsymbol{X})}}\Bigg|\boldsymbol{X},S_{i^*}^2,S_i^2\right)\\&=\Phi\left(\frac{h}{\sqrt{(n_0m-d)(\xi_{i^*}^{-1}+\xi_i^{-1})V(\boldsymbol{X})}}\right),\end{aligned} \tag{EC.3}$$

where the inequality follows the definitions of $\Omega(\boldsymbol{X})$ and N_i, and the last equality follows the definition of ξ_i.

Then, conditionally on $\boldsymbol{X}$, by the definition (2), the CS event must occur if alternative i^* eliminates all alternatives in $\Omega(\boldsymbol{X})$. Thus,

$$\begin{aligned}\mathrm{PCS}(\boldsymbol{X})&\geq\mathbb{P}\left(\bigcap_{i\in\Omega(\boldsymbol{X})}\left\{\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*}-\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i>0\right\}\Bigg|\boldsymbol{X}\right)\\&=\mathbb{E}\left[\mathbb{P}\left(\bigcap_{i\in\Omega(\boldsymbol{X})}\left\{\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*}-\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i>0\right\}\Bigg|\boldsymbol{X},S_{i^*}^2,\{S_i^2:i\in\Omega(\boldsymbol{X})\}\right)\Bigg|\boldsymbol{X}\right],\end{aligned} \tag{EC.4}$$

where the equality is due to the tower law of conditional expectation. Notice that conditionally on $\{\boldsymbol{X},S_{i^*}^2,\{S_i^2:i\in\Omega(\boldsymbol{X})\}\}$, $\{\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*}-\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i:i\in\Omega(\boldsymbol{X})\}$ is multivariate normal by (EC.2). Moreover, for $i,i'\in\Omega(\boldsymbol{X})$ and $i\neq i'$, due to the conditional independence between $\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i$ and $\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i'}$,

$$\mathrm{Cov}\left(\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*}-\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i,\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*}-\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i'}\Big|\boldsymbol{X},S_{i^*}^2,\{S_i^2:i\in\Omega(\boldsymbol{X})\}\right)=\mathrm{Var}\left(\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*}\Big|\boldsymbol{X},S_{i^*}^2\right)>0.$$

Therefore, applying (EC.4) and Lemma EC.2, we have

$$\begin{aligned}\mathrm{PCS}(\boldsymbol{X})&\geq\mathbb{E}\left[\prod_{i\in\Omega(\boldsymbol{X})}\mathbb{P}\left(\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*}-\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i>0\Big|\boldsymbol{X},S_{i^*}^2,S_i^2\right)\Bigg|\boldsymbol{X}\right]\\&\geq\mathbb{E}\left[\prod_{i\in\Omega(\boldsymbol{X})}\Phi\left(\frac{h}{\sqrt{(n_0m-d)(\xi_{i^*}^{-1}+\xi_i^{-1})V(\boldsymbol{X})}}\right)\Bigg|\boldsymbol{X}\right]\\&=\int_0^\infty\left[\int_0^\infty\Phi\left(\frac{h}{\sqrt{(n_0m-d)(t^{-1}+s^{-1})V(\boldsymbol{X})}}\right)\eta(s)\mathrm{d}s\right]^{|\Omega(\boldsymbol{X})|}\eta(t)\mathrm{d}t,\end{aligned} \tag{EC.5}$$

where the second inequality follows from (EC.3), and $|\Omega(\boldsymbol{X})|$ denotes the cardinality of $\Omega(\boldsymbol{X})$. Since $0 \le \Phi(\cdot) \le 1$ and $\eta(\cdot)$ is a pdf, the integral inside the square brackets in (EC.5) is no greater than 1. Moreover, since $|\Omega(\boldsymbol{X})| \le k-1$, hence,

$$\mathrm{PCS}(\boldsymbol{X}) \ge \int_0^\infty \left[\int_0^\infty \Phi\left(\frac{h}{\sqrt{(n_0 m - d)(t^{-1}+s^{-1})V(\boldsymbol{X})}}\right)\eta(s)\mathrm{d}s\right]^{k-1} \eta(t)\mathrm{d}t.$$

Then, it follows immediately from the definition of h in (4) that $\mathrm{PCS_E} = \mathbb{E}[\mathrm{PCS}(\boldsymbol{X})] \ge 1-\alpha$. □

EC.4. Proof of Theorem 2

The proof of Theorem 2 critically relies on Lemma EC.1.

Proof of Theorem 2. Under Assumption 2, for $i=1,\dots,k, j=1,\dots,m$, $\overline{Y}_{ij}$ is independent of S_{ij}^2; moreover, let $\sigma_{ij} = \sigma_i(\boldsymbol{x}_j)$, then $\xi_{ij} := (n_0-1)S_{ij}^2/\sigma_{ij}^2 \sim \chi^2_{n_0-1}$; see, e.g., Examples 5.6a and 5 in Rencher and Schaalje (2008). Let $\mathcal{S}_i := \{S_{i1}^2,\dots,S_{im}^2\}$, for $i=1,\dots,k$. Then, $\mathcal{S}_i$ is independent of $\sum_{\ell=1}^{n_0} Y_{i\ell}(\boldsymbol{x}_j)$ and of $\{Y_{i\ell}(\boldsymbol{x}_j) : \ell \ge n_0+1\}$. Since $N_{i1},\dots,N_{im}$ are integer-valued functions only of $\mathcal{S}_i$, by Lemma EC.1, for $i=1,\dots,k$,

$$\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i\Big| (\boldsymbol{X},\mathcal{S}_i) \sim \mathcal{N}\left(\boldsymbol{X}^\intercal\boldsymbol{\beta}_i, \boldsymbol{X}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\mathcal{X}^\intercal\Sigma_i\mathcal{X}(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{X}\right),$$

where $\Sigma_i = \mathrm{Diag}(\sigma_{i1}^2/N_{i1},\dots,\sigma_{im}^2/N_{im})$.

For notational simplicity, let $\boldsymbol{a} := (a_1,\dots,a_m)^\intercal := \mathcal{X}(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{X}$ and write $i^* = i^*(\boldsymbol{X})$ to suppress the dependence on $\boldsymbol{X}$. Then,

$$\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i\Big| (\boldsymbol{X},\mathcal{S}_i) \sim N\left(\boldsymbol{X}^\intercal\boldsymbol{\beta}_i, \sum_{j=1}^m a_j^2\sigma_{ij}^2/N_{ij}\right). \tag{EC.6}$$

Let $\Omega(\boldsymbol{x}) := \{i : \boldsymbol{X}^\intercal\boldsymbol{\beta}_{i^*} - \boldsymbol{X}^\intercal\boldsymbol{\beta}_i \ge \delta | \boldsymbol{X} = \boldsymbol{x}\}$ be the set of alternatives outside the IZ given $\boldsymbol{X}=\boldsymbol{x}$. For each $i \in \Omega(\boldsymbol{X})$, $\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*}$ is independent of $\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i$ given $\boldsymbol{X}$. It then follows from (EC.6) that

$$\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*} - \boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i\Big| (\boldsymbol{X},\mathcal{S}_{i^*},\mathcal{S}_i) \sim \mathcal{N}\left(\boldsymbol{X}^\intercal\boldsymbol{\beta}_{i^*} - \boldsymbol{X}^\intercal\boldsymbol{\beta}_i, \sum_{j=1}^m a_j^2(\sigma_{i^*j}^2/N_{i^*j} + \sigma_{ij}^2/N_{ij})\right). \tag{EC.7}$$

Hence, letting Z denote a standard normal random variable, for each $i \in \Omega(\boldsymbol{X})$, we have

$$\begin{aligned}
\mathbb{P}\left(\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*} - \boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i > 0 \Big| \boldsymbol{X},\mathcal{S}_{i^*},\mathcal{S}_i\right) &= \mathbb{P}\left(Z > \frac{-(\boldsymbol{X}^\intercal\boldsymbol{\beta}_{i^*} - \boldsymbol{X}^\intercal\boldsymbol{\beta}_i)}{\sqrt{\sum_{j=1}^m a_j^2\left(\sigma_{i^*j}^2/N_{i^*j} + \sigma_{ij}^2/N_{ij}\right)}}\Bigg| \boldsymbol{X},\mathcal{S}_{i^*},\mathcal{S}_i\right) \\
&\ge \mathbb{P}\left(Z > \frac{-\delta}{\sqrt{\delta^2 h_{\mathrm{Het}}^{-2}\sum_{j=1}^m a_j^2\left(\sigma_{i^*j}^2/S_{i^*j}^2 + \sigma_{ij}^2/S_{ij}^2\right)}}\Bigg| \boldsymbol{X},\mathcal{S}_{i^*},\mathcal{S}_i\right) \\
&= \Phi\left(\frac{h_{\mathrm{Het}}}{\sqrt{(n_0-1)\sum_{j=1}^m a_j^2\left(1/\xi_{i^*j} + 1/\xi_{ij}\right)}}\right),
\end{aligned} \tag{EC.8}$$

where the inequality follows the definition of $\Omega(\boldsymbol{X})$ and N_{ij}, and the last equality from that of ξ_{ij}.

Then, conditionally on $\boldsymbol{X}$, by the definition (2), the CS event must occur if alternative i^* eliminates all alternatives in $\Omega(\boldsymbol{X})$. Thus,

$$\begin{aligned}\mathrm{PCS}(\boldsymbol{X}) &\geq \mathbb{P}\left(\bigcap_{i\in\Omega(\boldsymbol{X})}\left\{\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*}-\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i>0\right\}\middle|\boldsymbol{X}\right)\\ &=\mathbb{E}\left[\mathbb{P}\left(\bigcap_{i\in\Omega(\boldsymbol{X})}\left\{\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*}-\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i>0\right\}\middle|\boldsymbol{X},\mathcal{S}_{i^*},\{\mathcal{S}_i:i\in\Omega(\boldsymbol{X})\}\right)\middle|\boldsymbol{X}\right],\end{aligned}\tag{EC.9}$$

where the equality is due to the tower law of conditional expectation. Notice that conditionally on $\{\boldsymbol{X},\mathcal{S}_{i^*},\{\mathcal{S}_i:i\in\Omega(\boldsymbol{X})\}\}$, $\{\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*}-\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i:i\in\Omega(\boldsymbol{X})\}$ is multivariate normal by (EC.7). Moreover, for $i,i'\in\Omega(\boldsymbol{X})$ and $i\neq i'$, due to the conditional independence between $\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i$ and $\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i'}$,

$$\mathrm{Cov}\left(\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*}-\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i,\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*}-\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i'}\middle|\boldsymbol{X},\mathcal{S}_{i^*},\{\mathcal{S}_i:i\in\Omega(\boldsymbol{X})\}\right)=\mathrm{Var}\left(\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*}\middle|\boldsymbol{X},\mathcal{S}_{i^*}\right)>0.$$

Therefore, applying (EC.9) and Lemma EC.2,

$$\begin{aligned}\mathrm{PCS}(\boldsymbol{X}) &\geq \mathbb{E}\left[\prod_{i\in\Omega(\boldsymbol{X})}\mathbb{P}\left(\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*}-\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i>0\middle|\boldsymbol{X},\mathcal{S}_{i^*},\mathcal{S}_i\right)\middle|\boldsymbol{X}\right]\\ &\geq\mathbb{E}\left[\prod_{i\in\Omega(\boldsymbol{X})}\Phi\left(\frac{h_{\mathrm{Het}}}{\sqrt{(n_0-1)\sum_{j=1}^m a_j^2\left(1/\xi_{i^*j}+1/\xi_{ij}\right)}}\right)\middle|\boldsymbol{X}\right],\end{aligned}\tag{EC.10}$$

where the second inequality follows from (EC.8).

