Additional Proofs and Numerical Results

EC.1. Proof for the Two-Stage RSB Procedure

Proof of Theorem 1. We first notice that if $\mu_{k1}-\mu_{11}\le\delta$, then by Assumption 1 $\mu_{i1}-\mu_{11}\le\delta$ for all $i=1,\dots,k$, which implies that $\mathbb{P}(\mu_{i^*1}-\mu_{11}\le\delta)=1$ and the theorem trivially holds. Hence, without loss of generality we assume that there exists $l=1,\dots,k-1$ for which $\mu_{l1}-\mu_{11}\le\delta$ and $\mu_{l+1,1}-\mu_{11}>\delta$. Then, a good selection (i.e., $\{\mu_{i^*1}-\mu_{11}\le\delta\}$) occurs if any alternative i, $i=l+1,\dots,k$ is not selected. It follows that

$$\begin{aligned}
\mathbb{P}\{\mu_{i^*1}-\mu_{11}\le\delta\} &\ge \mathbb{P}\left\{\bigcap_{i=l+1}^{k}\left\{\max_{1\le j\le m}\bar{X}_{ij}(N)>\max_{1\le j\le m}\bar{X}_{1j}(N)\right\}\right\}\\
&\ge \mathbb{P}\left\{\bigcap_{i=l+1}^{k}\left\{\max_{1\le j\le m}\bar{X}_{ij}(N)>\max_{1\le j\le m}\bar{X}_{1j}(N)\right\}\bigcap\left\{\max_{1\le j\le m}\bar{X}_{1j}(N)-\bar{X}_{11}(N)<\delta_I\right\}\right\}\\
&\ge \mathbb{P}\left\{\bigcap_{i=l+1}^{k}\left\{\bar{X}_{i1}(N)>\max_{1\le j\le m}\bar{X}_{1j}(N)\right\}\bigcap\left\{\max_{1\le j\le m}\bar{X}_{1j}(N)-\bar{X}_{11}(N)<\delta_I\right\}\right\}\\
&\ge \mathbb{P}\left\{\bigcap_{i=l+1}^{k}\left\{\bar{X}_{i1}(N)>\bar{X}_{11}(N)+\delta_I\right\}\bigcap\bigcap_{j=2}^{m}\left\{\bar{X}_{1j}(N)-\bar{X}_{11}(N)<\delta_I\right\}\right\}\\
&\ge 1-\sum_{i=l+1}^{k}\mathbb{P}\left\{\bar{X}_{i1}(N)\le\bar{X}_{11}(N)+\delta_I\right\}-\sum_{j=2}^{m}\mathbb{P}\left\{\bar{X}_{1j}(N)\ge\bar{X}_{11}(N)+\delta_I\right\}, \qquad \text{(EC.1)}
\end{aligned}$$

where the last step is due to the Bonferroni inequality. For each $i=l+1,\dots,k$,

$$\begin{aligned}
\mathbb{P}\{\bar{X}_{i1}(N)\le\bar{X}_{11}(N)+\delta_I\} &= \mathbb{P}\left\{\bar{X}_{i1}(N)-\bar{X}_{11}(N)-(\mu_{i1}-\mu_{11})\le-\mu_{i1}+\mu_{11}+\delta_I\right\}\\
&\le \mathbb{P}\left\{\bar{X}_{i1}(N)-\bar{X}_{11}(N)-(\mu_{i1}-\mu_{11})\le-\delta_O\right\}\\
&= \mathbb{P}\left\{\frac{\bar{X}_{i1}(N)-\bar{X}_{11}(N)-(\mu_{i1}-\mu_{11})}{\sqrt{S^2_{11,i1}/N}}\le-\frac{\delta_O}{\sqrt{S^2_{11,i1}/N}}\right\}\\
&\le \mathbb{P}\left\{\frac{\bar{X}_{i1}(N)-\bar{X}_{11}(N)-(\mu_{i1}-\mu_{11})}{\sqrt{S^2_{11,i1}/N}}\le-h\right\}, \qquad \text{(EC.2)}
\end{aligned}$$

where the first inequality holds because $\delta_I+\delta_O=\delta$ and $\mu_{i1}-\mu_{11}>\delta$ for each $i=l+1,\dots,k$ under Assumption 1, and the second inequality holds because $N\ge h^2S^2_{11,i1}/\delta_O^2$.

For notational simplicity, we suppress its dependence on i and set $Y_r=X_{i1,r}-X_{11,r}-(\mu_{i1}-\mu_{11})$ for $r=1,\dots,N$, $\sigma_Y^2=\mathrm{Var}[Y_r]$, and $S_Y^2=S^2_{11,i1}$. Applying (EC.2) and following a similar derivation in Stein (1945),

$$\mathbb{P}\{\bar{X}_{i1}(N)\le\bar{X}_{11}(N)+\delta_I\}\le\mathbb{P}\left\{\frac{\sum_{r=1}^{N}Y_r}{\sqrt{NS_Y^2}}\le-h\right\}=\mathbb{P}\{Z\le-h\}=\beta, \qquad \text{(EC.3)}$$

where Z has the distribution of Student's t with n_0-1 degrees of freedom, and the second inequality holds due to the definition of h. Likewise, we can show that

$$\mathbb{P}\left\{\bar{X}_{1j}(N)\geq\bar{X}_{11}(N)+\delta_I\right\}\leq\beta, \tag{EC.4}$$

for each $j=2,\ldots,m$. The proof is completed by combining (EC.1), (EC.3), and (EC.4) to obtain

$$\begin{aligned}\mathbb{P}\{\mu_{i^*1}-\mu_{11}>\delta\} &\leq \sum_{i=2}^{k}\mathbb{P}\left\{\bar{X}_{i1}(N)\leq\bar{X}_{11}(N)+\delta_I\right\}+\sum_{j=2}^{m}\mathbb{P}\left\{\bar{X}_{1j}(N)\geq\bar{X}_{11}(N)+\delta_I\right\}\\ &\leq (k+m-2)\beta=\alpha. \quad \square\end{aligned}$$

EC.2. Proofs for the Sequential RSB Procedure

We prove Proposition 1 and Theorem 2 in this section. To facilitate their proofs, we first present a result that characterizes the first-exit probability that a random walk exits from the region $(-g_c(t),g_c(t))$.

