e - c o m p a n i o n

ONLY AVAILABLE IN ELECTRONIC FORM

Electronic Companion—"Simulation of Coherent Risk Measures Based on Generalized Scenarios" by Vadim Lesnevski, Barry L. Nelson, and Jeremy Staum, *Management Science*, DOI 10.1287/mnsc.1070.0734.

Online Appendices

Appendix A. Algorithms

In this appendix we specify the algorithms used in the experiments reported in §5. The algorithms are implementations of the procedures developed in §4. All algorithms are stated for the case where at most q control variates are used for any system, but this includes the case $q = 0$ where control variates are not used, as explained in §4.2.

The algorithms are constructed for clarity rather than efficiency. They do not address computational issues such as how to update sample averages and variances, or the order in which to do the screening comparisons so as to reduce the number that actually have to be made.

A.1. The Standard Algorithm

This is a two-stage algorithm without screening. It is based on a procedure of Chen and Dudewicz (1976), but with ordinary sample means instead of generalized sample means, allowing for user-specified unequal error bounds associated with the lower and upper confidence limits, and using control variates.

(1) USER INPUT: The user specifies the fixed confidence interval width $L > 0$ and the lower and upper error bounds α_a and α_b in $(0, 1/2)$.

(2) ALGORITHM PARAMETERS: Choose the number of stage-0 replications $n_0 > q + 2$ and $\alpha_C < \min\{\alpha_a/k, \alpha_b\}$, the error component devoted to control variates.

(3) STAGE 0 SIMULATION: Simulate (X_{ij}, C_{ij}) for all $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n_0$.

(4) COMPUTE FINAL SAMPLE SIZES: Set $\alpha_a'' \leftarrow \alpha_a/k - \alpha_C$, $\alpha_b'' \leftarrow \alpha_b - \alpha_C$, and the scaling constant

$$c \leftarrow \frac{1}{L}(t_{n_0-q-1, \, 1-\alpha_a''} + t_{n_0-q-1, \, 1-\alpha_b''}).$$

For each $i = 1, 2, \ldots, k$, compute the residual variance $\hat{\tau}_i^2$ of regressing $X_{i1}, X_{i2}, \ldots, X_{in_0}$ on $C_{i1}, C_{i2}, \ldots, C_{in_0}$, according to Appendix D, and from it the final sample size

$$N_i \leftarrow \max\{n_0, \lceil c^2 \hat{\tau}_i^2 + \chi^2_{q_i, \, 1-\alpha_C} \rceil\}.$$

(5) STAGE 1 SIMULATION: Simulate (X_{ij}, C_{ij}) for all $i = 1, 2, \ldots, k$ and $j = n_0 + 1, n_0 + 2, \ldots, N_i$.

(6) COMPUTE CONFIDENCE INTERVAL: For each $i = 1, \ldots, k$, compute the estimate $\hat{\mu}_i$ from the regression of $X_{i1}, X_{i2}, \ldots, X_{iN_i}$ on $C_{i1}, C_{i2}, \ldots, C_{iN_i}$, according to Appendix D. Set

$$a \leftarrow \frac{1}{c} t_{n_0-q-1, \, 1-\alpha_a''} \qquad \text{and} \qquad b \leftarrow \frac{1}{c} t_{n_0-q-1, \, 1-\alpha_b''},$$

and the confidence interval is $[\max_{i=1,\ldots,k} \hat{\mu}_i - a, \max_{i=1,\ldots,k} \hat{\mu}_i + b]$.

A.2. A Two-Stage Algorithm with Screening

(1) USER INPUT: The user specifies the fixed confidence interval width $L > 0$ and the lower and upper error bounds α_a and α_b in $(0, 1/2)$.

(2) ALGORITHM PARAMETERS: Choose the number of stage-0 replications $n_0 > q + 2$, the error component $\alpha_I < \alpha_b$ devoted to screening, and $\alpha_C < \min\{\alpha_a/k, \alpha_b\}$, the error component devoted to control variates.