Notice that ξ_{ij}'s are i.i.d. $\chi^2_{n_0-1}$ random variables. Let $\xi_i^{(1)}=\min\{\xi_{i1},\ldots,\xi_{im}\}$ be their smallest order statistic. Then for each $i\in\Omega(\boldsymbol{X})$,

$$\sum_{j=1}^m a_j^2\left(1/\xi_{i^*j}+1/\xi_{ij}\right)\leq\sum_{j=1}^m a_j^2\left(1/\xi_{i^*}^{(1)}+1/\xi_i^{(1)}\right)=\left(1/\xi_{i^*}^{(1)}+1/\xi_i^{(1)}\right)\boldsymbol{a}^\intercal\boldsymbol{a}.\tag{EC.11}$$

It then follows from (EC.10) and (EC.11) that

$$\begin{aligned}\mathrm{PCS}(\boldsymbol{X}) &\geq \mathbb{E}\left[\prod_{i\in\Omega(\boldsymbol{X})}\Phi\left(\frac{h_{\mathrm{Het}}}{\sqrt{(n_0-1)(1/\xi_{i^*}^{(1)}+1/\xi_i^{(1)})\boldsymbol{a}^\intercal\boldsymbol{a}}}\right)\middle|\boldsymbol{X}\right]\\ &=\int_0^\infty\left[\int_0^\infty\Phi\left(\frac{h_{\mathrm{Het}}}{\sqrt{(n_0-1)(t^{-1}+s^{-1})\boldsymbol{a}^\intercal\boldsymbol{a}}}\right)\gamma_{(1)}(s)\mathrm{d}s\right]^{|\Omega(\boldsymbol{X})|}\gamma_{(1)}(t)\mathrm{d}t.\end{aligned}\tag{EC.12}$$

Since $0\leq\Phi(\cdot)\leq 1$ and $\gamma_{(1)}(\cdot)$ is a pdf, the integral inside the square brackets in (EC.12) is no greater than 1. Moreover, since $|\Omega(\boldsymbol{X})|\leq k-1$, hence,

$$\begin{aligned}\mathrm{PCS}(\boldsymbol{X}) &\geq \int_0^\infty\left[\int_0^\infty\Phi\left(\frac{h_{\mathrm{Het}}}{\sqrt{(n_0-1)(t^{-1}+s^{-1})\boldsymbol{a}^\intercal\boldsymbol{a}}}\right)\gamma_{(1)}(s)\mathrm{d}s\right]^{k-1}\gamma_{(1)}(t)\mathrm{d}t\\ &=\int_0^\infty\left[\int_0^\infty\Phi\left(\frac{h_{\mathrm{Het}}}{\sqrt{(n_0-1)(t^{-1}+s^{-1})\boldsymbol{X}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{X}}}\right)\gamma_{(1)}(s)\mathrm{d}s\right]^{k-1}\gamma_{(1)}(t)\mathrm{d}t,\end{aligned}$$

where the equality holds because

$$\boldsymbol{a}^\intercal \boldsymbol{a} = (\mathcal{X}(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{X})^\intercal \mathcal{X}(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{X} = \boldsymbol{X}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{X}.$$

It follows immediately from the definition of h_{Het} in (5) that $\text{PCS}_{\text{E}} = \mathbb{E}[\text{PCS}(\boldsymbol{X})] \geq 1-\alpha$. □

Remark EC.2. We have introduced the smallest order statistics in (EC.11) for computational feasibility. Without it, Procedure TS$^+$ would still be valid provided that we can compute the constant h_{Het} from the following equation,

$$\mathbb{E}\left\{\int_{\mathbb{R}_+^m}\left[\int_{\mathbb{R}_+^m} g(\boldsymbol{X}, h_{\text{Het}})\prod_{j=1}^m \gamma(s_j)\mathrm{d}s_1\cdots\mathrm{d}s_m\right]^{k-1}\prod_{j=1}^m\gamma(t_j)\mathrm{d}t_1\cdots\mathrm{d}t_m\right\} = 1-\alpha,$$

where

$$g(\boldsymbol{X}, h_{\text{Het}}) := \Phi\left(\frac{h_{\text{Het}}}{\sqrt{(n_0-1)\sum_{j=1}^m a_j^2(t_j^{-1}+s_j^{-1})}}\right).$$

However, it is prohibitively challenging to solve the above two equations numerically for $m \geq 3$. By introducing the smallest order statistic, we can instead solve (5) for h_{Het}, which is much easier computationally, while the price is h_{Het} will be a little larger then necessary as the lower bound of the PCS_{E} is further loosened. Also, in analogy to the discussion in §EC.2, when $\boldsymbol{X}$ is high-dimensional, h_{Het} in (5) can also be solved via Monte Carlo method or stochastic approximation method.

EC.5. Proof of Theorem 3

Proof of Theorem 3. First notice that, conditionally on S_i^2, $N_i \to \infty$ as $\delta \to 0$. Recall that $\widehat{\boldsymbol{\beta}}_i = \frac{1}{N_i}(\mathcal{X}^\intercal\mathcal{X})^{-1}\mathcal{X}^\intercal\sum_{\ell=1}^{N_i}\boldsymbol{Y}_{i\ell}$, and $\boldsymbol{Y}_{i\ell} = (Y_{i\ell}(\boldsymbol{x}_1),\ldots,Y_{i\ell}(\boldsymbol{x}_m))^\intercal$, $i=1,\ldots,k$. Under Assumption 3, $Y_{i\ell}(\boldsymbol{x})$ is independent of $Y_{i'\ell'}(\boldsymbol{x}')$ for any $(i,\ell,\boldsymbol{x}) \neq (i',\ell',\boldsymbol{x}')$; moreover, $Y_{i\ell}(\boldsymbol{x})$ and $Y_{i\ell'}(\boldsymbol{x})$ are identically distributed for $\ell = \ell'$. Recall that $N_i = \max\{\lceil h^2S_i^2/\delta^2\rceil, n_0\}$, and for small enough δ, $N_i = \lceil h^2 S_i^2/\delta^2\rceil$. We first establish the following convergence result by the central limit theorem, for each $i=1,\ldots,k$.

$$\frac{\sqrt{N_i}}{\sigma_i}\left(\frac{1}{N_i}\sum_{\ell=1}^{N_i}Y_{i\ell}(\boldsymbol{x}) - \boldsymbol{x}^\intercal\boldsymbol{\beta}_i\right)\Bigg| S_i^2 \Rightarrow Z, \tag{EC.13}$$

as $\delta \to 0$, where "⇒" denotes convergence in distribution, and Z is a standard normal random variable. To see (EC.13), we split the left-hand side of (EC.13) as follows.

$$\begin{aligned}
&\frac{\sqrt{N_i}}{\sigma_i}\left(\frac{1}{N_i}\sum_{\ell=1}^{N_i}Y_{i\ell}(\boldsymbol{x}) - \boldsymbol{x}^\intercal\boldsymbol{\beta}_i\right)\\
=&\frac{\sqrt{N_i}}{\sigma_i}\left\{\frac{n_0}{N_i}\left(\frac{1}{n_0}\sum_{\ell=1}^{n_0}Y_{i\ell}(\boldsymbol{x}) - \boldsymbol{x}^\intercal\boldsymbol{\beta}_i\right) + \frac{N_i-n_0}{N_i}\left(\frac{1}{N_i-n_0}\sum_{\ell=n_0+1}^{N_i}Y_{i\ell}(\boldsymbol{x}) - \boldsymbol{x}^\intercal\boldsymbol{\beta}_i\right)\right\}\\
=&\frac{n_0}{\sigma_i\sqrt{N_i}}\left(\frac{1}{n_0}\sum_{\ell=1}^{n_0}Y_{i\ell}(\boldsymbol{x}) - \boldsymbol{x}^\intercal\boldsymbol{\beta}_i\right) + \frac{\sqrt{N_i-n_0}}{\sqrt{N_i}}\frac{\sqrt{N_i-n_0}}{\sigma_i}\left(\frac{1}{N_i-n_0}\sum_{\ell=n_0+1}^{N_i}Y_{i\ell}(\boldsymbol{x}) - \boldsymbol{x}^\intercal\boldsymbol{\beta}_i\right).
\end{aligned}\tag{EC.14}$$

Conditionally on S_i^2, as $\delta \to 0$, $N_i \to \infty$, which implies that $\frac{n_0}{\sigma_i\sqrt{N_i}}\left(\frac{1}{n_0}\sum_{\ell=1}^{n_0} Y_{i\ell}(\boldsymbol{x}) - \boldsymbol{x}^\intercal\boldsymbol{\beta}_i\right) \to 0$ almost surely, $\frac{\sqrt{N_i-n_0}}{\sqrt{N_i}} \to 1$, and

$$\frac{\sqrt{N_i - n_0}}{\sigma_i}\left(\frac{1}{N_i - n_0}\sum_{\ell=n_0+1}^{N_i} Y_{i\ell}(\boldsymbol{x}) - \boldsymbol{x}^\intercal\boldsymbol{\beta}_i\right) \Rightarrow Z,$$

given by the central limit theorem. These three convergence results together with (EC.14) establish (EC.13).

It is then easy to see

$$\frac{\sqrt{N_i}}{\sigma_i}\left(\frac{1}{N_i}\sum_{\ell=1}^{N_i} \boldsymbol{Y}_{i\ell}(\boldsymbol{x}) - \mathcal{X}\boldsymbol{\beta}_i\right)\Big| S_i^2 \Rightarrow \boldsymbol{Z},$$

as $\delta \to 0$, where $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}, \mathcal{I})$ is a standard m-variate normal random vector. Hence,

$$\frac{\sqrt{N_i}}{\sqrt{V(\boldsymbol{X})}}\left(\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i - \boldsymbol{X}^\intercal\boldsymbol{\beta}_i\right)\Big|\{\boldsymbol{X}, S_i^2\} \Rightarrow \sigma_i Z, \tag{EC.15}$$

as $\delta \to 0$, where $V(\boldsymbol{X}) := \boldsymbol{X}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{X}$.

To simplify notation, we write $i^* = i^*(\mathbf{X})$ to temporarily suppress the dependence on $\mathbf{X}$. Let $\Omega(\boldsymbol{x}) := \{i : \boldsymbol{X}^\intercal\boldsymbol{\beta}_{i^*} - \boldsymbol{X}^\intercal\boldsymbol{\beta}_i \geq \delta | \boldsymbol{X} = \boldsymbol{x}\}$ be the set of alternatives outside the IZ given $\boldsymbol{X} = \boldsymbol{x}$. Let $U_{i^*} := \frac{\sqrt{N_{i^*}}}{\sqrt{V(\boldsymbol{X})}}\left(\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*} - \boldsymbol{X}^\intercal\boldsymbol{\beta}_{i^*}\right)$. Then, $U_{i^*}|\{\boldsymbol{X}, S_{i^*}^2\} \Rightarrow \sigma_{i^*} Z$, as $\delta \to 0$, by (EC.15). For $i \in \Omega(\boldsymbol{X})$, let

$$U_i := \frac{\sqrt{N_{i^*}}}{\sqrt{V(\boldsymbol{X})}}\left(\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i - \boldsymbol{X}^\intercal\boldsymbol{\beta}_i\right) = \frac{\sqrt{N_{i^*}}}{\sqrt{N_i}}\frac{\sqrt{N_i}}{\sqrt{V(\boldsymbol{X})}}\left(\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i - \boldsymbol{X}^\intercal\boldsymbol{\beta}_i\right).$$

Then, $U_i|\{\boldsymbol{X}, S_{i^*}^2, S_i^2\} \Rightarrow \frac{S_{i^*}}{S_i}\sigma_i Z$, as $\delta \to 0$, due to (EC.15) and that $\sqrt{N_{i^*}}/\sqrt{N_i} \to S_{i^*}/S_i$ as $\delta \to 0$.

For notational simplicity, we temporarily let s denote the cardinality of $\Omega(\boldsymbol{X})$, and refer to the s alternatives in $\Omega(\boldsymbol{X})$ as alternatives $1, \ldots, s$, without loss of generality. As $U_{i^*}, U_1, \ldots, U_s$ are independent of each other given $\{\boldsymbol{X}, S_{i^*}^2, S_1^2, \ldots, S_k^2\}$, as $\delta \to 0$,

$$(U_{i^*}, U_1, \ldots, U_s)^\intercal\big|\{\boldsymbol{X}, S_{i^*}^2, S_1^2, \ldots, S_s^2\} \Rightarrow \left(\sigma_{i^*}Z_0, \frac{S_{i^*}}{S_1}\sigma_1 Z_1, \ldots, \frac{S_{i^*}}{S_s}\sigma_s Z_s\right)^\intercal,$$

where $Z_0, Z_1, \ldots, Z_s$ are independent standard normal random variables. Hence, by the continuous mapping theorem, as $\delta \to 0$,

$$(U_{i^*} - U_1, \ldots, U_{i^*} - U_s)^\intercal\big|\{\boldsymbol{X}, S_{i^*}^2, S_1^2, \ldots, S_s^2\} \Rightarrow \left(\sigma_{i^*}Z_0 - \frac{S_{i^*}}{S_1}\sigma_1 Z_1, \ldots, \sigma_{i^*}Z_0 - \frac{S_{i^*}}{S_s}\sigma_s Z_s\right)^\intercal, \tag{EC.16}$$

where the limit is multivariate normal, and for $i, j \in \{1, \ldots, s\}$ and $i \neq j$,

$$\mathrm{Cov}\left(\sigma_{i^*}Z_0 - \frac{S_{i^*}}{S_i}\sigma_i Z_i,\ \sigma_{i^*}Z_0 - \frac{S_{i^*}}{S_j}\sigma_j Z_j \Big| \boldsymbol{X}, S_{i^*}^2, S_1^2, \ldots, S_s^2\right) = \sigma_{i^*}^2 > 0. \tag{EC.17}$$