LEMMA EC.1. *Let $\{Y_n:i=n=1,2,\ldots\}$ be a sequence of independent and identically distributed (i.i.d.) random variables with mean 0 and variance $\sigma^2<\infty$. Let $\bar{Y}(n)$ and $S^2(n)$ denote the sample mean and the sample variance of $(Y_i:i=1,\ldots,n)$, respectively. Let $g_c(t)=\sqrt{[c+\log(t+1)](t+1)}$ with $c=-2\log(2\beta)$ for some $\beta\in(0,1)$. Assume that the moment generating function of Y_1 is finite in a neighborhood of zero.*

(i) Define $t(n)=n/\sigma^2$ and $N=\min\{n\geq n_0:t(n)|\bar{Y}(n)|\geq g_c(t(n))\}$. If $n_0\to\infty$ as $\beta\to 0$, then,

$$\limsup_{\beta\to 0}\frac{1}{\beta}\mathbb{P}\left\{t(N)\bar{Y}(N)\leq -g_c(t(N)),N<\infty\right\}\leq 1.$$

(ii) Define $\tau(n)=n/S^2(n)$ and $N'=\min\{n\geq n_0:\tau(n)|\bar{Y}(n)|\geq g_c(\tau(n))\}$. If $n_0\to\infty$ as $\beta\to 0$, then,

$$\limsup_{\beta\to 0}\frac{1}{\beta}\mathbb{P}\left\{\tau(N')\bar{Y}(N')\leq -g_c(\tau(N')),N'<\infty\right\}\leq 1.$$

Proof of Lemma EC.1. We provide a proof sketch here. The complete proof can be found in the proof for Theorem 2 in Fan et al. (2016). Let $(B(t):t\geq 0)$ be a standard Brownian motion. By virtue of the functional central limit theorem (Whitt 2002, p.102), it can be shown that

$$\mathbb{P}\left\{t(N)\bar{Y}(N)\leq -g_c(t(N)),N<\infty\right\}\leq\mathbb{P}\left\{B(T_c)\leq -g_c(T_c),T_c<\infty\right\},$$

where $T_c=\inf\{t\geq 0:|B(t)|\geq g_c(t)\}$. Moreover, Example 6 in Jennen and Lerche (1981) shows that

$$\mathbb{P}\left\{B(T_c)\leq -g_c(T_c),T_c<\infty\right\}=\frac{1}{2}e^{-c/2}=\beta,$$

and thus (i) follows immediately. On the other hand, (ii) can be shown by noting

$$\limsup_{\beta\to 0}\frac{1}{\beta}\mathbb{P}\left\{\tau(N')\bar{Y}(N')\leq -g_c(\tau(N')),N'<\infty\right\}\leq\limsup_{\beta\to 0}\frac{1}{\beta}\mathbb{P}\left\{t(N)\bar{Y}(N)\leq -g_c(t(N)),N<\infty\right\}. \quad \square$$

EC.2.1. Proof of Proposition 1

Proof of Proposition 1. If $(i,1)\in\mathcal{S}_i(n)$, then

$$U_{ii'}(n) = \max_{(i,j)\in\mathcal{S}_i(n)} \bar{X}_{ij}(n) - \max_{(i',j)\in\mathcal{S}_{i'}(n)} \bar{X}_{i'j}(n) + C_{i'}(n) + D_{ii'}(n)$$
$$\geq \bar{X}_{i1}(n) - \max_{(i',j)\in\mathcal{S}_{i'}(n)} \bar{X}_{i'j}(n) + C_{i'}(n) + \frac{g_c(t_{i1,i'1}(n))}{t_{i1,i'1}(n)}. \tag{EC.5}$$

Let $(i', j^*_{i'})$ denote the system having the largest mean performance among $\mathcal{S}_{i'}(n)$. If $(i',1)\in\mathcal{S}_{i'}(n)$, then system $(i',1)$ has not been eliminated by $(i', j^*_{i'})$. According to the mechanism of the inner-layer elimination and its related discussion in Section 4.1,

$$\max_{(i',j)\in\mathcal{S}_{i'}(n)} \bar{X}_{i'j}(n) - \bar{X}_{i'1}(n) = \bar{X}_{i'j^*_{i'}}(n) - \bar{X}_{i'1}(n) < \frac{g_c(t_{i'1,i'j^*_{i'}}(n))}{t_{i'1,i'j^*_{i'}}(n)} \leq C_{i'}(n). \tag{EC.6}$$

Plugging (EC.6) into (EC.5) yields

$$U_{ii'}(n) \geq \bar{X}_{i1}(n) - \bar{X}_{i'1}(n) + \frac{g_c(t_{i1,i'1}(n))}{t_{i1,i'1}(n)}. \tag{EC.7}$$

Likewise, we can show that

$$L_{ii'}(n) \leq \bar{X}_{i1}(n) - \bar{X}_{i'1}(n) - \frac{g_c(t_{i1,i'1}(n))}{t_{i1,i'1}(n)}. \tag{EC.8}$$

By (EC.7) and (EC.8), if $(i,1)\in\mathcal{S}_i(n)$ and $(i',1)\in\mathcal{S}_{i'}(n)$ for all $n\geq 1$, then

$$\begin{aligned}
&\mathbb{P}\left\{\mu_{i1}-\mu_{i'1}\notin (L_{ii'}(n), U_{ii'}(n)) \text{ for some } n\geq 1\right\}\\
&\leq \mathbb{P}\left\{\mu_{i1}-\mu_{i'1}\notin \left(\bar{X}_{i1}(n)-\bar{X}_{i'1}(n)-\frac{g_c(t_{i1,i'1}(n))}{t_{i1,i'1}(n)}, \bar{X}_{i1}(n)-\bar{X}_{i'1}(n)+\frac{g_c(t_{i1,i'1}(n))}{t_{i1,i'1}(n)}\right) \text{ for some } n\geq 1\right\}\\
&= \mathbb{P}\left\{t_{i1,i'1}(n)|\bar{X}_{i1}(n)-\bar{X}_{i'1}(n)-(\mu_{i1}-\mu_{i'1})| \geq g_c(t_{i1,i'1}(n)) \text{ for some } n\geq 1\right\}\\
&= \mathbb{P}\left\{N_{i1,i'1}<\infty\right\},
\end{aligned} \tag{EC.9}$$

where $N_{i1,i'1} = \min\{n\geq 1: t_{i1,i'1}(n)|\bar{X}_{i1}(n)-\bar{X}_{i'1}(n)-(\mu_{i1}-\mu_{i'1})|\geq g_c(t_{i1,i'1}(n))\}$. It follows from Lemma EC.1(i) that, letting $\bar{Y}_{ii'}(n) = \bar{X}_{i1}(n)-\bar{X}_{i'1}(n)-(\mu_{i1}-\mu_{i'1})$,

$$\limsup_{\beta\to 0}\frac{1}{\beta}\mathbb{P}\left\{t_{i1,i'1}(n)\bar{Y}_{ii'}(n)\leq -g_c(t_{i1,i'1}(n)),\ N_{i1,i'1}<\infty\right\}\leq 1.$$

By the symmetry of the random walk paths,

$$\limsup_{\beta\to 0}\frac{1}{\beta}\mathbb{P}\left\{t_{i1,i'1}(n)\bar{Y}_{ii'}(n)\geq g_c(t_{i1,i'1}(n)),\ N_{i1,i'1}<\infty\right\}\leq 1.$$

Hence, the proof is completed in the light of (EC.9). More specifically,

$$\limsup_{\beta\to 0}\frac{1}{\beta}\mathbb{P}\left\{\mu_{i1}-\mu_{i'1}\notin (L_{ii'}(n), U_{ii'}(n)) \text{ for some } n\geq 1\right\}$$

$$
\begin{aligned}
&\le \limsup_{\beta\to 0}\frac{1}{\beta}\mathbb{P}\{N_{i1,i'1}<\infty\}\\
&\le \limsup_{\beta\to 0}\frac{1}{\beta}\mathbb{P}\left\{t_{i1,i'1}(N_{i1,i'1})\bar{Y}_{ii'}(N_{i1,i'1})\le -g_c(t_{i1,i'1}(N_{i1,i'1})),\ N_{i1,i'1}<\infty\right\}\\
&\quad +\limsup_{\beta\to 0}\frac{1}{\beta}\mathbb{P}\left\{t_{i1,i'1}(N_{i1,i'1})\bar{Y}_{ii'}(N_{i1,i'1})\ge g_c(t_{i1,i'1}(N_{i1,i'1})),\ N_{i1,i'1}<\infty\right\}\\
&\le 2. \quad \square
\end{aligned}
$$