(3) STAGE 0 SIMULATION: Simulate (X_{ij}, C_{ij}) for all $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n_0$.
For each $h, i = 1, 2, \ldots, k$ such that $h \neq i$, set

$$\bar{D}_{hi} \leftarrow \frac{1}{n_0} \sum_{j=1}^{n_0} (X_{hj} - X_{ij}),$$

$$S_{hi}^2 \leftarrow \frac{1}{n_0 - 1} \sum_{j=1}^{n_0} (X_{hj} - X_{ij} - \bar{D}_{hi})^2, \qquad \text{and}$$

$$W_{hi} \leftarrow t_{n_0 - 1, \, 1 - \alpha_I/(k-1)} \frac{S_{hi}}{\sqrt{n_0}}.$$

Set $I \leftarrow \{h = 1, 2, \ldots, k \mid \forall i \in I(l), \bar{D}_{hi} \geq -W_{hi}\}$.

(4) COMPUTE FINAL SAMPLE SIZES: Set $\alpha_a'' \leftarrow \alpha_a/k - \alpha_C$, $\alpha_b'' \leftarrow \alpha_b - \alpha_I - \alpha_C$, and the scaling constant

$$c \leftarrow \frac{1}{L} (t_{n_0 - q - 1, \, 1 - \alpha_a''} + t_{n_0 - q - 1, \, 1 - \alpha_b''}),$$

where $q := \max_{i \in I} q_i$ and q_i is the number of control variates in C_i.

For each $i \in I$, compute the residual variance $\hat{\tau}_i^2$ of regressing $X_{i1}, X_{i2}, \ldots, X_{i, n_0}$ on $C_{i1}, C_{i2}, \ldots, C_{i, n_0}$, according to Appendix D, and from it the final sample size

$$N_i \leftarrow \max\{n_0, \lceil c^2 \hat{\tau}_i^2 + \chi_{q_i, \, 1 - \alpha_C}^2 \rceil\}.$$

(5) FINAL STAGE SIMULATION: Simulate (X_{ij}, C_{ij}) for all $i \in I$ and $j = n_0 + 1, n_0 + 2, \ldots, N_i$.

(6) COMPUTE CONFIDENCE INTERVAL: For each $i \in I$, compute the estimate $\hat{\mu}_i$ from the regression of $X_{i1}, X_{i2}, \ldots, X_{iN_i}$ on $C_{i1}, C_{i2}, \ldots, C_{iN_i}$, according to Appendix D. Set

$$a \leftarrow \frac{1}{c} t_{n_0 - q - 1, \, 1 - \alpha_a''} \qquad \text{and} \qquad b \leftarrow \frac{1}{c} t_{n_0 - q - 1, \, 1 - \alpha_b''},$$

and the confidence interval is $[\max_{i \in I} \hat{\mu}_i - a, \max_{i \in I} \hat{\mu}_i + b]$.

A.3. A Multistage Algorithm with Early Stopping

(1) USER INPUT: The user specifies the fixed confidence interval width $L > 0$ and the lower and upper error bounds α_a and α_b in $(0, 1/2)$.

(2) ALGORITHM PARAMETERS: Choose
   (a) the number of stage-0 replications $n_0 = N(0) > q + 2$,
   (b) the maximum number m of screening stages,
   (c) the number $l^* \in \{1, 2, \ldots, m - 1\}$ of screening stages at which early stopping is not allowed,
   (d) the factor $R > 1$ by which the sample size grows at each screening stage,
   (e) the error component $\alpha_I < \alpha_b$ devoted to screening, and
   (f) the error component $\alpha_C < \min\{\alpha_a/k, \alpha_b - \alpha_I\}$ devoted to control variates.

(3) INITIALIZATION: Set $l \leftarrow 0$, $I(0) \leftarrow \{1, 2, \ldots, k\}$, and $N(-1) \leftarrow 0$.

(4) SCREENING STAGE SIMULATION: Simulate (X_{ij}, C_{ij}) for all $i \in I(l)$ and $j = N(l-1) + 1, N(l-1) + 2, \ldots, N(l)$.

For each $h, i \in I(l)$ such that $h \neq i$, set

$$\bar{\bar{D}}_{hi} \leftarrow \frac{1}{N(l)} \sum_{j=1}^{N(l)} (X_{hj} - X_{ij}),$$

$$S_{hi}^2 \leftarrow \frac{1}{N(l) - 1} \sum_{j=1}^{N(l)} (X_{hj} - X_{ij} - \bar{\bar{D}}_{hi})^2, \qquad \text{and}$$

$$W_{hi} \leftarrow t_{N(l) - 1, \, 1 - \alpha_I/(m(k-1))} \frac{S_{hi}}{\sqrt{N(l)}}.$$

Set $I(l+1) \leftarrow \{h \in I(l) \mid \forall i \in I(l), \bar{\bar{D}}_{hi} \geq -W_{hi}\}$.