Now we have

$$
\begin{aligned}
&\liminf_{\delta\to 0} \mathrm{PCS}(\boldsymbol{X}) \\
&\geq \liminf_{\delta\to 0} \mathbb{P}\left(\bigcap_{i\in\Omega(\boldsymbol{X})} \left\{\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*} - \boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i > 0\right\}\middle|\boldsymbol{X}\right) && \text{(EC.18)}\\
&= \liminf_{\delta\to 0} \mathbb{E}\left[\mathbb{P}\left(\bigcap_{i\in\Omega(\boldsymbol{X})} \left\{\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*} - \boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i > 0\right\}\middle|\boldsymbol{X}, S_{i^*}^2, \{S_i^2 : i\in\Omega(\boldsymbol{X})\}\right)\middle|\boldsymbol{X}\right] && \text{(EC.19)}\\
&\geq \mathbb{E}\left[\liminf_{\delta\to 0} \mathbb{P}\left(\bigcap_{i\in\Omega(\boldsymbol{X})} \left\{\boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_{i^*} - \boldsymbol{X}^\intercal\widehat{\boldsymbol{\beta}}_i > 0\right\}\middle|\boldsymbol{X}, S_{i^*}^2, \{S_i^2 : i\in\Omega(\boldsymbol{X})\}\right)\middle|\boldsymbol{X}\right] && \text{(EC.20)}\\
&= \mathbb{E}\left[\liminf_{\delta\to 0} \mathbb{P}\left(\bigcap_{i\in\Omega(\boldsymbol{X})} \left\{U_{i^*} - U_i > \frac{-(\boldsymbol{X}^\intercal\boldsymbol{\beta}_{i^*} - \boldsymbol{X}^\intercal\boldsymbol{\beta}_i)}{\sqrt{V(\boldsymbol{X})/N_{i^*}}}\right\}\middle|\boldsymbol{X}, S_{i^*}^2, \{S_i^2 : i\in\Omega(\boldsymbol{X})\}\right)\middle|\boldsymbol{X}\right] && \text{(EC.21)}\\
&= \mathbb{E}\left[\mathbb{P}\left(\bigcap_{i\in\Omega(\boldsymbol{X})} \left\{\sigma_{i^*}Z_0 - \frac{S_{i^*}}{S_i}\sigma_i Z_i > \frac{-(\boldsymbol{X}^\intercal\boldsymbol{\beta}_{i^*} - \boldsymbol{X}^\intercal\boldsymbol{\beta}_i)}{\sqrt{V(\boldsymbol{X})/N_{i^*}}}\right\}\middle|\boldsymbol{X}, S_{i^*}^2, \{S_i^2 : i\in\Omega(\boldsymbol{X})\}\right)\middle|\boldsymbol{X}\right] && \text{(EC.22)}\\
&\geq \mathbb{E}\left[\prod_{i\in\Omega(\boldsymbol{X})} \mathbb{P}\left(\sigma_{i^*}Z_0 - \frac{S_{i^*}}{S_i}\sigma_i Z_i > \frac{-(\boldsymbol{X}^\intercal\boldsymbol{\beta}_{i^*} - \boldsymbol{X}^\intercal\boldsymbol{\beta}_i)}{\sqrt{V(\boldsymbol{X})/N_{i^*}}}\middle|\boldsymbol{X}, S_{i^*}^2, S_i^2\right)\middle|\boldsymbol{X}\right] && \text{(EC.23)}\\
&\geq \mathbb{E}\left[\prod_{i\in\Omega(\boldsymbol{X})} \mathbb{P}\left(\left(\sigma_{i^*}^2 + \frac{S_{i^*}^2}{S_i^2}\sigma_i^2\right)^{1/2} Z > \frac{-\delta}{\sqrt{V(\boldsymbol{X})\delta^2/(h^2 S_{i^*}^2)}}\middle|\boldsymbol{X}, S_{i^*}^2, S_i^2\right)\middle|\boldsymbol{X}\right] && \text{(EC.24)}\\
&= \mathbb{E}\left[\prod_{i\in\Omega(\boldsymbol{X})} \mathbb{P}\left(Z > \frac{-h}{\sqrt{[\sigma_{i^*}^2/S_{i^*}^2 + \sigma_i^2/S_i^2]V(\mathbf{X})}}\right)\middle|\boldsymbol{X}\right]\\
&= \mathbb{E}\left[\prod_{i\in\Omega(\boldsymbol{X})} \Phi\left(\frac{h}{\sqrt{(n_0 m - d)(\xi_{i^*}^{-1} + \xi_i^{-1})V(\boldsymbol{X})}}\right)\middle|\boldsymbol{X}\right],
\end{aligned}
$$

where (EC.18) holds because the CS event must occur if alternative i^* eliminates all alternatives in $\Omega(\boldsymbol{X})$, (EC.19) is due to the tower law of conditional expectation, (EC.20) is due to Fatou's Lemma, (EC.21) is by the definitions of U_{i^*} and U_i, (EC.22) is by (EC.16), (EC.23) is obtained by Lemma EC.2 together with (EC.17), and (EC.24) follows from the definitions of $\Omega(\boldsymbol{X})$ and N_{i^*}.

The rest of the proof follows the same argument as that in the proof of Theorem 1. □

EC.6. Choosing $\mathrm{PCS}_{\min}$ as Target

EC.6.1. Two-Stage Procedures

If we use $\mathrm{PCS}_{\min} = \min_{\boldsymbol{x}\in\Theta} \mathrm{PCS}(\boldsymbol{x})$ to measure correct selection across the population, and set the pre-specified target as $\mathrm{PCS}_{\min} \geq 1-\alpha$, instead of $\mathrm{PCS}_{\mathrm{E}} \geq 1-\alpha$, we are in a more conservative case wherein we require the selection policy produced by the selection procedure to make correct selection with probability at least $1-\alpha$ for *all* values of the covariates. In this case, both Procedure TS and Procedure TS$^+$ can be revised slightly to retain statistical validity under the new criterion. In particular, we only need change the definition of the constant h (resp., h_{Het}) in Procedure TS (resp., Procedure TS$^+$), while keeping the other parts of the procedure the same. The following results are parallel to those for PCS_E, that is, Theorems 1–4. The proofs are essentially the same and thus we omit the details.

THEOREM EC.1. *Suppose that Procedure TS is used to solve the R&S-C problem with the constant h in the procedure being solved from*

$$\min_{\boldsymbol{x}\in\Theta}\left\{\int_0^\infty\left[\int_0^\infty \Phi\left(\frac{h}{\sqrt{(n_0m-d)(t^{-1}+s^{-1})\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}}}\right)\eta(s)\mathrm{d}s\right]^{k-1}\eta(t)\mathrm{d}t\right\}=1-\alpha. \quad \text{(EC.25)}$$

- *If Assumption 1 is satisfied, then* $\text{PCS}_{\min}\ge 1-\alpha$.
- *If Assumption 3 is satisfied, then* $\liminf_{\delta\to 0}\text{PCS}_{\min}\ge 1-\alpha$.

THEOREM EC.2. *Suppose that Procedure TS$^+$ is used to solve the R&S-C problem with the constant h_{Het} in the procedure being solved from*

$$\min_{\boldsymbol{x}\in\Theta}\left\{\int_0^\infty\left[\int_0^\infty \Phi\left(\frac{h_{\text{Het}}}{\sqrt{(n_0-1)(t^{-1}+s^{-1})\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}}}\right)\gamma_{(1)}(s)\mathrm{d}s\right]^{k-1}\gamma_{(1)}(t)\mathrm{d}t\right\}=1-\alpha. \quad \text{(EC.26)}$$

- *If Assumption 2 is satisfied, then* $\text{PCS}_{\min}\ge 1-\alpha$.
- *If Assumption 4 is satisfied, then* $\liminf_{\delta\to 0}\text{PCS}_{\min}\ge 1-\alpha$.

REMARK EC.3. It is computationally easier to solve for h from (EC.25) than from (4). First, there is no need to compute the expectation with respect to the distribution of $\mathbf{x}$ in (EC.25), which amounts to multidimensional numerical integration. Second, note that the minimizer of the left-hand side of (EC.25) is the same as the maximizer of $\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}$, since the function $\Phi(\cdot)$ is increasing. Since $\mathcal{X}^\intercal\mathcal{X}$ is nonsingular, it is easy to see that $\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}$ is convex in $\boldsymbol{x}$. Thus, if Θ is a bounded closed set, the maximizer must lie in the set of all extreme points of the convex hull of Θ; see, for example, Theorem 32.2 and Corollary 32.3.3 in Rockafellar (1970). A similar argument can be made for the computation of h_{Het} by comparing (EC.26) with (5).

EC.6.2. Numerical Results

Define the achieved $\text{PCS}_{\min}$ as

$$\widehat{\text{PCS}}_{\min}:=\min_{\boldsymbol{x}\in\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_T\}}\frac{1}{R}\sum_{r=1}^R\mathbb{I}\left\{\mu_{i^*(\boldsymbol{x})}(\boldsymbol{x})-\mu_{\widehat{i_r^*}(\boldsymbol{x})}(\boldsymbol{x})<\delta\right\}.$$

Table EC.1 reports $\widehat{\mathrm{PCS}}_{\mathrm{E}}$ and $\widehat{\mathrm{PCS}}_{\min}$ when the target is $\mathrm{PCS}_{\min} \geq 95\%$, while Table EC.2 reports the case when the target is $\mathrm{PCS}_{\mathrm{E}} \geq 95\%$.

First, results in Table EC.1 show that Procedure TS and Procedure TS$^+$ with h and h_{Het} computed from (EC.25) and (EC.26), respectively, can deliver the target $\mathrm{PCS}_{\min}$ in their respective domains. In particular, Procedure TS using h in (EC.25) can deliver the target $\mathrm{PCS}_{\min}$ if the simulation errors are homoscedastic, while Procedure TS$^+$ using h in (EC.26) can do the same even when the simulation errors are heteroscedastic. Moreover, the achieved $\mathrm{PCS}_{\min}$ is higher than

Table EC.1 Results When the Target is $\mathrm{PCS}_{\min} \geq 95\%$.

	Procedure TS (using h in (EC.25))				Procedure TS$^+$ (using h in (EC.26))			
Problem	h	Sample	$\widehat{\mathrm{PCS}}_{\mathrm{E}}$	$\widehat{\mathrm{PCS}}_{\min}$	h_{Het}	Sample	$\widehat{\mathrm{PCS}}_{\mathrm{E}}$	$\widehat{\mathrm{PCS}}_{\min}$
(0) Benchmark	5.927	140,543	0.9989	0.9609	6.990	195,337	0.9997	0.9840
(1) $k=2$	4.362	30,447	0.9958	0.9481	5.132	42,164	0.9987	0.9709
(2) $k=8$	6.481	268,749	0.9993	0.9657	7.651	374,716	0.9999	0.9852
(3) Non-GSC	5.927	140,542	1.0000	0.9958	6.990	195,337	1.0000	0.9980
(4) IV	5.927	158,139	0.9989	0.9590	6.990	219,871	0.9998	0.9869
(5) DV	5.927	158,100	0.9990	0.9628	6.990	219,741	0.9998	0.9837
(6) Het	5.927	175,698	0.9952	[0.9032]	6.990	244,488	0.9999	**0.9904**
(7) $d=2$	7.155	51,161	0.9954	0.9600	7.648	58,493	0.9971	0.9708
(8) $d=6$	3.792	230,221	0.9994	0.9667	4.804	369,307	1.0000	0.9944
(9) Normal Dist	5.927	140,550	0.9990	0.9623	6.990	195,404	0.9997	0.9851
(10) $k=100$	7.385	3,272,127	0.9999	0.9754	8.678	4,518,029	1.0000	0.9941
(11) $d=50$	9.444	4,370,569	1.0000	0.9998	12.631	7,818,201	1.0000	1.0000
(12) $k=100$, $d=50$	13.875	1.89×10^8	1.0000	1.0000	18.970	3.53×10^8	1.0000	1.0000

Note. In the presence of heteroscedasticity, the boxed number suggests that Procedure TS fails to deliver the target $\mathrm{PCS}_{\min}$, whereas the bold number suggests that Procedure TS$^+$ succeeds to do so.

Table EC.2 Results When the Target is $\mathrm{PCS}_{\mathrm{E}} \geq 95\%$.