EC.2.2. Proof of Theorem 2

In order to prove Theorem 2, we characterize various scenarios that can lead to an incorrect selection (ICS) event that alternative 1 is not ultimately selected. One such scenario is that in step 3.2, i.e., outer-layer elimination, alternative 1 may be eliminated because the approximate dynamic confidence interval for $\mu_{11}-\mu_{i1}$, which is constructed in the spirit of Proposition 1, is entirely to the right of the origin. In this scenario, we say "alternative 1 is *eliminated* by alternative i".

The other possible scenario for ICS is that in step 4, i.e., stopping, the stopping criterion is met with both alternative 1 and alternative i having survived, but alternative 1 has a larger worst-case sample mean than alternative i. From now on, when we say "alternative 1 is *killed* by alternative i", we mean that *either* of the above two scenarios occurs.

LEMMA EC.2. *Assume that $(1,1)\in\mathcal{S}_1(n)$ and $(i,1)\in\mathcal{S}_i(n)$ for all $n\ge 1$. Then,*

$$\limsup_{\beta\to 0}\frac{1}{\beta}\mathbb{P}\{\textit{alternative 1 is eliminated by alternative } i\}\le 1.$$

Proof of Lemma EC.2. Let M denote the sample size n when the stopping criterion is met. Define

$$\tilde{N}_{11,i1}=\min\{n\ge n_0:\tau^*_{1i}(n)[W_{1i}(n)-C_1(n)]\ge g_c(\tau^*_{1i}(n)) \text{ or } \tau^*_{1i}(n)[W_{1i}(n)+C_i(n)]\le -g_c(\tau^*_{1i}(n))\},$$

where $\tau^*_{1i}(n)=\min_{(1,j)\in\mathcal{S}_1(n),(i,j')\in\mathcal{S}_i(n)}\tau_{1j,ij'}(n), W_{1i}(n)=\max\limits_{(1,j)\in\mathcal{S}_1(n)}\bar{X}_{1j}(n)-\max\limits_{(i,j)\in\mathcal{S}_i(n)}\bar{X}_{ij}(n)$, and $C_i(n)=\max_{(i,j),(i,j')\in\mathcal{S}_i(n)}g_c(\tau_{ij,ij'}(n))/\tau_{ij,ij'}(n)$. Then, alternative 1 is eliminated by alternative i if and only if

$$
\begin{aligned}
&\{\tau^*_{1i}(\tilde{N}_{11,i1})[W_{1i}(\tilde{N}_{11,i1})-C_1(\tilde{N}_{11,i1})]\ge g_c(\tau^*_{1i}(\tilde{N}_{11,i1})),\ \tilde{N}_{11,i1}\le M<\infty\}\\
&\subseteq \{\tau^*_{1i}(\tilde{N}_{11,i1})[W_{1i}(\tilde{N}_{11,i1})-C_1(\tilde{N}_{11,i1})]\ge g_c(\tau^*_{1i}(\tilde{N}_{11,i1})),\ \tilde{N}_{11,i1}<\infty\}. \qquad \text{(EC.10)}
\end{aligned}
$$

Since $(1,1)\in\mathcal{S}_1(n)$, following the argument for deriving (EC.6) we can show that $\max_{(1,j)\in\mathcal{S}_1(n)}\bar{X}_{1j}(n)-\bar{X}_{11}<C_1(n)$. Hence,

$$W_{1i}(n)-C_1(n)=\max_{(1,j)\in\mathcal{S}_1(n)}\bar{X}_{1j}(n)-\max_{(i,j)\in\mathcal{S}_i(n)}\bar{X}_{ij}(n)-C_1(n)<\bar{X}_{11}(n)-\bar{X}_{i1}(n) \qquad \text{(EC.11)}$$

It is easy to see that $g_c(t)/t$ is decreasing in $t > 0$, and thus

$$\frac{g_c\tau^*_{1i}(n)}{\tau^*_{1i}(n)} = \max_{(1,j)\in\mathcal{S}_1(n),(i,j')\in\mathcal{S}_i(n)} \frac{g_c(\tau_{1j,ij'}(n))}{\tau_{1j,ij'}(n)} \geq \frac{g_c(\tau_{11,i1}(n))}{\tau_{11,i1}(n)}, \tag{EC.12}$$

where the inequality holds because $(1,1)\in\mathcal{S}_1(n)$ and $(i,1)\in\mathcal{S}_i(n)$. It follows from (EC.10), (EC.11) and (EC.12) that

$$0 \leq [W_{1i}(n) - C_1(n)] - \frac{g_c(\tau^*_{1i}(n))}{\tau^*_{1i}(n)} < [\bar{X}_{11}(n) - \bar{X}_{i1}(n)] - \frac{g_c(\tau_{11,i1}(n))}{\tau_{11,i1}(n)}.$$

Therefore,

$$\begin{aligned}
&(\text{EC.10})\\
&\subseteq \left\{\tau_{11,i1}(\tilde{N}_{11,i1})[\bar{X}_{11}(\tilde{N}_{11,i1}) - \bar{X}_{i1}(\tilde{N}_{11,i1})] \geq g_c(\tau_{11,i1}(\tilde{N}_{11,i1})),\ \tilde{N}_{11,i1} < \infty\right\}\\
&\subseteq \left\{\tau_{11,i1}(\tilde{N}_{11,i1})[\bar{X}_{11}(\tilde{N}_{11,i1}) - \bar{X}_{i1}(\tilde{N}_{11,i1}) - (\mu_{11}-\mu_{i1})] \geq g_c(\tau_{11,i1}(\tilde{N}_{11,i1})),\ \tilde{N}_{11,i1} < \infty\right\}.
\end{aligned} \tag{EC.13}$$

Define $N_{11,i1} = \min\{n \geq n_0 : \left|\tau_{11,i1}(n)[\bar{X}_{11}(n) - \bar{X}_{i1}(n) - (\mu_{11}-\mu_{i1})]\right| \geq g_c(\tau_{11,i1}(n))\}$. By (EC.10) and (EC.13), in order to prove Lemma EC.2 it suffices to show

$$\limsup_{\beta\to 0} \frac{1}{\beta}\mathbb{P}\{\tau_{11,i1}(\tilde{N}_{11,i1})[\bar{X}_{11}(\tilde{N}_{11,i1}) - \bar{X}_{i1}(\tilde{N}_{11,i1}) - (\mu_{11}-\mu_{i1})] \geq g_c(\tilde{N}_{11,i1})),\ \tilde{N}_{11,i1} < \infty\} \leq 1. \tag{EC.14}$$