(5) PROCEED TO NEXT STAGE: Increment $l \leftarrow l+1$.
If $l \leq l^*$, or if $l < m$ and $|I(l)| > 1$, set $N(l) \leftarrow \lceil n_0 R^l \rceil$ and return to Step 4.
Otherwise, set $M \leftarrow l$.
(6) COMPUTE FINAL SAMPLE SIZES: Set $\alpha_a'' \leftarrow \alpha_a/k - \alpha_C$, $\alpha_b'' \leftarrow \alpha_b - \alpha_I - \alpha_C$, and the scaling constant

$$c \leftarrow \frac{1}{L}(t_{N(l^*)-q-1,1-\alpha_a''} + t_{N(l^*)-q-1,1-\alpha_b''}).$$

For each $i \in I(M)$, compute the residual variance $\hat{\sigma}_i^2$ of regressing $X_{i1}, X_{i2}, \ldots, X_{iN(l^*)}$ on $C_{i1}, C_{i2}, \ldots,$ $C_{iN(l^*)}$, according to Appendix D, and from it the final sample size

$$N_i \leftarrow \max\{N(M-1), \lceil c^2\hat{\sigma}_i^2 + \chi_{q_i,1-\alpha_C}^2 \rceil\}.$$

(7) FINAL STAGE SIMULATION: Simulate (X_{ij}, C_{ij}) for all $i \in I(M)$ and $j = N(M-1)+1, \ldots, N_i$.
(8) COMPUTE CONFIDENCE INTERVAL: For each $i \in I(M)$, compute the estimate $\hat{\mu}_i$ from the regression of $X_{i1}, X_{i2}, \ldots, X_{iN_i}$ on $C_{i1}, C_{i2}, \ldots, C_{iN_i}$, according to Appendix D. Set

$$a \leftarrow \frac{1}{c}t_{N(l^*)-q-1,1-\alpha_a''} \qquad \text{and} \qquad b \leftarrow \frac{1}{c}t_{N(l^*)-q-1,1-\alpha_b''},$$

and the confidence interval is $[\max_{i\in I(M)} \hat{\mu}_i - a, \max_{i\in I(M)} \hat{\mu}_i + b]$.

A.4. A Multistage Algorithm with Restarting
(1) USER INPUT: The user specifies the fixed confidence interval width $L > 0$ and the lower and upper error bounds α_a and α_b in $(0, 1/2)$.
(2) ALGORITHM PARAMETERS: Choose
    (a) the number of stage-0 replications $n_0 = N(0) > q + 2$,
    (b) the maximum number m of screening stages,
    (c) the factor $R > 1$ by which the sample size grows at each screening stage,
    (d) the error component $\alpha_I < \alpha_b$ devoted to screening,
    (e) the error component $\alpha_C < \min\{\alpha_a/k, \alpha_b - \alpha_I\}$ devoted to control variates, and
    (f) the prediction confidence level $0 < \epsilon < 1/2$ for use in choosing the number of replications in the first stage of the restarted procedure.
(3) INITIALIZATION: Set $l \leftarrow 0$, $I(0) \leftarrow \{1, 2, \ldots, k\}$, and $N(-1) \leftarrow 0$.
(4) SCREENING STAGE SIMULATION: Simulate (X_{ij}, C_{ij}) for all $i \in I(l)$ and $j = N(l-1) + 1,$ $N(l-1)+2, \ldots, N(l)$.
For each $h, i \in I(l)$ such that $h \neq i$, set

$$\bar{\bar{D}}_{hi} \leftarrow \frac{1}{N(l)} \sum_{j=1}^{N(l)} (X_{hj} - X_{ij}),$$

$$S_{hi}^2 \leftarrow \frac{1}{N(l)-1} \sum_{j=1}^{N(l)} (X_{hj} - X_{ij} - \bar{\bar{D}}_{hi})^2, \qquad \text{and}$$