	Procedure TS (using h in (4))				Procedure TS$^+$ (using h in (5))			
Problem	h	Sample	$\widehat{\mathrm{PCS}}_{\mathrm{E}}$	$\widehat{\mathrm{PCS}}_{\min}$	h_{Het}	Sample	$\widehat{\mathrm{PCS}}_{\mathrm{E}}$	$\widehat{\mathrm{PCS}}_{\min}$
(0) Benchmark	3.423	46,865	0.9610	0.7476	4.034	65,138	0.9801	0.8120
(1) $k=2$	2.363	8,947	0.9501	0.8094	2.781	12,380	0.9702	0.8541
(2) $k=8$	3.822	93,542	0.9650	0.7290	4.510	130,200	0.9842	0.8098
(3) Non-GSC	3.423	46,865	0.9987	0.9400	4.034	65,138	0.9994	0.9599
(4) IV	3.423	52,698	0.9618	0.7589	4.034	73,265	0.9807	0.8184
(5) DV	3.423	52,720	0.9614	0.7544	4.034	73,246	0.9806	0.8143
(6) Het	3.423	58,626	[0.9232]	0.6368	4.034	81,555	**0.9846**	0.8625
(7) $d=2$	4.612	21,288	0.9593	0.7941	4.924	24,266	0.9662	0.8223
(8) $d=6$	2.141	73,428	0.9656	0.7662	2.710	117,626	0.9895	0.8589
(9) Normal Dist	3.447	47,529	0.9626	0.7579	4.063	66,061	0.9821	0.8230
(10) $k=100$	4.346	1,133,384	0.9758	0.5952	5.117	1,570,911	0.9918	0.7218
(11) $d=50$	3.222	508,977	0.9583	0.7522	4.312	911,326	0.9926	0.8749
(12) $k=100$, $d=50$	4.886	23,400,677	0.9765	0.6189	6.702	44,024,486	0.9991	0.8854

Note. In the presence of heteroscedasticity, the boxed number suggests that Procedure TS fails to deliver the target $\mathrm{PCS}_{\mathrm{E}}$, whereas the bold number suggests that Procedure TS$^+$ succeeds to do so.

the target in general; see, e.g., the column "$\widehat{\text{PCS}}_{\text{min}}$" under "Procedure TS" of Table EC.1, except the entry for Problem (6). This kind of conservativeness is also observed in Table EC.2, where Procedure TS using h in (4) and Procedure TS$^+$ using h in (5) are used when the objective is to meet the target PCS_{E}.

Second, the numerical results show that PCS_{min} is a much more conservative criterion than PCS_{E}. In particular, if the target is $\text{PCS}_{\text{E}} \geq 1-\alpha$, then $\widehat{\text{PCS}}_{\text{min}}$ is significantly lower than $1-\alpha$, except for Problem (3), in which the non-GSC amplifies the procedures' conservativeness stemming from the IZ formulation and provides the "extra" sample size needed for making $\widehat{\text{PCS}}_{\text{min}}$ reach the target; see Table EC.2. By contrast, if the target is $\text{PCS}_{\text{min}} \geq 1-\alpha$, then $\widehat{\text{PCS}}_{\text{E}}$ is virtually 1 for each problem-procedure combination; see Table EC.1. Another indication of the conservativeness of PCS_{min} is that in each problem-procedure combination, the sample size when using PCS_{min} as the criterion is about three times larger than that when using PCS_{E}. For example, in Table EC.2 the sample size for Problem (0) with Procedure TS is 46,865, whereas the corresponding sample size in Table EC.1 is 140,543.

EC.7. Asymptotic Sample Size Analysis

For ease of presentation, we relax the integrality constraint of the sample size, but it has no essential impact on the asymptotic analysis of the sample size.

EC.7.1. Procedure TS

The expected total sample size of Procedure TS is

$$N_{\text{TS}} = \mathbb{E}\left[\sum_{i=1}^{k} m N_i\right] = m h^2 \sum_{i=1}^{k} \mathbb{E}\left[\max\left\{S_i^2/\delta^2, n_0/h^2\right\}\right], \tag{EC.27}$$

where h is solved from (4) if PCS_{E} is used as the criterion, whereas from (EC.25) if PCS_{min} is used. We will provide an asymptotic upper bound on N_{TS} in two asymptotic regimes, i.e., as $k \to \infty$ and as $\alpha \to 0$. Note that the expression of N_{TS} involves both h^2 and $1/h^2$, which depend on k and α. We first establish in Lemma EC.3 both a lower bound and a upper bound on h. Its proof is deferred to §EC.7.3.

Lemma EC.3. *Let h be the constant solved from either* (4) *or* (EC.25)*, and let $\underline{\alpha} \in (0, 1/2)$ be a constant. Then,*

$$0 < \underline{h} \leq h \leq \left\{ 2(n_0 m - d)\left[\left(\frac{2(k-1)}{\alpha}\right)^{\frac{2}{n_0 m - d}} - 1\right] \times \max_{\boldsymbol{x}\in\Theta} \boldsymbol{x}^\intercal (\mathcal{X}^\intercal \mathcal{X})^{-1} \boldsymbol{x} \right\}^{1/2},$$

for all $k \geq 2$ and $\alpha \leq \underline{\alpha}$, where $\underline{h}$ is a solution of (4) *for $k=2$ and $\alpha = \underline{\alpha}$.*

Then, it follows immediately from (EC.27) and Lemma EC.3 that for any k and small α,

$$N_{\rm TS} \le 2m(n_0m-d)\Big[\Big(\frac{2(k-1)}{\alpha}\Big)^{\frac{2}{n_0m-d}}-1\Big]\times\max_{\boldsymbol{x}\in\Theta}\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}\times\sum_{i=1}^k\mathbb{E}\big[\max\big\{S_i^2/\delta^2, n_0/\underline{h}^2\big\}\big]$$
$$\le C_{\rm TS}\times k\Big(\frac{k}{\alpha}\Big)^{\frac{2}{n_0m-d}},$$

where $C_{\rm TS}$ is a constant independent of k and α, given by

$$C_{\rm TS} = 2m(n_0m-d)\times 2^{\frac{2}{n_0m-d}}\times\max_{\boldsymbol{x}\in\Theta}\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}\times\max_{1\le i\le k}\mathbb{E}\big[\max\big\{S_i^2/\delta^2, n_0/\underline{h}^2\big\}\big].$$

Hence, we conclude that $N_{\rm TS}=\mathcal{O}(k^{1+\frac{2}{n_0m-d}})$ as $k\to\infty$, and $N_{\rm TS}=\mathcal{O}(\alpha^{-\frac{2}{n_0m-d}})$ as $\alpha\to 0$.

EC.7.2. Procedure TS$^+$

The expected total sample size of Procedure TS$^+$ is

$$N_{\rm TS^+}=\mathbb{E}\Big[\sum_{i=1}^k\sum_{j=1}^m N_{ij}\Big]=h_{\rm Het}^2\sum_{i=1}^k\sum_{j=1}^m\mathbb{E}\big[\max\big\{S_{ij}^2/\delta^2, n_0/h_{\rm Het}^2\big\}\big], \tag{EC.28}$$

where $h_{\rm Het}$ is solved from (5) if $\rm PCS_E$ is used as the criterion, whereas from (EC.26) if $\rm PCS_{min}$ is used. Similar to the analysis in §EC.7.1, we first give bounds on $h_{\rm Het}$ in Lemma EC.4. The proof is presented in §EC.7.4.

LEMMA EC.4. *Let $h_{\rm Het}$ be the constant solved either from* (5) *or* (EC.26)*, and let $\underline{\alpha}\in(0,1/2)$ be a constant. Then,*

$$0<\underline{h_{\rm Het}}\le h_{\rm Het}\le\Big\{2(n_0-1)\Big[\Big(\frac{2m(k-1)}{\alpha}\Big)^{\frac{2}{n_0-1}}-1\Big]\times\max_{\boldsymbol{x}\in\Theta}\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}\Big\}^{1/2},$$

for all $k\ge 2$ and $\alpha\le\underline{\alpha}$, where $\underline{h_{\rm Het}}$ is a solution of (5) *when $k=2$ and $\alpha=\underline{\alpha}$.*

Then, it follows from (EC.27) and Lemma EC.4 that for any k and small α,

$$N_{\rm TS^+}\le 2(n_0-1)\Big[\Big(\frac{2m(k-1)}{\alpha}\Big)^{\frac{2}{n_0-1}}-1\Big]\times\max_{\boldsymbol{x}\in\Theta}\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}\times\sum_{i=1}^k\sum_{j=1}^m\mathbb{E}\big[\max\big\{S_{ij}^2/\delta^2, n_0/\underline{h_{\rm Het}}^2\big\}\big],$$
$$\le C_{\rm TS^+}\times k\Big(\frac{k}{\alpha}\Big)^{\frac{2}{n_0-1}},$$

where $C_{\rm TS^+}$ is a constant independent of k and α, given by

$$C_{\rm TS^+}=2(n_0-1)\times(2m)^{\frac{2}{n_0-1}}\times\max_{\boldsymbol{x}\in\Theta}\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}\times\max_{1\le i\le k}\sum_{j=1}^m\mathbb{E}\big[\max\big\{S_{ij}^2/\delta^2, n_0/\underline{h_{\rm Het}}^2\big\}\big].$$

Hence, we conclude that $N_{\rm TS^+}=\mathcal{O}(k^{1+\frac{2}{n_0-1}})$ as $k\to\infty$, and $N_{\rm TS}=\mathcal{O}(\alpha^{-\frac{2}{n_0-1}})$ as $\alpha\to 0$.

REMARK EC.4. It is straightforward to see that with the same design matrix $\mathcal{X}$ and initial sample size n_0, $N_{\rm TS^+}=\mathcal{O}(k^{1+\frac{2}{n_0-1}}\alpha^{-\frac{2}{n_0-1}})$ has a higher order of magnitude than $N_{\rm TS}=\mathcal{O}(k^{1+\frac{2}{n_0m-d}}\alpha^{-\frac{2}{n_0m-d}})$ as $k\to\infty$ or $\alpha\to 0$, since $n_0m-d\ge n_0-1$ for all $m\ge d\ge 1$.

REMARK EC.5. For Procedure TS$^+$, m is not involved in the order of magnitude of N_{TS^+} as $k\to\infty$ or $\alpha\to 0$, but only takes effect in the leading constant C_{TS^+}. By contrast, a larger value of m leads to a lower order of magnitude of N_{TS} for Procedure TS. An intuitive explanation for the above difference is that, increasing m will result in a more accurate estimation of the common variance σ_i^2 in Procedure TS, while it does not affect estimation of the variances in Procedure TS$^+$ since they are estimated separately. This suggests that Procedure TS$^+$ for the linear models will favor the minimal m, that is, $m=d$.

EC.7.3. Proof of Lemma EC.3

We first prove the lower bound. Let

$$f(\boldsymbol{x},h):=\int_0^\infty\left[\int_0^\infty \Phi\left(\frac{h}{\sqrt{(n_0m-d)(t^{-1}+s^{-1})\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}}}\right)\eta(s)\mathrm{d}s\right]^{k-1}\eta(t)\mathrm{d}t.$$

Then, h solved from (4) satisfies $\mathbb{E}[f(\boldsymbol{X},h)]=1-\alpha$, whereas h solved from (EC.25) satisfies $\min_{\boldsymbol{x}\in\Theta} f(\boldsymbol{x},h)=1-\alpha$. Note that both $\mathbb{E}[f(\boldsymbol{X},h)]$ and $\min_{\boldsymbol{x}\in\Theta} f(\boldsymbol{x},h)$ are increasing functions in h, so a smaller α will yield a larger h. It is also clear that a larger k will yield a larger h. Hence, $\underline{h}$ defined in Lemma EC.3 is smaller than h solved from (4) for all $k\ge 2$ and $\alpha\le\underline{\alpha}$. Note that h solved from (4) is smaller than h solved from (EC.25) with everything else the same. The lower bound is then established.

The proof for the upper bound is similar to the proof of Lemma 4 in Zhong and Hong (2020). Specifically, let

$$h^*:=\left\{2(n_0m-d)\Big[\Big(\frac{2(k-1)}{\alpha}\Big)^{\frac{2}{n_0m-d}}-1\Big]\times\max_{\boldsymbol{x}\in\Theta}\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}\right\}^{1/2}.$$

To show that h, which is solved from either (4) or (EC.25), is no larger than h^*, it suffices to show that $\mathbb{E}[f(\boldsymbol{X},h^*)]\ge 1-\alpha$ and $\min_{\boldsymbol{x}\in\Theta} f(\boldsymbol{x},h^*)\ge 1-\alpha$, which is clearly true if we can show that $f(\boldsymbol{x},h^*)\ge 1-\alpha$ for any $\boldsymbol{x}\in\Theta$.