Following the proof of Theorem 2 of Fan et al. (2016) and the functional central limit theorem, it can be shown that the left-hand-side of the inequality (EC.14) is upper bounded by its counterpart for the standard Brownian motion, i.e.,

$$\begin{aligned}
&\limsup_{\beta\to 0} \frac{1}{\beta}\mathbb{P}\{\tau_{11,i1}(\tilde{N}_{11,i1})[\bar{X}_{11}(\tilde{N}_{11,i1}) - \bar{X}_{i1}(\tilde{N}_{11,i1}) - (\mu_{11}-\mu_{i1})] \geq g_c(\tilde{N}_{11,i1})),\ \tilde{N}_{11,i1} < \infty\}\\
&\leq \limsup_{\beta\to 0} \mathbb{P}\{B(\tilde{T}_{11,i1}) \geq g_c(\tilde{T}_{11,i1}),\ \tilde{T}_{11,i1} < \infty\},
\end{aligned} \tag{EC.15}$$

where $\tilde{T}_{11,i1}$ is the random time that can be seen as the limit of $\tilde{N}_{11,i1}$ as $n\to\infty$. Its explicit form can be written by applying the functional central limit theorem, but we omit it since it is quite involved. Moreover, (EC.13) implies that $T \leq \tilde{T}_{11,i1}$ and $|B(\tilde{T}_{11,i1})| \geq g_c(\tilde{T}_{11,i1})$, where $T = \inf\{t > 0 : |B(t)| \geq g_c(t)\}$. By the symmetry of standard Brownian motion $B(\cdot)$, we have

$$\mathbb{P}\{B(\tilde{T}_{11,i1}) \geq g_c(\tilde{T}_{11,i1}),\ \tilde{T}_{11,i1} < \infty\} \leq \mathbb{P}\{B(T) \geq g_c(T),\ T < \infty\}. \tag{EC.16}$$

Combining (EC.15) and (EC.16),

$$\begin{aligned}
&\limsup_{\beta\to 0} \frac{1}{\beta}\mathbb{P}\{\tau_{11,i1}(\tilde{N}_{11,i1})[\bar{X}_{11}(\tilde{N}_{11,i1}) - \bar{X}_{i1}(\tilde{N}_{11,i1}) - (\mu_{11}-\mu_{i1})] \geq g_c(\tilde{N}_{11,i1})),\ \tilde{N}_{11,i1} < \infty\}\\
&\leq \limsup_{\beta\to 0} \frac{1}{\beta}\mathbb{P}\{B(T) \geq g_c(T),\ T < \infty\}\\
&= \limsup_{\beta\to 0} \frac{1}{\beta}\cdot\frac{1}{2}e^{-c/2} = 1,
\end{aligned}$$

where the equality follows from Example 6 in Jennen and Lerche (1981). Therefore, (EC.14) is true and the proof is complete. $\square$

LEMMA EC.3. *Assume that $(1,1)\in\mathcal{S}_1(n)$ and $(i,1)\in\mathcal{S}_i(n)$ for all $n\geq 1$. If $\mu_{i1}-\mu_{11}\geq\delta$, then*

$$\limsup_{\beta\to 0}\frac{1}{\beta}\mathbb{P}\{\textit{alternative 1 is killed by alternative } i\}\leq 1.$$

Proof of Lemma EC.3. We follow the notation in the proof of Lemma EC.2. Notice that alternative 1 is killed by alternative i either immediately after the stopping criterion is met, i.e.,

$$\{W_{1i}(M)>0,\ M<\tilde{N}_{11,i1}\wedge\infty\}\bigcap\{\tau_{1i}^*(M)[\delta-C_1(M)\vee C_i(M)]\geq g_c(\tau_{1i}^*(M))\},\tag{EC.17}$$

or before the stopping criterion is met, i.e.,

$$\{\tau_{1i}^*(\tilde{N}_{11,i1})[W_{1i}(\tilde{N}_{11,i1})-C_1(\tilde{N}_{11,i1})]\geq g_c(\tau_{1i}^*(\tilde{N}_{11,i1})),\ \tilde{N}_{11,i1}\leq M<\infty\}.\tag{EC.18}$$

Let $W_{1i}^0(n)=W_{1i}(n)-(\mu_{11}-\mu_{i1})$ and

$$\tilde{N}_{11,i1}^0=\min\{n\geq n_0:\tau_{1i}^*(n)[W_{1i}^0(n)-C_1(n)]\geq g_c(\tau_{1i}^*(n))\text{ or }\tau_{1i}^*(n)[W_{1i}^0(n)+C_i(n)]\leq -g_c(\tau_{1i}^*(n))\}.$$

Then,

$$\begin{aligned}
\text{(EC.17)} &= \{W_{1i}^0(M)>\mu_{i1}-\mu_{11},\ M<\tilde{N}_{11,i1}\wedge\infty\}\bigcap\{\tau_{1i}^*(M)[\delta-C_1(M)\vee C_i(M)]\geq g_c(\tau_{1i}^*(M))\}\\
&\subseteq \{W_{1i}^0(M)>\delta,\ M<\tilde{N}_{11,i1}\wedge\infty\}\bigcap\{\tau_{1i}^*(M)[\delta-C_1(M)]\geq g_c(\tau_{1i}^*(M))\}\\
&\subseteq \left\{\tau_{1i}^*(M)[W_{1i}^0(M)-C_1(M)]\geq g_c(\tau_{1i}^*(M)),\ M<\tilde{N}_{11,i1}\wedge\infty\right\}\\
&\subseteq \left\{\tau_{1i}^*(M)[W_{1i}^0(M)-C_1(M)]\geq g_c(\tau_{1i}^*(M)),\ \tilde{N}_{11,i1}^0\leq M<\tilde{N}_{11,i1}\wedge\infty\right\},
\end{aligned}\tag{EC.19}$$

where the last step follows from the definition of $\tilde{N}_{11,i1}^0$. Moreover, notice that

$$\begin{aligned}
\{M<\tilde{N}_{11,i1}\} &\subseteq \{\tau_{1i}^*(n)[W_{1i}(n)+C_i(n)]>-g_c(\tau_{1i}^*(n))\text{ for all }n\leq M\}\\
&\subseteq \{\tau_{1i}^*(n)[W_{1i}^0(n)+C_i(n)]>-g_c(\tau_{1i}^*(n))\text{ for all }n\leq M\},
\end{aligned}$$

since $W_{1i}^0(n)>W_{1i}(n)$. It then follows from (EC.19) that

$$\begin{aligned}
&\text{(EC.17)}\\
&\subseteq \left\{\tau_{1i}^*(N_{11,i1}^0)[W_{1i}^0(N_{11,i1}^0)-C_1(N_{11,i1}^0)]\geq g_c(\tau_{1i}^*(N_{11,i1}^0)),\ \tilde{N}_{11,i1}^0\leq M<\tilde{N}_{11,i1}\wedge\infty\right\}.
\end{aligned}\tag{EC.20}$$