$$W_{hi} \leftarrow t_{N(l)-1,1-\alpha_I/(m(k-1))} \frac{S_{hi}}{\sqrt{N(l)}}.$$

Set $I(l+1) \leftarrow \{h \in I(l) \mid \forall i \in I(l), \bar{\bar{D}}_{hi} \geq -W_{hi}\}$.
(5) PROCEED TO NEXT STAGE: Increment $l \leftarrow l+1$.
If $l < m$ and $|I(l)| > 1$, set $N(l) \leftarrow \lceil n_0 R^l \rceil$ and return to Step 4.
Otherwise, set $M \leftarrow l$.
(6) FIRST STAGE OF MEAN ESTIMATION: Set $\alpha_a'' \leftarrow \alpha_a/|I(M)| - \alpha_C$ and $\alpha_b'' \leftarrow \alpha_b - \alpha_I - \alpha_C$.
For each $i \in I(M)$, compute the residual variance $\hat{\tau}_i^2$ of regressing $X_{i1}, X_{i2}, \ldots, X_{iN(M-1)}$ on $C_{i1}, C_{i2}, \ldots, C_{iN(M-1)}$, according to Appendix D, and set

$$n_i \leftarrow \max\left\{q+3, \left\lceil \left(\frac{\Phi^{-1}(1-\alpha_a'') + \Phi^{-1}(1-\alpha_b'')}{L}\right)^2 \frac{(N(M-1)-1)\hat{\tau}_i^2}{\chi_{N(M-1)-1,1-\epsilon}^2} + \chi_{q_i,1-\alpha_C}^2 \right\rceil\right\}.$$

Set $n \leftarrow \min_{i \in I(M)} n_i$ and the scaling constant

$$c \leftarrow \frac{1}{L}(t_{n-q-1,\,1-\alpha_a''} + t_{n-q-1,\,1-\alpha_b''}).$$

Simulate (X_{ij}, C_{ij}) for all $i \in I(M)$ and $j = N(M-1)+1, \ldots, N(M-1)+n_i$.

For each $i \in I(M)$, compute the residual variance $\hat{\sigma}_i^2$ of regressing $X_{i,\,N(M-1)+1}, X_{i,\,N(M-1)+2}, \ldots,$ $X_{i,\,N(M-1)+n_i}$ on $C_{i,\,N(M-1)+1}, C_{i,\,N(M-1)+2}, \ldots, C_{i,\,N(M-1)+n_i}$, according to Appendix D, and set the final sample size

$$N_i \leftarrow \max\{n_i, \lceil c^2 \hat{\sigma}_i^2 + \chi_{q_i,\,1-\alpha_C}^2 \rceil\}.$$

(7) SECOND STAGE OF MEAN ESTIMATION: Simulate (X_{ij}, C_{ij}) for all $i \in I(M)$ and $j = N(M-1)+n_i+1, \ldots, N(M-1)+N_i$.

(8) COMPUTE CONFIDENCE INTERVAL: For each $i \in I(M)$, compute the estimate $\hat{\mu}_i$ from the regression of $X_{i,\,N(M-1)+1}, X_{i,\,N(M-1)+2}, \ldots, X_{i,\,N(M-1)+N_i}$ on $C_{i,\,N(M-1)+1}, C_{i,\,N(M-1)+2}, \ldots, C_{i,\,N(M-1)+N_i}$, according to Appendix D. Set

$$a \leftarrow \frac{1}{c} t_{n-q-1,\,1-\alpha_a''} \qquad \text{and} \qquad b \leftarrow \frac{1}{c} t_{n-q-1,\,1-\alpha_b''},$$

and the confidence interval is $[\max_{i \in I(M)} \hat{\mu}_i - a, \max_{i \in I(M)} \hat{\mu}_i + b]$.

Appendix B. Proofs

The proofs rely on Proposition 3.1. We show that inequalities (6) and (10) hold. Inequality (6) bounds the probability of wrongly discarding the best system during screening. We must specify some increasing functions G_a and G_b defined on the positive part of the real line and show that they satisfy inequality (10), which involves the error in estimating μ_i, the mean of system i.

PROPOSITION B.1. *If for each $i = 1, 2, \ldots, k$, the observations $X_{i1}, X_{i2}, \ldots$ are independent and identically distributed (i.i.d.) normal random variables, then the standard procedure (Algorithm A.1) without control variates makes inequalities (2) and (3) hold.*

PROOF. This procedure has no screening, so $\alpha_I = 0$, $I = \{1, 2, \ldots, k\}$, and inequality (6) holds trivially.