Let $Z_1,\dots,Z_{k-1}$ be $(k-1)$ independent standard normal random variables. Let $\xi_1,\dots,\xi_k$ be k independent chi-square random variables with (n_0m-d) degrees of freedom. Moreover, assume that Z_i is independent of $\xi_{i'}$, for $1\le i\le k-1$, $1\le i'\le k$. Then, for any $\boldsymbol{x}\in\Theta$,

$$\begin{aligned}
f(\boldsymbol{x},h^*)&=\mathbb{E}\left[\mathbb{P}\left(\bigcap_{i=1}^{k-1}\left\{Z_i\le\frac{h^*}{\sqrt{(n_0m-d)(\xi_k^{-1}+\xi_i^{-1})\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}}}\right\}\middle|\xi_k\right)\right]\\
&\ge\mathbb{E}\left[1-\sum_{i=1}^{k-1}\mathbb{P}\left(Z_i>\frac{h^*}{\sqrt{(n_0m-d)(\xi_k^{-1}+\xi_i^{-1})\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}}}\middle|\xi_k\right)\right]\\
&=1-(k-1)\times\mathbb{P}\left(Z_1>\frac{h^*}{\sqrt{(n_0m-d)(\xi_k^{-1}+\xi_1^{-1})\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}}}\right).
\end{aligned}\tag{EC.29}$$

By the Chernoff bound, $\mathbb{P}(Z_1 > a) \le \exp\{-\frac{a^2}{2}\}$ for all $a > 0$. Hence,

$$\begin{aligned}
&\mathbb{P}\left(Z_1 > \frac{h^*}{\sqrt{(n_0m-d)(\xi_k^{-1}+\xi_1^{-1})\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}}}\right)\\
\le &\mathbb{E}\left[\exp\left\{-\frac{(h^*)^2}{2(n_0m-d)(\xi_k^{-1}+\xi_1^{-1})\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}}\right\}\right]\\
\le &\mathbb{E}\left[\exp\left\{-\frac{\left(\frac{2(k-1)}{\alpha}\right)^{\frac{2}{n_0m-d}}-1}{\xi_k^{-1}+\xi_1^{-1}}\right\}\right]\\
\le &\mathbb{E}\left[\exp\left\{-\frac{\left(\frac{2(k-1)}{\alpha}\right)^{\frac{2}{n_0m-d}}-1}{2}\xi_{(1)}\right\}\right],
\end{aligned} \tag{EC.30}$$

where $\xi_{(1)} := \min\{\xi_k, \xi_1\}$ is the smallest order statistic of two independent chi-square random variables with (n_0m-d) degrees of freedom. Here, the second inequality holds by the definition of h^*. Let $f_{n_0m-d}(\cdot)$ and $F_{n_0m-d}(\cdot)$ denote the pdf and cdf of the chi-square random variables with (n_0m-d) degrees of freedom, respectively. Then the pdf of $\xi_{(1)}$ is known as $2f_{n_0m-d}(t)(1-F_{n_0m-d}(t))$. Hence, following (EC.30), we have

$$\begin{aligned}
&\mathbb{P}\left(Z_1 > \frac{h^*}{\sqrt{(n_0m-d)(\xi_k^{-1}+\xi_1^{-1})\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}}}\right)\\
\le &\int_0^\infty \exp\left\{-\frac{\left(\frac{2(k-1)}{\alpha}\right)^{\frac{2}{n_0m-d}}-1}{2}t\right\}\times 2f_{n_0m-d}(t)(1-F_{n_0m-d}(t))\mathrm{d}t\\
\le &\,2\int_0^\infty \exp\left\{-\frac{\left(\frac{2(k-1)}{\alpha}\right)^{\frac{2}{n_0m-d}}-1}{2}t\right\}f_{n_0m-d}(t)\mathrm{d}t\\
= &\,2\mathbb{E}\left[\exp\left\{-\frac{\left(\frac{2(k-1)}{\alpha}\right)^{\frac{2}{n_0m-d}}-1}{2}\xi_1\right\}\right]\\
= &\,2\left[1+\left(\frac{2(k-1)}{\alpha}\right)^{\frac{2}{n_0m-d}}-1\right]^{-(n_0m-d)/2}\\
= &\,\alpha/(k-1),
\end{aligned} \tag{EC.31}$$

where the second equality is due to the moment generating function of the chi-square random variables with (n_0m-d) degrees of freedom. Combining (EC.29) and (EC.31), we can conclude that $f(\boldsymbol{x}, h^*) \ge 1-\alpha$ for any $\boldsymbol{x} \in \Theta$, which completes the proof. □

EC.7.4. Proof of Lemma EC.4

The lower bound can be proved using the same argument for proving the lower bound in Lemma EC.3. Let

$$h^*_{\text{Het}} := \left\{2(n_0-1)\left[\left(\frac{2m(k-1)}{\alpha}\right)^{\frac{2}{n_0-1}}-1\right]\times\max_{\boldsymbol{x}\in\Theta}\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}\right\}^{1/2},$$

and let

$$f(\boldsymbol{x}, h_{\text{Het}}) := \int_0^\infty \left[\int_0^\infty \Phi\left(\frac{h_{\text{Het}}}{\sqrt{(n_0-1)(t^{-1}+s^{-1})\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}}} \right) \gamma_{(1)}(s)\mathrm{d}s \right]^{k-1} \gamma_{(1)}(t)\mathrm{d}t.$$

It then suffices to show that $f(\boldsymbol{x}, h^*_{\text{Het}}) \geq 1-\alpha$ for any $\boldsymbol{x} \in \Theta$, in order to prove the upper bound.

Let $Z_1, \ldots, Z_{k-1}$ be $(k-1)$ independent standard normal random variables. Let $\xi_1, \ldots, \xi_k$ be k independent random variables, each of which is the smallest order statistic of m chi-square random variables with (n_0-1) degrees of freedom. Moreover, assume that Z_i is independent of $\xi_{i'}$, for $1 \leq i \leq k-1$, $1 \leq i' \leq k$. With the same argument leading to (EC.29), we have

$$f(\boldsymbol{x}, h^*_{\text{Het}}) \geq 1-(k-1) \times \mathbb{P}\left(Z_1 > \frac{h^*_{\text{Het}}}{\sqrt{(n_0-1)(\xi_k^{-1}+\xi_1^{-1})\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}}} \right). \tag{EC.32}$$

With the same argument leading to (EC.30), we have

$$\mathbb{P}\left(Z_1 > \frac{h^*_{\text{Het}}}{\sqrt{(n_0-1)(\xi_k^{-1}+\xi_1^{-1})\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}}} \right) \leq \mathbb{E}\left[\exp\left\{ -\frac{\left(\frac{2m(k-1)}{\alpha}\right)^{\frac{2}{n_0-1}}-1}{2}\xi_{(1)} \right\} \right],$$

where $\xi_{(1)} := \min\{\xi_k, \xi_1\}$ is the smallest order statistic of $2m$ independent chi-square random variables with (n_0-1) degrees of freedom, and its pdf is $2m f_{n_0-1}(t)(1-F_{n_0-1}(t))^{2m-1}$. Then, with the same argument leading to (EC.31), we have

$$\begin{aligned}\mathbb{P}\left(Z_1 > \frac{h^*_{\text{Het}}}{\sqrt{(n_0-1)(\xi_k^{-1}+\xi_1^{-1})\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}}} \right) &\leq 2m\int_0^\infty \exp\left\{ -\frac{\left(\frac{2m(k-1)}{\alpha}\right)^{\frac{2}{n_0-1}}-1}{2}t \right\} f_{n_0-1}(t)\mathrm{d}t \\ &= 2m\mathbb{E}\left[\exp\left\{ -\frac{\left(\frac{2m(k-1)}{\alpha}\right)^{\frac{2}{n_0-1}}-1}{2}\xi_0 \right\} \right] \\ &= \alpha/(k-1), \end{aligned} \tag{EC.33}$$

where ξ_0 is a chi-square random variables with (n_0-1) degrees of freedom. Combining (EC.32) and (EC.33), we can conclude that $f(\boldsymbol{x}, h^*_{\text{Het}}) \geq 1-\alpha$ for any $\boldsymbol{x} \in \Theta$, which completes the proof. $\square$

EC.8. Proof of Theorem 5

Theorem 5 can be viewed as a corollary of the following Theorem EC.3. Therefore we only provide the proof of Theorem EC.3 and remark that Theorem 5 holds immediately.

Theorem EC.3. *Let N_{ij} denote the number of samples of alternative i taken at design point $\boldsymbol{x}_j$, and $\widehat{Y}_{ij}$ denote their means, for $i=1,\ldots,k$, $j=1,\ldots,m$. Let $\widehat{\boldsymbol{Y}}_i = (\widehat{Y}_{i1}, \ldots, \widehat{Y}_{im})^\intercal$ and $\widehat{\boldsymbol{\beta}}_i = (\mathcal{X}^\intercal\mathcal{X})^{-1}\mathcal{X}^\intercal\widehat{\boldsymbol{Y}}_i$ for $i=1,\ldots,k$. Under Assumption 1 or 2, the GSC defined in (7) is the LFC for a selection procedure of the R&S-C problem with the IZ formulation and a fixed design, if all the following properties hold:*

(i) The selected alternative is $\widehat{i^*}(\boldsymbol{x}) = \arg\max_{1\le i\le k} \left\{\boldsymbol{x}^\intercal \widehat{\boldsymbol{\beta}}_i\right\}$.

(ii) Conditionally on $\{N_{ij} : 1 \le i \le k, 1 \le j \le m\}$, $\widehat{Y}_{ij} \sim \mathcal{N}\left(\boldsymbol{x}_j^\intercal \boldsymbol{\beta}_i, \sigma_i^2(\boldsymbol{x}_j)/N_{ij}\right)$ *for all* $i = 1,\ldots,k$, $j = 1,\ldots,m$, *and* $\widehat{Y}_{ij}$ *is independent of* $\widehat{Y}_{i'j'}$ *if* $(i,j) \ne (i',j')$.

(iii) N_{ij} *is independent of the configuration of the means, for all* $i = 1,\ldots,k$, $j = 1,\ldots,m$.

Proof. Suppose that $\boldsymbol{\beta} = (\boldsymbol{\beta}_i : 1 \le i \le k)$ follows the GSC. Then, $i^*(\boldsymbol{x}) \equiv 1$ and by Property (i), conditionally on $\boldsymbol{X} = \boldsymbol{x}$,

$$\begin{aligned}\mathrm{PCS}(\boldsymbol{x};\boldsymbol{\beta}) &= \mathbb{P}\left(\boldsymbol{x}^\intercal \widehat{\boldsymbol{\beta}}_1 - \boldsymbol{x}^\intercal \widehat{\boldsymbol{\beta}}_i > 0,\ \forall i = 2,\ldots,k\right)\\ &= \mathbb{E}\left[\mathbb{P}\left(\boldsymbol{x}^\intercal \widehat{\boldsymbol{\beta}}_1 - \boldsymbol{x}^\intercal \widehat{\boldsymbol{\beta}}_i > 0,\ \forall i = 2,\ldots,k \,\middle|\, N_{ij}, 1 \le i \le k, 1 \le j \le m\right\}\right], \qquad \text{(EC.34)}\end{aligned}$$

where the expectation is taken with respect to the N_{ij}'s and we write $\mathrm{PCS}(\boldsymbol{x};\boldsymbol{\beta})$ to stress its dependence on $\boldsymbol{\beta}$ since we will consider a different configuration of the means later.

By Property (ii), conditionally on $\boldsymbol{X} = \boldsymbol{x}$ and $\{N_{ij} : 1 \le i \le k, 1 \le j \le m\}$, $\boldsymbol{x}^\intercal \widehat{\boldsymbol{\beta}}_i$ is independent of $\boldsymbol{x}^\intercal \widehat{\boldsymbol{\beta}}_{i'}$ for $i \ne i'$; moreover,

$$\boldsymbol{x}^\intercal \widehat{\boldsymbol{\beta}}_i \big| \{N_{i,j} : 1 \le i \le k, 1 \le j \le m\} \sim \mathcal{N}\left(\boldsymbol{x}^\intercal \boldsymbol{\beta}_i, \tilde{\sigma}^2(\boldsymbol{x}, \Sigma_i)\right),$$

where $\tilde{\sigma}^2(\boldsymbol{x},\Sigma_i) := \boldsymbol{x}^\intercal (\mathcal{X}^\intercal \mathcal{X})^{-1} \mathcal{X}^\intercal \Sigma_i \mathcal{X} (\mathcal{X}^\intercal \mathcal{X})^{-1} \boldsymbol{x}$ and $\Sigma_i := \mathrm{Diag}(\sigma_i^2(\boldsymbol{x}_1)/N_{i1}, \ldots, \sigma_i^2(\boldsymbol{x}_m)/N_{im})$. In particular, $\tilde{\sigma}^2(\boldsymbol{x},\Sigma_i)$ does not depend on $\boldsymbol{\beta}$ by Property (iii). Hence, if we let $\phi(\cdot;\mu,\sigma^2)$ denote the pdf of $\mathcal{N}(\mu,\sigma^2)$, it follows from (EC.34) that

$$\mathrm{PCS}(\boldsymbol{x};\boldsymbol{\beta}) = \mathbb{E}\left[\int_{-\infty}^{+\infty} \prod_{i=2}^{k} \Phi\left(\frac{t - \boldsymbol{x}^\intercal \boldsymbol{\beta}_i}{\tilde{\sigma}(\boldsymbol{x},\Sigma_i)}\right) \phi(t; \boldsymbol{x}^\intercal \boldsymbol{\beta}_1, \tilde{\sigma}^2(\boldsymbol{x},\Sigma_1))\, \mathrm{d}t\right]. \qquad \text{(EC.35)}$$