For (EC.18), the other scenario that can lead to ICS, we notice that since $W_{1i}^0(n)>W_{1i}(n)$,

$$\begin{aligned}
&\text{(EC.18)}\\
&\subseteq \{\tau_{1i}^*(\tilde{N}_{11,i1})[W_{1i}^0(\tilde{N}_{11,i1})-C_1(\tilde{N}_{11,i1})]\geq g_c(\tau_{1i}^*(\tilde{N}_{11,i1})),\ \tilde{N}_{11,i1}\leq M<\infty\}\\
&\subseteq \{\tau_{1i}^*(\tilde{N}_{11,i1})[W_{1i}^0(\tilde{N}_{11,i1})-C_1(\tilde{N}_{11,i1})]\geq g_c(\tau_{1i}^*(\tilde{N}_{11,i1})),\ \tilde{N}_{11,i1}^0\leq\tilde{N}_{11,i1}\leq M<\infty\}.
\end{aligned}\tag{EC.21}$$

Moreover,

$$\{\tilde{N}^0_{11,i1} \le \tilde{N}_{11,i1}\} \subseteq \{\tau^*_{1i}(n)[W_{1i}(n)+C_i(n)] > -g_c(\tau^*_{1i}(n)) \text{ for all } n < \tilde{N}^0_{1i}\}$$
$$\subseteq \{\tau^*_{1i}(n)[W^0_{1i}(n)+C_i(n)] > -g_c(\tau^*_{1i}(n)) \text{ for all } n < \tilde{N}^0_{1i}\}.$$

It then follows from (EC.21) that

$$\text{(EC.18)} \subseteq \{\tau^*_{1i}(\tilde{N}^0_{11,i1})[W^0_{1i}(\tilde{N}^0_{11,i1}) - C_1(\tilde{N}^0_{11,i1})] \ge g_c(\tau^*_{1i}(\tilde{N}^0_{11,i1})),\ \tilde{N}^0_{11,i1} \le \tilde{N}_{11,i1} \le M < \infty\}. \tag{EC.22}$$

By (EC.20) and (EC.22),

$$\mathbb{P}\{\text{alternative 1 is killed by alternative } i\}$$
$$\le \mathbb{P}\{\tau^*_{1i}(\tilde{N}^0_{11,i1})[W^0_{1i}(\tilde{N}^0_{11,i1}) - C_1(\tilde{N}^0_{11,i1})] \ge g_c(\tau^*_{1i}(\tilde{N}^0_{11,i1})),\ \tilde{N}^0_{11,i1} < \infty\}.$$

Hence, in order to prove Lemma EC.3, it suffices to show

$$\limsup_{\beta\to 0} \frac{1}{\beta}\mathbb{P}\{\tau^*_{1i}(\tilde{N}^0_{11,i1})[W^0_{1i}(\tilde{N}^0_{11,i1}) - C_1(\tilde{N}^0_{11,i1})] \ge g_c(\tau^*_{1i}(\tilde{N}^0_{11,i1})),\ \tilde{N}^0_{11,i1} < \infty\} \le 1.$$

This can be done by adopting a proof that is essentially identical to the discussion between (EC.10) and the end of the proof of Lemma EC.2. □

To establish Theorem 2, we need one additional building block. Notice that a common assumption shared by both Lemma EC.2 and Lemma EC.3 is that when alternative i is compared with another alternative, system $(i,1)$, the worst system of alternative i, is not yet eliminated in the inner-layer elimination. This assumption essentially guarantees that the worst-case mean performance of alternative i can be accurately estimated via a dynamic confidence interval; see Proposition 1. Therefore, we need to characterize the probability that system $(i,1)$ is eliminated by some other system (i,j).

Lemma EC.4. *In the inner-layer selection process of the sequential RSB procedure,*

$$\limsup_{\beta\to 0} \frac{1}{\beta}\mathbb{P}\{\textit{system } (i,1) \textit{ is eliminated by system } (i,j)\} \le 1,$$

for each $i=1,2,\ldots,k$ *and each* $j=2,3,\ldots,m$.

Proof of Lemma EC.4. Define $N'_{i1,ij} = \min\{n \ge n_0 : \tau_{i1,ij}(n)|\bar{X}_{i1}(n) - \bar{X}_{ij}(n)| \ge g_c(\tau_{i1,ij}(n))\}$, for each $i=1,2,\ldots,k$ and each $j=2,3,\ldots,m$. Then,

$$\mathbb{P}\{\text{system } (i,1) \text{ is eliminated by system } (i,j)\}$$
$$= \mathbb{P}\left\{\bar{X}_{i1}(N_{i1,ij}) - \bar{X}_{ij}(N'_{i1,ij}) \le -\frac{g_c(\tau_{i1,ij}(N'_{i1,ij}))}{\tau_{i1,ij}(N'_{i1,ij})}, N'_{i1,ij} < \infty\right\}$$
$$\le \mathbb{P}\left\{\bar{X}_{i1}(N_{i1,ij}) - \bar{X}_{ij}(N'_{i1,ij}) - (\mu_{i1} - \mu_{ij}) \le -\frac{g_c(\tau_{i1,ij}(N'_{i1,ij}))}{\tau_{i1,ij}(N'_{i1,ij})}, N'_{i1,ij} < \infty\right\},$$

where the inequality holds because $\mu_{i1}-\mu_{ij}>0$. The proof is completed by applying Lemma EC.1(ii). $\square$

Proof of Theorem 2. We first notice that if $\mu_{k1}-\mu_{11}\le\delta$, then by Assumption 1 $\mu_{i1}-\mu_{11}\le\delta$ for all $i=1,\dots,k$, which implies that $\mathbb{P}(\mu_{i^*1}-\mu_{11}\le\delta)=1$ and the theorem trivially holds. Hence, without loss of generality we assume that there exists $l=1,\dots,k-1$ for which $\mu_{l1}-\mu_{11}\le\delta$ and $\mu_{l+1,1}-\mu_{11}>\delta$. Then, a good selection event (i.e., alternative i is selected for any $i=1,\dots,l$) occurs if $\bigcap_{i=l+1}^k\{\text{alternative 1 kills alternative } i\}$. We denote

$$\begin{aligned}
A &= \bigcap_{i=l+1}^{k}\{\text{alternative 1 kills alternative } i\}\\
B &= \bigcap_{i=2}^{l}\{\text{alternative 1 is not eliminated by alternative } i\}\\
C &= \bigcap_{i=1}^{k}\bigcap_{j=2}^{m}\{\text{system } (i,1) \text{ is not eliminated by system } (i,j)\}.
\end{aligned}$$

Clearly, $\mathbb{P}(A\cap B|C)\ge 1-\mathbb{P}(A^{\mathsf{c}}|C)-\mathbb{P}(B^{\mathsf{c}}|C)$. Multiplying $\mathbb{P}(C)$ on both sides this inequality yields

$$\mathbb{P}(A\cap B\cap C)\ge\mathbb{P}(C)-\mathbb{P}(A^{\mathsf{c}}\cap C)-\mathbb{P}(B^{\mathsf{c}}\cap C)=1-\mathbb{P}(C^{\mathsf{c}})-\mathbb{P}(A^{\mathsf{c}}\cap C)-\mathbb{P}(B^{\mathsf{c}}\cap C).$$