We now consider a different configuration of the means, $\boldsymbol{\beta}^\dagger = (\boldsymbol{\beta}_i^\dagger : 1 \le i \le k)$. We will show below that $\mathrm{PCS}(\boldsymbol{x};\boldsymbol{\beta}^\dagger) \ge \mathrm{PCS}(\boldsymbol{x};\boldsymbol{\beta})$ for all $\boldsymbol{x} \in \Theta$. For each $i = 1,\ldots,k$, we define sets $\Theta_i^{(1)}$ and $\Theta_i^{(2)}$ as follows,

$$\begin{aligned}\Theta_i^{(1)} &= \{\boldsymbol{x} \in \Theta : \boldsymbol{x}^\intercal \boldsymbol{\beta}_i^\dagger - \boldsymbol{x}^\intercal \boldsymbol{\beta}_j^\dagger \ge \delta \text{ for all } j \ne i\},\\ \Theta_i^{(2)} &= \{\boldsymbol{x} \in \Theta : \boldsymbol{x}^\intercal \boldsymbol{\beta}_i^\dagger - \boldsymbol{x}^\intercal \boldsymbol{\beta}_j^\dagger \ge 0 \text{ for all } j \ne i, \text{ and } \boldsymbol{x}^\intercal \boldsymbol{\beta}_i^\dagger - \boldsymbol{x}^\intercal \boldsymbol{\beta}_j^\dagger < \delta \text{ for some } j \ne i\}.\end{aligned}$$

Clearly, $\{\Theta_i^{(1)}, \Theta_i^{(2)} : i = 1,\ldots,k\}$ are mutually exclusive and $\Theta = \bigcup_{i=1}^k \left(\Theta_i^{(1)} \bigcup \Theta_i^{(2)}\right)$. We next conduct our analysis for each $\Theta_i^{(1)}$ and $\Theta_i^{(2)}$, respectively.

- **Case 1:** $\Theta_1^{(1)} \ne \emptyset$**.** For any $\boldsymbol{x} \in \Theta_1^{(1)}$, $\boldsymbol{x}^\intercal \boldsymbol{\beta}_1^\dagger - \boldsymbol{x}^\intercal \boldsymbol{\beta}_i^\dagger \ge \delta$ for each $i = 2,\ldots,k$. By the same analysis that leads to (EC.35), we can show that for any $\boldsymbol{x} \in \Theta_1^{(1)}$,

$$\begin{aligned}&\mathrm{PCS}(\boldsymbol{x};\boldsymbol{\beta}^\dagger)\\ &= \mathbb{E}\left[\int_{-\infty}^{+\infty} \prod_{i=2}^{k} \Phi\left(\frac{t - \boldsymbol{x}^\intercal \boldsymbol{\beta}_i^\dagger}{\tilde{\sigma}(\boldsymbol{x},\Sigma_i)}\right) \phi(t; \boldsymbol{x}^\intercal \boldsymbol{\beta}_1^\dagger, \tilde{\sigma}^2(\boldsymbol{x},\Sigma_1))\, \mathrm{d}t\right]\end{aligned}$$

$$= \mathbb{E}\left[\int_{-\infty}^{+\infty}\prod_{i=2}^{k}\Phi\left(\frac{t+(\boldsymbol{x}^\intercal\boldsymbol{\beta}_1^\dagger-\boldsymbol{x}^\intercal\boldsymbol{\beta}_1)-\boldsymbol{x}^\intercal\boldsymbol{\beta}_i^\dagger}{\tilde{\sigma}(\boldsymbol{x},\Sigma_i)}\right)\phi(t+(\boldsymbol{x}^\intercal\boldsymbol{\beta}_1^\dagger-\boldsymbol{x}^\intercal\boldsymbol{\beta}_1);\boldsymbol{x}^\intercal\boldsymbol{\beta}_1^\dagger,\tilde{\sigma}^2(\boldsymbol{x},\Sigma_1))\,\mathrm{d}t\right]$$
$$= \mathbb{E}\left[\int_{-\infty}^{+\infty}\prod_{i=2}^{k}\Phi\left(\frac{t-(\boldsymbol{x}^\intercal\boldsymbol{\beta}_1-\boldsymbol{x}^\intercal\boldsymbol{\beta}_1^\dagger+\boldsymbol{x}^\intercal\boldsymbol{\beta}_i^\dagger)}{\tilde{\sigma}(\boldsymbol{x},\Sigma_i)}\right)\phi(t;\boldsymbol{x}^\intercal\boldsymbol{\beta}_1,\tilde{\sigma}^2(\boldsymbol{x},\Sigma_1))\,\mathrm{d}t\right].$$

Due to (7) and the fact that $\boldsymbol{x}^\intercal\boldsymbol{\beta}_1^\dagger-\boldsymbol{x}^\intercal\boldsymbol{\beta}_i^\dagger\geq\delta$ for each $i=2,\dots,k$, $(\boldsymbol{x}^\intercal\boldsymbol{\beta}_1-\boldsymbol{x}^\intercal\boldsymbol{\beta}_1^\dagger+\boldsymbol{x}^\intercal\boldsymbol{\beta}_i^\dagger)\leq\boldsymbol{x}^\intercal\boldsymbol{\beta}_i$, for $i=2,\dots,k$. Since $\Phi(\cdot)$ is an increasing function, it is straightforward to see that $\mathrm{PCS}(\boldsymbol{x};\boldsymbol{\beta}^\dagger)\geq\mathrm{PCS}(\boldsymbol{x};\boldsymbol{\beta})$ for any $\boldsymbol{x}\in\Theta_1^{(1)}$.

- **Case 2:** $\Theta_1^{(2)}\neq\emptyset$**.** Fix an arbitrary $\boldsymbol{x}\in\Theta_1^{(2)}$, let $\Omega(\boldsymbol{x}):=\{i=2,\dots,k:\boldsymbol{x}^\intercal\boldsymbol{\beta}_1^\dagger-\boldsymbol{x}^\intercal\boldsymbol{\beta}_i^\dagger\geq\delta\}$. Then, $\Omega(\boldsymbol{x})\subset\{2,\dots,k\}$ by the definition of $\Theta_1^{(2)}$. If $\Omega(\boldsymbol{x})=\emptyset$, then each alternative i, $i=2,\dots,k$, is in the IZ, and thus $\mathrm{PCS}(x,\boldsymbol{\beta}^\dagger)=1$. Otherwise, $(\boldsymbol{x}^\intercal\boldsymbol{\beta}_1-\boldsymbol{x}^\intercal\boldsymbol{\beta}_1^\dagger+\boldsymbol{x}^\intercal\boldsymbol{\beta}_i^\dagger)\leq\boldsymbol{x}^\intercal\boldsymbol{\beta}_i$ for each $i\in\Omega(\boldsymbol{x})$. Hence,

$$\begin{aligned}\mathrm{PCS}(\boldsymbol{x};\boldsymbol{\beta}^\dagger)&\geq\mathbb{P}\left(\boldsymbol{x}^\intercal\widehat{\boldsymbol{\beta}}_1^\dagger-\boldsymbol{x}^\intercal\widehat{\boldsymbol{\beta}}_i^\dagger>0,\ \forall i\in\Omega(\boldsymbol{x})\right)\\&=\mathbb{E}\left[\int_{-\infty}^{+\infty}\prod_{i\in\Omega(\boldsymbol{x})}\Phi\left(\frac{t-\boldsymbol{x}^\intercal\boldsymbol{\beta}_i^\dagger}{\tilde{\sigma}(\boldsymbol{x},\Sigma_i)}\right)\phi(t;\boldsymbol{x}^\intercal\boldsymbol{\beta}_1^\dagger,\tilde{\sigma}^2(\boldsymbol{x},\Sigma_1))\,\mathrm{d}t\right]\\&=\mathbb{E}\left[\int_{-\infty}^{+\infty}\prod_{i\in\Omega(\boldsymbol{x})}\Phi\left(\frac{t-(\boldsymbol{x}^\intercal\boldsymbol{\beta}_1-\boldsymbol{x}^\intercal\boldsymbol{\beta}_1^\dagger+\boldsymbol{x}^\intercal\boldsymbol{\beta}_i^\dagger)}{\tilde{\sigma}(\boldsymbol{x},\Sigma_i)}\right)\phi(t;\boldsymbol{x}^\intercal\boldsymbol{\beta}_1,\tilde{\sigma}^2(\boldsymbol{x},\Sigma_1))\,\mathrm{d}t\right]\\&\geq\mathbb{E}\left[\int_{-\infty}^{+\infty}\prod_{i\in\Omega(\boldsymbol{x})}\Phi\left(\frac{t-\boldsymbol{x}^\intercal\boldsymbol{\beta}_i}{\tilde{\sigma}(\boldsymbol{x},\Sigma_i)}\right)\phi(t;\boldsymbol{x}^\intercal\boldsymbol{\beta}_1,\tilde{\sigma}^2(\boldsymbol{x},\Sigma_1))\,\mathrm{d}t\right]\\&\geq\mathrm{PCS}(\boldsymbol{x};\boldsymbol{\beta}),\end{aligned}$$

where the last inequality holds because $0\leq\Phi(\cdot)\leq1$ and $|\Omega(\boldsymbol{x})|<k-1$.

- **Other Cases.** For each $i=2,\dots,k$, if $\Theta_i^{(1)}\neq\emptyset$, then we can simply swap the indexes of alternative 1 and alternative i, and follow the same analysis as in Case 1. Likewise, for each $i=2,\dots,k$, if $\Theta_i^{(2)}\neq\emptyset$, we can follow the analysis in Case 2.

Therefore, we conclude that $\mathrm{PCS}(\boldsymbol{x};\boldsymbol{\beta}^\dagger)\geq\mathrm{PCS}(\boldsymbol{x};\boldsymbol{\beta})$ for any $\boldsymbol{x}\in\Theta$. So $\mathbb{E}[\mathrm{PCS}(\boldsymbol{X};\boldsymbol{\beta}^\dagger)]\geq\mathbb{E}[\mathrm{PCS}(\boldsymbol{X};\boldsymbol{\beta})]$. Moreover, the foregoing analysis also shows that, the equality may hold only if random vector $\boldsymbol{X}$ is degenerate to a constant vector. Thus, the GSC is the LFC. □

REMARK EC.6. Obviously, both Procedure TS and TS$^+$ possess those properties specified in Theorem EC.3, so Theorem 5 holds immediately as a corollary of Theorem EC.3.

EC.9. Proof of Theorem 6

Proof of Theorem 6. It suffices to show that the extreme design yields the minimal value of the solution h to (4) among all symmetric designs. We first notice that by (4), the design matrix $\mathcal{X}$ takes effect on the total sample size of Procedure TS only through the form $\mathcal{X}^\intercal\mathcal{X}$. In the sequel, two design matrices $\mathcal{X}$ and $\tilde{\mathcal{X}}$ are said to be *equivalent* if $\mathcal{X}^\intercal\mathcal{X} = \tilde{\mathcal{X}}^\intercal\tilde{\mathcal{X}}$. For instance, swapping any two rows of $\mathcal{X}$ leads to an equivalent design matrix since it does not change $\mathcal{X}^\intercal\mathcal{X}$.

Let $\mathcal{X}_*$ denote the design matrix corresponding to the extreme design $\mathcal{S}^1 = \cdots = \mathcal{S}^b = \mathcal{S}^0$, and $\mathcal{X}_\dagger$ denote a nonequivalent design matrix corresponding to a symmetric design. The key of the proof is to show that

$$\boldsymbol{x}^\intercal(\mathcal{X}_*^\intercal\mathcal{X}_*)^{-1}\boldsymbol{x} \le \boldsymbol{x}^\intercal(\mathcal{X}_\dagger^\intercal\mathcal{X}_\dagger)^{-1}\boldsymbol{x}, \quad \boldsymbol{x}\in\Theta, \tag{EC.36}$$

where the equality holds if and only if $\boldsymbol{x} = \left(1, \frac{l_2+u_2}{2}, \ldots, \frac{l_d+u_d}{2}\right)^\intercal$, which is the center of Θ.

To see this, let h_* and $h_\dagger$ denote the solution h of (4) for $\mathcal{X}_*$ and $\mathcal{X}_\dagger$, respectively. Notice that the double integral on the left-hand side of (4) is strictly increasing in h whereas strictly decreasing in $\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}$. Hence, if (EC.36) holds, then $h_* \le h_\dagger$, where the equality holds if and only if the random vector $\boldsymbol{X} \equiv \left(1, \frac{l_2+u_2}{2}, \ldots, \frac{l_d+u_d}{2}\right)^\intercal$.

Now we prove (EC.36). For ease of presentation, we first consider a design matrix that corresponds to a general symmetric design. Since the first element of the covariates is always 1, the $b(2^{d-1}) \times d$ design matrix $\mathcal{X}$ is

$$\mathcal{X} = \begin{pmatrix} (\boldsymbol{a}_1^1)^\intercal \\ \vdots \\ (\boldsymbol{a}_{2^{d-1}}^1)^\intercal \\ \vdots \\ (\boldsymbol{a}_1^b)^\intercal \\ \vdots \\ (\boldsymbol{a}_{2^{d-1}}^b)^\intercal \end{pmatrix} = \begin{pmatrix} 1 & a_{1,2}^1 & \cdots & a_{1,d}^1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & a_{2^{d-1},2}^1 & \cdots & a_{2^{d-1},d}^1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & a_{1,2}^b & \cdots & a_{1,d}^b \\ \vdots & \vdots & \vdots & \vdots \\ 1 & a_{2^{d-1},2}^b & \cdots & a_{2^{d-1},d}^b \end{pmatrix} \triangleq \left(\boldsymbol{1}\ \boldsymbol{v}_2 \cdots \boldsymbol{v}_d\right),$$

where $\boldsymbol{1}$ denotes the $b(2^{d-1}) \times 1$ vector of ones. We further set $\mathcal{Z} := (\boldsymbol{v}_2, \ldots, \boldsymbol{v}_d)$. Then $\mathcal{X} = (\boldsymbol{1}, \mathcal{Z})$, and

$$\mathcal{X}^\intercal\mathcal{X} = \begin{pmatrix} \boldsymbol{1}^\intercal\boldsymbol{1} & \boldsymbol{1}^\intercal\mathcal{Z} \\ \mathcal{Z}^\intercal\boldsymbol{1} & \mathcal{Z}^\intercal\mathcal{Z} \end{pmatrix}. \tag{EC.37}$$

Notice that $m = b(2^{d-1}) = \boldsymbol{1}^\intercal\boldsymbol{1}$. Then, for any $\boldsymbol{x} = (1, \boldsymbol{z}^\intercal)^\intercal$, where $\boldsymbol{z} \in \mathbb{R}^{d-1}$, standard matrix calculation (Horn and Johnson 2013, §0.8.5) yields that

$$\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x} = \left(\boldsymbol{z} - \mathcal{Z}^\intercal\boldsymbol{1}m^{-1}\right)^\intercal \mathcal{A}(\mathcal{Z})^{-1}\left(\boldsymbol{z} - \mathcal{Z}^\intercal\boldsymbol{1}m^{-1}\right) + m^{-1},$$

where $\mathcal{A}(\mathcal{Z}) := \mathcal{Z}^\intercal\mathcal{Z} - \mathcal{Z}^\intercal\boldsymbol{1}m^{-1}\boldsymbol{1}^\intercal\mathcal{Z}$ is the Schur complement of the block $\boldsymbol{1}^\intercal\boldsymbol{1}$ of $\mathcal{X}^\intercal\mathcal{X}$ in (EC.37) and it is nonsingular because $\mathcal{X}^\intercal\mathcal{X}$ is nonsingular. The symmetry of design points implies that $\boldsymbol{v}_w^\intercal\boldsymbol{1}m^{-1} = \frac{l_w+u_w}{2}$, for $w = 2, \ldots, d$. So, by letting $\boldsymbol{s} := \left(\frac{l_2+u_2}{2}, \ldots, \frac{l_d+u_d}{2}\right)^\intercal$, we have $\mathcal{Z}^\intercal\boldsymbol{1}m^{-1} = \boldsymbol{s}$ and

$$\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x} = (\boldsymbol{z} - \boldsymbol{s})^\intercal \mathcal{A}(\mathcal{Z})^{-1}(\boldsymbol{z} - \boldsymbol{s}) + m^{-1}. \tag{EC.38}$$

Hence, if $\boldsymbol{x} = \left(1, \frac{l_2+u_2}{2}, \ldots, \frac{l_d+u_d}{2}\right)^\intercal$, then $\boldsymbol{z} = \boldsymbol{s}$ and thus $\boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x} = m^{-1}$. Since both $\mathcal{X}_*$ and $\mathcal{X}_\dagger$ are symmetric designs, $\boldsymbol{x}^\intercal(\mathcal{X}_*^\intercal\mathcal{X}_*)^{-1}\boldsymbol{x} = \boldsymbol{x}^\intercal(\mathcal{X}_\dagger^\intercal\mathcal{X}_\dagger)^{-1}\boldsymbol{x}$ if $\boldsymbol{x} = \left(1, \frac{l_2+u_2}{2}, \ldots, \frac{l_d+u_d}{2}\right)^\intercal$.

It remains to prove the strict inequality in (EC.36) for $\boldsymbol{z} \neq \boldsymbol{s}$. Let $\mathcal{X}_* = (\mathbf{1}, \mathcal{Z}_*)$ and $\mathcal{X}_\dagger = (\mathbf{1}, \mathcal{Z}_\dagger)$. Due to (EC.38), it suffices to show that $\mathcal{A}(\mathcal{Z}_\dagger)^{-1} - \mathcal{A}(\mathcal{Z}_*)^{-1}$ is positive definite. This is equivalent to showing that $\mathcal{A}(\mathcal{Z}_*) - \mathcal{A}(\mathcal{Z}_\dagger)$ is positive definite (Horn and Johnson 2013, Corollary 7.7.4), i.e., for any nonzero $\boldsymbol{z} \in \mathbb{R}^{d-1}$,

$$\boldsymbol{z}^\intercal \mathcal{A}(\mathcal{Z}_*) \boldsymbol{z} > \boldsymbol{z}^\intercal \mathcal{A}(\mathcal{Z}_\dagger) \boldsymbol{z}. \tag{EC.39}$$

Let $\mathcal{I}$ denote the $b(2^{d-1}) \times b(2^{d-1})$ identity matrix. Then,

$$\mathcal{A}(\mathcal{Z}) = \mathcal{Z}^\intercal \left(\mathcal{I} - \mathbf{1}m^{-1}\mathbf{1}^\intercal\right) \mathcal{Z} = \mathcal{Z}^\intercal \left(\mathcal{I} - \mathbf{1}m^{-1}\mathbf{1}^\intercal\right)\left(\mathcal{I} - \mathbf{1}m^{-1}\mathbf{1}^\intercal\right) \mathcal{Z} = \left(\mathcal{Z} - \mathbf{1}\boldsymbol{s}^\intercal\right)^\intercal \left(\mathcal{Z} - \mathbf{1}\boldsymbol{s}^\intercal\right),$$

since $\mathcal{Z}^\intercal \mathbf{1} m^{-1} = \boldsymbol{s}$. Denote $\boldsymbol{z} := (z_2, \dots, z_d)^\intercal$ and $\boldsymbol{s} := (s_2, \dots, s_d)^\intercal$. Then

$$\boldsymbol{z}^\intercal \mathcal{A}(\mathcal{Z}) \boldsymbol{z} = \left[\left(\mathcal{Z} - \mathbf{1}\boldsymbol{s}^\intercal\right) \boldsymbol{z}\right]^\intercal \left[\left(\mathcal{Z} - \mathbf{1}\boldsymbol{s}^\intercal\right) \boldsymbol{z}\right] = \sum_{w=2}^{d} \sum_{q=2}^{d} z_w z_q (\boldsymbol{v}_w - \mathbf{1}s_w)^\intercal (\boldsymbol{v}_q - \mathbf{1}s_q).$$

Thanks to the symmetry of the design points, it is easy to verify that

$$(\boldsymbol{v}_w - \mathbf{1}s_w)^\intercal (\boldsymbol{v}_q - \mathbf{1}s_q) = \begin{cases} 0, & \text{if } w \neq q, \\ 2^{d-1} \sum_{j=1}^{b} (\rho_w^j)^2, & \text{if } w = q, \end{cases}$$

where $\rho_w^j := \left|a_{1,w}^j - s_w\right| = \cdots = \left|a_{2^{d-1},w}^j - s_w\right|$ is the common distance of $\boldsymbol{a}_1^j, \dots, \boldsymbol{a}_{2^{d-1}}^j$ to the center of Θ along coordinate x_w, for $w = 2, \dots, d$ and $j = 1, \dots, b$. Hence,

$$\boldsymbol{z}^\intercal \mathcal{A}(\mathcal{Z}) \boldsymbol{z} = 2^{d-1} \sum_{w=2}^{d} z_w^2 \sum_{j=1}^{b} (\rho_w^j)^2.$$

Since $\rho_w^j \in \left(0, \frac{u_w - l_w}{2}\right]$, obviously, $\{\rho_w^j = \frac{u_w - l_w}{2} | w = 2, \dots, d, j = 1, \dots, b\}$ maximizes $\boldsymbol{z}^\intercal \mathcal{A}(\mathcal{Z}) \boldsymbol{z}$. Because $\boldsymbol{z}$ is nonzero, i.e., at least one z_w is not zero, the solution is unique in terms of ρ_w^j. Notice that this solution exactly means that $\mathcal{S}^1 = \cdots = \mathcal{S}^b = \mathcal{S}^0$, i.e., the extreme design. Hence, (EC.39) is proved and the proof of (EC.36) is completed. □

EC.10. Proof of Theorem 7

Before the proof, we first introduce formally the D-optimality and the G-optimality of experimental designs in the linear regression setting. Consider the linear regression model

$$Y(\boldsymbol{x}) = \boldsymbol{x}^\intercal \boldsymbol{\beta} + \epsilon,$$

where $\boldsymbol{\beta}, \boldsymbol{x} \in \mathbb{R}^d$ and ϵ is random error with mean 0 and variance σ^2. Assuming the design region is Θ, we choose m design points with $\boldsymbol{x}_i \in \Theta$, $i = 1, \dots, m$. Let $\boldsymbol{Y} = (Y(\boldsymbol{x}_1), \dots, Y(\boldsymbol{x}_m))^\intercal$, $\mathcal{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_m)^\intercal$, and $\Upsilon = \{\mathcal{X} : \operatorname{rank}(\mathcal{X}^\intercal \mathcal{X}) = d, \boldsymbol{x}_i \in \Theta, i = 1, \dots, m\}$. It is known that if $\mathcal{X} \in \Upsilon$, the OLS estimator of $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}} = (\mathcal{X}^\intercal \mathcal{X})^{-1} \mathcal{X}^\intercal \boldsymbol{Y}$. Moreover, $\operatorname{Var}(\widehat{\boldsymbol{\beta}}) = \sigma^2 (\mathcal{X}^\intercal \mathcal{X})^{-1}$ and $\operatorname{Var}(\boldsymbol{x}^\intercal \widehat{\boldsymbol{\beta}}) = \sigma^2 \boldsymbol{x}^\intercal (\mathcal{X}^\intercal \mathcal{X})^{-1} \boldsymbol{x}$. The D-optimality and the G-optimality are related to the two variances, respectively.

A design $\mathcal{X}_*$ is said to be *D-optimal* if

$$\mathcal{X}_* = \arg\max_{\mathcal{X}\in\Upsilon} \det\left(\mathcal{X}^\intercal\mathcal{X}\right).$$

The D-optimal design aims to minimize the volume of confidence ellipsoid for $\boldsymbol{\beta}$ given a fixed confidence level under the assumption that the errors are normally distributed.

A design $\mathcal{X}_*$ is said to be *G-optimal* if

$$\mathcal{X}_* = \arg\min_{\mathcal{X}\in\Upsilon} \left\{\max_{\boldsymbol{x}\in\Theta} \boldsymbol{x}^\intercal(\mathcal{X}^\intercal\mathcal{X})^{-1}\boldsymbol{x}\right\}.$$

The G-optimal design aims to minimize the maximum variance of the fitted response over the design region.

Theorem 7 is an application of the general equivalence theory; see, e.g., Silvey (1980, Chapter 3) for a careful discussion on this subject. Some concepts need to be introduced before the proof.

The first concept is the *continuous design*, also called approximate design. Suppose that we relax the constraint that the number of samples at each design point must be an integer, that is, we can allocate any portion of a given total sample size m to any point in Θ. Formally speaking, the allocation can be described by a probability distribution ψ on Θ, which can be either continuous or discrete. Let $\boldsymbol{X}$ be a random vector with distribution ψ, and define $\mathcal{M}(\psi) := \mathbb{E}_\psi(\boldsymbol{X}\boldsymbol{X}^\intercal)$. For example, if $\mathcal{X}$ is an *exact* design which contains *distinct* points $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$, having $m_1,\ldots,m_n$ samples, respectively, where $m_1+\cdots+m_n = m$, then the distribution ψ for sample allocation is defined by $\mathbb{P}(\boldsymbol{X}=\boldsymbol{x}_i) = m_i/m$, $i=1,\ldots,n$, and thus $\mathcal{M}(\psi) = m^{-1}\mathcal{X}^\intercal\mathcal{X}$. However, a continuous design may not be an exact design due to the integrality constraint.