Since $\mathbb{P}\{\mu_{i^*1}-\mu_{11}\le\delta\}\ge\mathbb{P}(A)\ge\mathbb{P}(A\cap B\cap C)$, it follows that

$$\mathbb{P}\{\mu_{i^*1}-\mu_{11}>\delta\}\le\mathbb{P}(A^{\mathsf{c}}\cap C)+\mathbb{P}(B^{\mathsf{c}}\cap C)+\mathbb{P}(C^{\mathsf{c}}). \tag{EC.23}$$

Notice that

$$\begin{aligned}
\mathbb{P}(A^{\mathsf{c}}\cap C) &\le \sum_{i=l+1}^{k}\mathbb{P}\{\text{alternative 1 is killed by alternative } i,\ (1,1)\in\mathcal{S}_1(n) \text{ and } (i,1)\in\mathcal{S}_i(n) \text{ for all n}\}\\
\mathbb{P}(B^{\mathsf{c}}\cap C) &\le \sum_{i=1}^{l}\mathbb{P}\{\text{alternative 1 is eliminated by alternative } i,\ (1,1)\in\mathcal{S}_1(n) \text{ and } (i,1)\in\mathcal{S}_i(n) \text{ for all n}\}\\
\mathbb{P}(C^{\mathsf{c}}) &\le \sum_{i=1}^{k}\sum_{j=2}^{m}\mathbb{P}\{\text{system } (i,1) \text{ is eliminated by system } (i,j)\}.
\end{aligned}$$

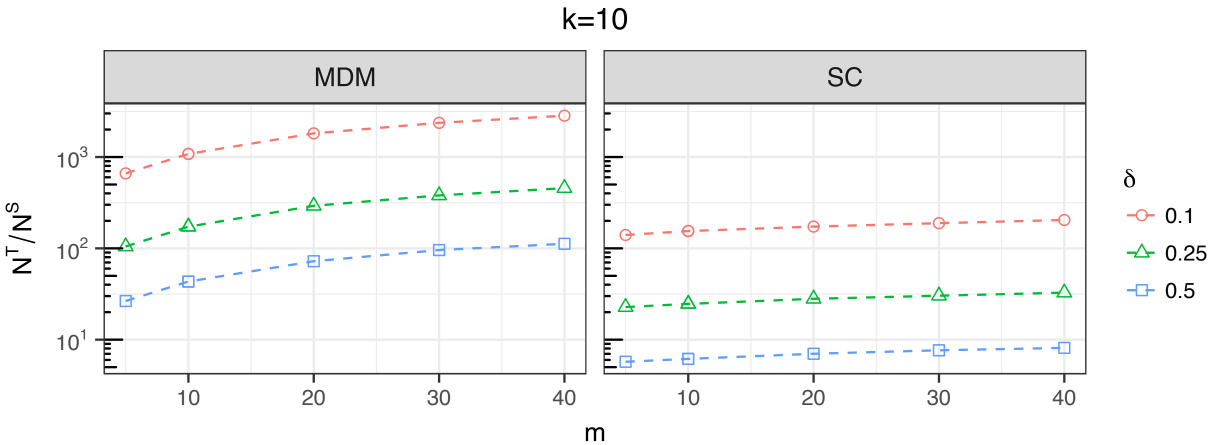
The proof concludes by combining Lemma EC.2, Lemma EC.3, Lemma EC.4, and (EC.23) to obtain

$$\limsup_{\alpha\to0}\frac{1}{\alpha}\mathbb{P}\{\mu_{i^*1}-\mu_{11}>\delta\}=\limsup_{\beta\to0}\frac{\beta}{\alpha}\cdot\frac{1}{\beta}\mathbb{P}\{\mu_{i^*1}-\mu_{11}>\delta\}\le\frac{1}{km-1}[(k-l)+l+k(m-1)]=1. \quad\square$$

EC.3. Additional Comparison Between Procedure T and Procedure S

We fix $k=10$ and plot in Figure 3 the ratio of the average sample size of Procedure T to that of Procedure S (i.e., N^T/N^S) as a function of m. The result associated with the case of fixing $m=10$ and varying k is almost identical and thus it is omitted.

Figure EC.1 Average Sample Sizes of Procedure T and Procedure S Under the EV Configuration



Note. The vertical axis is on a logarithmic scale with base 10.

EC.4. Comparison Between Procedure S and Procedure V

Procedure 3 in Fan et al. (2016) requires an IZ parameter and we set it to be $\delta/2$ when the procedure is applied to both the inner-layer and outer-layer selection of Procedure V. This is inspired by the decomposition of the IZ parameter in Section 3.1 for the two-stage RSB procedure.

Procedure V

0. *Setup.* Specify the error allowance $\beta = \alpha/(km-1)$ and the first-stage sample size $n_0 \geq 2$. Set $c = -2\log(2\beta)$ and $g_c(t) = \sqrt{[c+\log(t+1)](t+1)}$.
1. *Initialization:* Set $n = n_0$. Set $\mathcal{S} = \{1, 2, \ldots, k\}$ to be the set of surviving alternatives. Set $\mathcal{S}_i = \{(i,j) : j = 1, 2, \ldots, m\}$ to be the set of surviving systems of alternative i, $i = 1, \ldots, k$. Take n independent replications $X_{ij,1}, \ldots, X_{ij,n}$ of each system (i,j). Solve $T\delta/2 - g_c(T) = 0$ for T^*.
2. *Inner-layer Elimination.* For each $i \in \mathcal{S}$, do the following.

   2.1 *Updating.* Compute

$$\bar{X}_{ij}(n) = \frac{1}{n}\sum_{r=1}^{n} X_{ij,r}, \quad i \in \mathcal{S},\ (i,j) \in \mathcal{S}_i,$$

$$S^2_{ij,ij'}(n) = \frac{1}{n-1}\sum_{r=1}^{n}\left[X_{ij,r} - X_{ij',r} - (\bar{X}_{ij}(n) - \bar{X}_{ij'}(n))\right]^2, \quad (i,j),(i,j') \in \mathcal{S}_i.$$

   2.2 *Screening.* Compute

$$\tau_{ij,ij'}(n) = \frac{n}{S^2_{ij,ij'}(n)} \quad \text{and} \quad Z_{ij,ij'}(n) = \tau_{ij,ij'}(n)[\bar{X}_{ij}(n) - \bar{X}_{ij'}(n)], \quad (i,j),(i,j') \in \mathcal{S}_i.$$

   Assign $\mathcal{S}_i \leftarrow \mathcal{S}_i \setminus \{(i,j) \in \mathcal{S}_i : Z_{ij,ij'}(n) \leq -g_c(\tau_{ij,ij'}(n))$ for some $(i,j') \in \mathcal{S}_i\}$.

   2.3 *Stopping.* If either $|\mathcal{S}_i| = 1$ or $\tau_{ij,ij'}(n) \geq T^*$ for all $(i,j),(i,j') \in \mathcal{S}_i$ with $j \neq j'$, then stop and select $j_i^* = \arg\max_{j:(i,j)\in\mathcal{S}_i} \bar{X}_{ij}(n)$ as the worst system. Otherwise, take one additional replication of each $(i,j) \in \mathcal{S}_i$ with $i \in \mathcal{S}$, assign $n \leftarrow n+1$, and return to step 2.1.

3. *Outer-layer Elimination.*

3.1 *Updating.* Compute

$$\bar{X}_{ij_i^*}(n) = \frac{1}{n}\sum_{r=1}^{n} X_{ij,r}, \quad i \in \mathcal{S},$$

$$S^2_{ij_i^*, i'j_{i'}^*}(n) = \frac{1}{n-1}\sum_{r=1}^{n}\left[X_{ij_i^*,r} - X_{i'j_{i'}^*,r} - (\bar{X}_{ij_i^*}(n) - \bar{X}_{i'j_{i'}^*}(n))\right]^2, \quad i,i' \in \mathcal{S}.$$

3.2 *Screening.* For each $i, i' \in \mathcal{S}$ with $i \neq i'$, compute

$$\tau_{ij_i^*, i'j_{i'}^*}(n) = \frac{n}{S^2_{ij_i^*, i'j_{i'}^*}(n)} \quad \text{and} \quad Z_{ij_i^*, i'j_{i'}^*}(n) = \tau_{ij_i^*, i'j_{i'}^*}(n)[\bar{X}_{ij_i^*}(n) - \bar{X}_{i'j_{i'}^*}(n)], \quad i,i' \in \mathcal{S}.$$

Assign $\mathcal{S} \leftarrow \mathcal{S} \setminus \{(i,j) \in \mathcal{S} : Z_{ij_i^*, i'j_{i'}^*}(n) \geq g_c(\tau_{ij_i^*, i'j_{i'}^*}(n))$ for some $i' \in \mathcal{S}\}$.

3.3 *Stopping.* If either $|\mathcal{S}| = 1$ or $\tau_{ij_i^*, i'j_{i'}^*}(n) \geq T^*$ for all $i, i' \in \mathcal{S}$ with $i \neq i'$, then stop and select $i^* = \arg\min_{i\in\mathcal{S}} \bar{X}_{ij_i^*}(n)$ as the best alternative. Otherwise, take one additional replication of each (i, j_i^*) with $i \in \mathcal{S}$, assign $n \leftarrow n+1$, and return to step 3.1. □

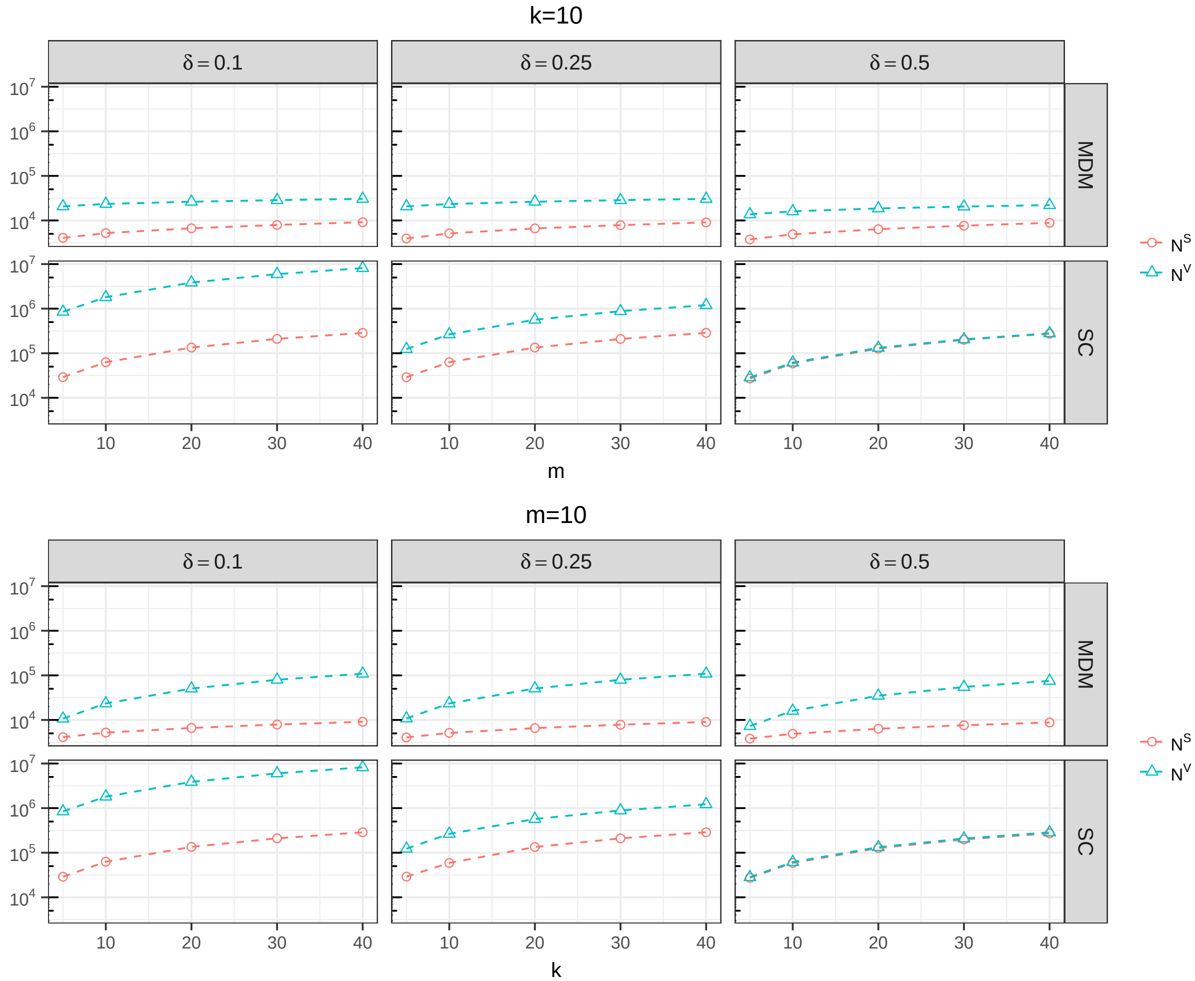
The numerical results for the EV configuration are presented in Figure EC.2. The results for the other two configurations of the variances are very similar so we omit them.

First, as expected, Procedure S requires significantly fewer samples than Procedure V in general. In particular, under SC, if the IZ parameter δ happens to be the difference between the best and the second-best worst-case mean performances (i.e., $\delta = \mu_{21} - \mu_{11} = 0.5$), then the average sample sizes required by the two procedures are almost the same, regardless of the problem scale. This implies that in this case, simultaneous elimination of the surviving systems of an alternative that is unlikely to be the best rarely happens in Procedure S, which diminishes its advantage over Procedure V. This is because under SC, the outer-layer selection process deals with the worst-case mean performances $(0, 0.5, \ldots, 0.5)$. With $\delta = 0.5$, the alternatives are hard to differentiate in early iterations of Procedure S when the sample size is large enough.

Second, the average sample size of Procedure V is more heavily affected by the configurations of the means than Procedure S. With everything else the same, the average sample size required by Procedure V (denoted by N^{V}) increases faster than that of Procedure S (denoted by N^{S}) under MDM than under SC. This suggests that there are a significantly larger number of early outer-layer eliminations in Procedure S under MDM than under SC.