By allowing continuous designs, the D-optimal design is extended to be a distribution

$$\psi_* = \arg\max_{\psi\in\Psi} \det\left(\mathcal{M}(\psi)\right),$$

where Ψ denotes the set of all ψ such that $\mathcal{M}(\psi)$ is nonsingular. Notice that Ψ is also the set of all ψ such that $\mathcal{M}(\psi)$ is positive definite. Likewise, the G-optimal design can be extended as follows

$$\psi_* = \arg\min_{\psi\in\Psi} \left\{\max_{\boldsymbol{x}\in\Theta} \boldsymbol{x}^\intercal(\mathcal{M}(\psi))^{-1}\boldsymbol{x}\right\}.$$

More generally, consider a function f of positive definite matrices. A continuous design ψ_* is said to be f-optimal if

$$\psi_* = \arg\max_{\psi\in\Psi} f\left(\mathcal{M}(\psi)\right).$$

For instance, the D-optimal design and the G-optimal design can be obtained by setting $f(\mathcal{M}) = \log\det\left(\mathcal{M}\right)$ and $f(\mathcal{M}) = -\max_{\boldsymbol{x}\in\Theta} \boldsymbol{x}^\intercal(\mathcal{M})^{-1}\boldsymbol{x}$, respectively. It is easy to verify that both functions are concave in $\mathcal{M}$. The use of concavity will become clear in Lemma EC.5 below.

At last, we introduce two kinds of derivatives. The *Gâteaux derivative* of f at $\mathcal{M}_1$ in the direction of $\mathcal{M}_2$ is defined as

$$G_f(\mathcal{M}_1, \mathcal{M}_2) := \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \left[f(\mathcal{M}_1 + \varepsilon \mathcal{M}_2) - f(\mathcal{M}_1) \right].$$

We say f is *differentiable* at $\mathcal{M}_1$ if $G_f(\mathcal{M}_1, \mathcal{M}_2)$ is well defined. The *Fréchet derivative* of f at $\mathcal{M}_1$ in the direction of $\mathcal{M}_2$ is defined as

$$F_f(\mathcal{M}_1, \mathcal{M}_2) := \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \left[f\left\{(1-\varepsilon)\mathcal{M}_1 + \varepsilon \mathcal{M}_2\right\} - f(\mathcal{M}_1) \right] = G_f(\mathcal{M}_1, \mathcal{M}_2 - \mathcal{M}_1).$$

We state Theorem 3.7 of Silvey (1980) as Lemma EC.5 and will apply it to prove Theorem 7.

LEMMA EC.5. *If f is a concave function of positive definite matrices and is differentiable at $\mathcal{M}(\psi_*)$, then ψ_* if f-optimal if and only if $F_f(\mathcal{M}(\psi_*), \boldsymbol{x}\boldsymbol{x}^\intercal) \leq 0$ for all $\boldsymbol{x} \in \Theta$.*

Now we are ready to prove Theorem 7.

Proof of Theorem 7. Consider the continuous design ψ_0 that assigns probability $1/(2^{d-1})$ at each corner point of Θ. Since $m = b(2^{d-1})$, ψ_0 is indeed the exact extreme design defined in §7.1. Hence, it suffices to prove that ψ_0 is both D-optimal and G-optimal in the continuous case.

We first prove the D-optimality of ψ_0. Let $\mathcal{X}_0$ denote the design matrix corresponding to ψ_0. Then, $\mathcal{X}_0$ is a $2^{d-1} \times d$ matrix with each row corresponding one of the 2^{d-1} distinct corners of Θ. For example, if $d = 3$,

$$\mathcal{X}_0 = \begin{pmatrix} 1 & l_2 & l_3 \\ 1 & l_2 & u_3 \\ 1 & u_2 & l_3 \\ 1 & u_2 & u_3 \end{pmatrix}.$$

It is easy to see that $\mathcal{M}(\psi_0) = \frac{1}{2^{d-1}} \mathcal{X}_0^\intercal \mathcal{X}_0$. In the sequel, we verify the conditions of Lemma EC.5 for $f = \log\det$ to prove the D-optimality.

The concavity of f is trivial; see, e.g., Theorem 7.6.6 of Horn and Johnson (2013). For the differentiability, notice that for any positive definite matrix $\mathcal{M}_1$,

$$\log\det(\mathcal{M}_1 + \varepsilon\mathcal{M}_2) - \log\det(\mathcal{M}_1) = \log\det(\mathcal{I} + \varepsilon\mathcal{M}_1^{-1}\mathcal{M}_2) = \sum_{w=1}^{d} \log(1 + \varepsilon\lambda_w),$$

where $\mathcal{I}$ is the $d \times d$ identity matrix and $\lambda_1, \ldots, \lambda_d$ are the eigenvalues of $\mathcal{M}_1^{-1}\mathcal{M}_2$ which are all real (Horn and Johnson 2013, Corollary 7.6.2). Hence,

$$G_f(\mathcal{M}_1, \mathcal{M}_2) = \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \left[\log\det(\mathcal{M}_1 + \varepsilon\mathcal{M}_2) - \log\det(\mathcal{M}_1) \right] = \sum_{w=1}^{d} \lambda_w = \operatorname{tr}(\mathcal{M}_1^{-1}\mathcal{M}_2),$$

is well defined, so f is differentiable at $\mathcal{M}_1$.

Moreover,

$$F_f(\mathcal{M}_1, \mathcal{M}_2) = G_f(\mathcal{M}_1, \mathcal{M}_2 - \mathcal{M}_1) = \operatorname{tr}(\mathcal{M}_1^{-1}\mathcal{M}_2 - \mathcal{I}) = \operatorname{tr}(\mathcal{M}_1^{-1}\mathcal{M}_2) - d.$$

Hence, for any $\boldsymbol{x}\in\Theta$,

$$F_f(\mathcal{M}(\psi_0),\boldsymbol{x}\boldsymbol{x}^\intercal)=\operatorname{tr}(\mathcal{M}(\psi_0)^{-1}\boldsymbol{x}\boldsymbol{x}^\intercal)-d=\operatorname{tr}(\boldsymbol{x}^\intercal\mathcal{M}(\psi_0)^{-1}\boldsymbol{x})-d=\boldsymbol{x}^\intercal\mathcal{M}(\psi_0)^{-1}\boldsymbol{x}-d.$$

Since $\mathcal{M}(\psi_0)^{-1}$ is positive definite, $F_f(\mathcal{M}(\psi_0),\boldsymbol{x}\boldsymbol{x}^\intercal)$ is convex in $\boldsymbol{x}$, thereby achieving its maximum only if $\boldsymbol{x}$ is one of the corners of Θ, i.e., $\boldsymbol{x}\in\mathcal{S}^0$. Therefore, to verify $F_f(\mathcal{M}(\psi_0),\boldsymbol{x}\boldsymbol{x}^\intercal)\le 0$, it suffices to show that

$$\boldsymbol{x}^\intercal\mathcal{M}(\psi_0)^{-1}\boldsymbol{x}=d,\quad \boldsymbol{x}\in\mathcal{S}^0. \tag{EC.40}$$

We denote $\mathcal{X}_0\coloneqq(\mathbf{1},\mathcal{Z}_0)\triangleq(\mathbf{1},\boldsymbol{v}_2,\ldots,\boldsymbol{v}_d)$, where $\mathbf{1}$ is the $2^{d-1}\times 1$ vector of ones. Following the standard matrix calculations similar to those in the proof of Theorem 6, we can have that, for any $\boldsymbol{x}=(1,\boldsymbol{z}^\intercal)^\intercal$, where $\boldsymbol{z}\coloneqq(z_2,\ldots,z_d)^\intercal\in\mathbb{R}^{d-1}$,

$$\boldsymbol{x}^\intercal\mathcal{M}(\psi_0)^{-1}\boldsymbol{x}=2^{d-1}\boldsymbol{x}^\intercal(\mathcal{X}_0^\intercal\mathcal{X}_0)^{-1}\boldsymbol{x}=2^{d-1}\left[(\boldsymbol{z}-\boldsymbol{s})^\intercal\left\{(\mathcal{Z}_0-\mathbf{1}\boldsymbol{s}^\intercal)^\intercal(\mathcal{Z}_0-\mathbf{1}\boldsymbol{s}^\intercal)\right\}^{-1}(\boldsymbol{z}-\boldsymbol{s})+1/\left(2^{d-1}\right)\right], \tag{EC.41}$$

where $\boldsymbol{s}=(s_2,\ldots,s_d)^\intercal\coloneqq\left(\frac{l_2+u_2}{2},\ldots,\frac{l_d+u_d}{2}\right)^\intercal$. Notice that $(\mathcal{Z}_0-\mathbf{1}\boldsymbol{s}^\intercal)^\intercal(\mathcal{Z}_0-\mathbf{1}\boldsymbol{s}^\intercal)$ is a $(d-1)\times(d-1)$ matrix whose $(w-1,q-1)$-th element is $(\boldsymbol{v}_w-\mathbf{1}s_w)^\intercal(\boldsymbol{v}_q-\mathbf{1}s_q)$, for $w,q=2,\ldots,d$, and that

$$(\boldsymbol{v}_w-\mathbf{1}s_w)^\intercal(\boldsymbol{v}_q-\mathbf{1}s_q)=\begin{cases}0, & \text{if } w\neq q,\\ 2^{d-1}\left(\frac{u_w-l_w}{2}\right)^2, & \text{if } w=q.\end{cases}$$

Hence,

$$(\mathcal{Z}_0-\mathbf{1}\boldsymbol{s}^\intercal)^\intercal(\mathcal{Z}_0-\mathbf{1}\boldsymbol{s}^\intercal)=2^{d-1}\operatorname{Diag}\left\{\left(\frac{u_2-l_2}{2}\right)^2,\ldots,\left(\frac{u_d-l_d}{2}\right)^2\right\},$$

and thus

$$\left\{(\mathcal{Z}_0-\mathbf{1}\boldsymbol{s}^\intercal)^\intercal(\mathcal{Z}_0-\mathbf{1}\boldsymbol{s}^\intercal)\right\}^{-1}=\frac{1}{2^{d-1}}\operatorname{Diag}\left\{\left(\frac{2}{u_2-l_2}\right)^2,\ldots,\left(\frac{2}{u_d-l_d}\right)^2\right\}.$$

Moreover, for any $\boldsymbol{x}\in\mathcal{S}^0$, $\boldsymbol{z}\in\{l_2,u_2\}\times\cdots\times\{l_d,u_d\}$. Hence, $(z_w-s_w)^2=\left(\frac{u_w-l_w}{2}\right)^2$, for $w=2,\ldots,d$. Then,

$$(\boldsymbol{z}-\boldsymbol{s})^\intercal\left\{(\mathcal{Z}_0-\mathbf{1}\boldsymbol{s}^\intercal)^\intercal(\mathcal{Z}_0-\mathbf{1}\boldsymbol{s}^\intercal)\right\}^{-1}(\boldsymbol{z}-\boldsymbol{s})=\frac{1}{2^{d-1}}\sum_{w=2}^{d}\left(\frac{2}{u_w-l_w}\right)^2(z_w-s_w)^2=\frac{d-1}{2^{d-1}}. \tag{EC.42}$$

Then, (EC.40) follows immediately from (EC.41) and (EC.42), proving the D-optimality by Lemma EC.5.

The G-optimality of ψ_0 can be proved similarly, by taking $f(\mathcal{M}(\psi))=-\max_{\boldsymbol{x}\in\Theta}\boldsymbol{x}^\intercal(\mathcal{M}(\psi))^{-1}\boldsymbol{x}$. Or, we can conclude this immediately by applying the known equivalence between the D-optimality and the G-optimality for continuous designs established in Kiefer and Wolfowitz (1960); see also Silvey (1980, §3.11). □

References

Horn RA, Johnson CR (2013) *Matrix Analysis* (Cambridge University Press), 2nd edition.

Kiefer J, Wolfowitz J (1960) The equivalence of two extremum problems. *Canad. J. Math.* 12:363–366.

Rencher AC, Schaalje GB (2008) *Linear Models in Statistics* (John Wiley & Sons, Inc.), 2nd edition.

Robbins H, Monro S (1951) A stochastic approximation method. *Ann. Math. Stat.* 22(3):400–407.

Rockafellar RT (1970) *Convex Analysis* (Princeton University Press).

Silvey SD (1980) *Optimal Design: An Introduction to the Theory for Parameter Estimation* (Chapman and Hall).

Slepian D (1962) The one-sided barrier problem for Gaussian noise. *Bell System Tech. J.* 41(2):463–501.

Zhong Y, Hong LJ (2020) Knockout-tournament procedures for large-scale ranking and selection in parallel computing environments. *Working Paper.* URL `https://www.researchgate.net/publication/339200267`.