Third, the average sample size of Procedure V is much more sensitive to δ under SC than that of Procedure S. For instance, with $k = 30$ and $m = 10$, N^{V} increases from about 8.82×10^5 to 5.96×10^6 as δ drops from 0.25 to 0.1, respectively, whereas N^{S} almost remains the value 2.10×10^5. The reason is as follows. The inner-layer selection process of Procedure V that relies on Procedure 3 in Fan et al. (2016) faces systems with equal means under SC. It does not terminate until its stopping

Figure EC.2 Average Sample Sizes of Procedure S and Procedure V Under the EV Configuration



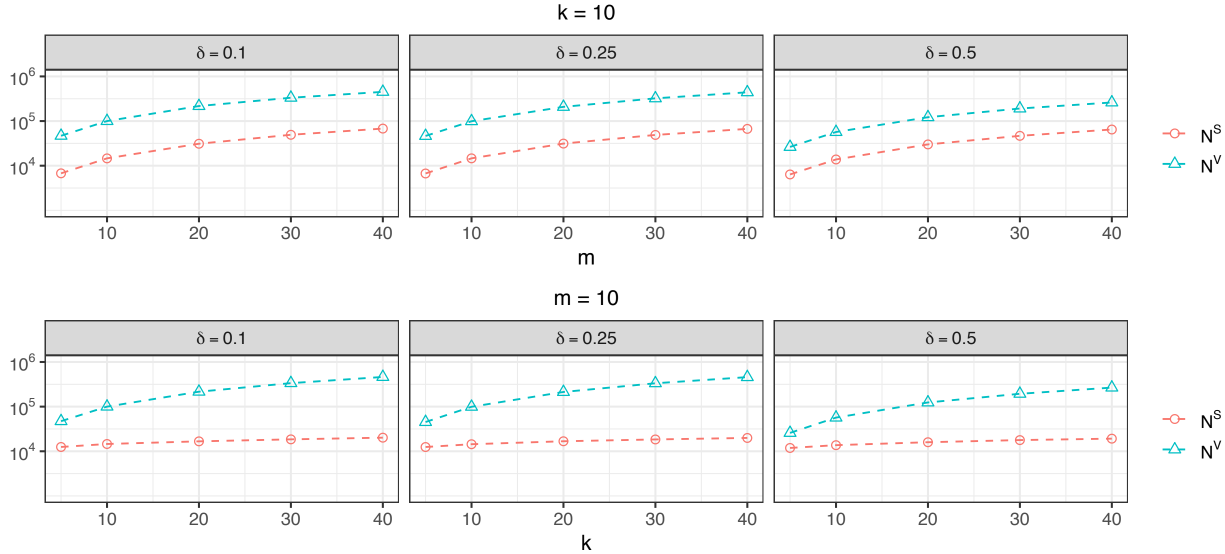
Note. Top: m varies with $k = 10$; Bottom: k varies with $m = 10$. The vertical axis is on a logarithmic scale.

criterion that depends on the IZ parameter δ is met. This stopping criterion is harder to meet for a smaller value of δ. Hence, a smaller δ implies that the inner-layer selection process of Procedure V needs more time to terminate, resulting in more required samples.

Last, N^{V} grows much faster than N^{S} as the problem scale k or m increases. For instance, with $\delta = 0.25$, $k = 10$ and MDM, N^{V} increases from about 2.90×10^4 to 2.87×10^5 as m increases from 5 to 40, whereas N^{S} increases from about 3.94×10^3 to 9.04×10^3. This is because, as the problem scale increases, there are more opportunities for Procedure S to eliminate alternatives early, leading to a slower growth in N^{S}.

The above numerical comparison between Procedure S and Procedure V indicates that the inferior performance of the latter stems from its non-fully sequential nature – its outer-layer selection cannot begin unless all the inner-layer eliminations are completed. Hence, an alternative having a configuration of the means that is close to the SC will dominate the inner-elimination time, even if

Figure EC.3 Average Sample Sizes of Procedure S and Procedure V.



Note. Top: m varies with $k = 10$; Bottom: k varies with $m = 10$. The vertical axis is on a logarithmic scale.

it were otherwise a poor alternative for outer elimination. This suggests an additional comparison between Procedure S and Procedure V using a configuration of the means that somewhat combines MDM and SC as follows

$$[\mu_{ij}]_{k\times m} = \begin{pmatrix} 0 & -0.2 & -0.2 & \dots & -0.2 \\ 0.5 & 0.3 & 0.3 & \dots & 0.3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.5(k-1) & 0.5(k-1)-0.2 & 0.5(k-1)-0.2 & \dots & 0.5(k-1)-0.2 \end{pmatrix}.$$

Here, the alternatives are ordered as MDM, but the systems of each alternative are ordered as SC.

We adopt the EV configuration of the variances. The other experiment specifications remain the same. The results are given in Figure EC.3 and they are consistent with the findings revealed by Figure EC.2.

EC.5. Realized PCS of the $G/G/s+G$ Queueing Example

This section assesses efficacy of the two proposed RSB procedures as to whether they can achieve the target PCS as promised. This complements the analysis in Section 5, since the samples in queueing simulation are not normally distributed in general, in contrast to the normal assumptions of the numerical experiments there.

With $\sigma = 2$ and $\ell = 50$, we run 1,000 macro-replications of the experiment below.

(i) Generate a sample of service times from P_0.

(ii) Construct an ambiguity set $\mathcal{P}$ based on the sample.

(iii) Compute the expected cost $\mathbb{E}[f(s,\xi)]$ with 10,000 samples for each pair (s,P), $s = 1,\dots,k$, $P \in \mathcal{P}$ so that the estimation errors are negligible and find the best alternative,

(iv) Run the two RSB procedures 1,000 times independently on $\mathcal{P}$ and estimate their respective PCS. The IZ parameter δ is set to be small enough so that the indifference zone contains only the best alternative.

In summary, there are 100 ambiguity sets constructed in total, each from one macro-replication. Hence, PCS is estimated 100 times for each RSB procedure and some statistics of these estimated probabilities are reported in Table EC.1. Clearly, both procedures can achieve the target PCS even by a large margin in general, despite the samples' non-normal distribution. Moreover, the two-stage RSB procedure is significantly more conservative than the sequential RSB procedure, producing a larger realized PCS. This reflects that the former requires a larger number of samples, which is consistent with the findings in Section 5.

Table EC.1 Realized PCS

Procedure	Statistics				
	Min	25% Quantile	Median	75% Quantile	Max
Two-stage	0.992	0.998	0.999	1.000	1.000
Sequential	0.951	0.980	0.991	0.996	1.000

Target PCS: 0.95.

References

Fan W, Hong LJ, Nelson BL (2016) Indifference-zone-free selection of the best. *Oper. Res.* 64(6):1499–1514.

Jennen C, Lerche HR (1981) First exit densities of Brownian motion through one-sided moving boundaries. of origin cheeses during ripening. *Z. Wahrsch. Verw. Gebiete* 55(2):133–148.

Stein, C (1945) A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Stat.* 16(3):243–258.

Whitt, W (2002) *Stochastic-process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues* (Springer).