Let G_a and G_b be the cumulative distribution function $F_{t_{n_0-1}}$ of the t distribution with $n_0 - 1$ degrees of freedom. Because the error probability bounds α_a' and α_b' are both in $(0, 1/2)$, while $F_{t_{n_0-1}}(0) = 1/2$ and $\lim_{x \to \infty} F_{t_{n_0-1}}(x) = 1$, $G_a(0) < 1 - \alpha_a' < \lim_{x \to \infty} G_a(x)$ and $G_b(0) < 1 - \alpha_b' < \lim_{x \to \infty} G_b(x)$.

In the absence of control variates, $\hat{\mu}_i = \sum_{j=1}^{N_i} X_{ij}/N_i$. The distribution of $(\hat{\mu}_i - \mu_i)/(S_i/\sqrt{N_i})$ is t with $n_0 - 1$ degrees of freedom (Hochberg and Tamhane 1987, Theorem 2.1). By Equation (14), $cS_i/\sqrt{N_i} \leq 1$. Thus, for $x \geq 0$,

$$\Pr\{\hat{\mu}_i - \mu_i \leq x\} \geq \Pr\left\{\hat{\mu}_i - \mu_i \leq \frac{xcS_i}{\sqrt{N_i}}\right\} = \Pr\left\{\frac{\hat{\mu}_i - \mu_i}{S_i/\sqrt{N_i}} \leq xc\right\} = F_{t_{n_0-1}}(xc).$$

Similar reasoning provides the other half of inequality (10). $\square$

When we employ control variates, the terminal sample size in our procedures is of the form

$$N_i = \max\{n_0, \lceil c^2 \hat{\tau}^2 + \chi_{q,\,1-\alpha_C}^2 \rceil\}.$$

However, this formula is a convenient approximation for the exact required sample size

$$\min_{n \geq n_0}\left\{n \,\middle|\, \left(\frac{n-q}{q}\right)\left(\frac{n}{c^2 \hat{\tau}^2} - 1\right) \geq \mathscr{F}_{1-\alpha_C,\,q,\,n-q}\right\},$$

where $\mathscr{F}_{1-\alpha_C,\,q,\,n-q}$ is the $1 - \alpha_C$ quantile of the F distribution with $(q, n-q)$ degrees of freedom (Nelson and Staum 2006). Although the proofs that follow refer to algorithms incorporating the approximation, they depend on having the exact required sample size.

PROPOSITION B.2. *Suppose that for each $i = 1, 2, \ldots, k$, $X_{ij} = \mu_i + (C_{ij} - \xi_i)'\beta_i + \eta_{ij}$, where the residuals $\{\eta_{ij}, j = 1, 2, \ldots\}$ and controls $\{C_{ij}, j = 1, 2, \ldots\}$ are independent sets of i.i.d. normal random variables, β_i is an unknown constant vector, $E[C_{i1}] = \xi_i$, and $E[\eta_{i1}] = 0$. Also suppose that for each $j = 1, 2, \ldots,$ the vector*

$(X_{ij}, X_{2j}, \ldots, X_{kj})$ *is multivariate normal. Thenn the standard procedure (Algorithm* A.1) *and the two-stage procedure with screening (Algorithm* A.2) *make inequalities* (2) *and* (3) *hold.*

Proof. Inequality (10) follows from Proposition 4 of Nelson and Staum (2006), using $G_a(x) = G_b(x) = F_{t_{n_0-q-1}}(x) - \alpha_C$, where q is the number of controls.

The error probability bounds α'_a and α'_b are both in $(0, 1/2)$, while $\alpha_C < \min\{\alpha'_a, \alpha'_b\}$. From $G_a(0) = G_b(0) = 1/2 - \alpha_C$ and $\lim_{x \to \infty} G_a(x) = \lim_{x \to \infty} G_b(x) = 1 - \alpha_C$, it follows that $G_a(0) < 1 - \alpha'_a < \lim_{x \to \infty} G_a(x)$ and $G_b(0) < 1 - \alpha'_b < \lim_{x \to \infty} G_b(x)$.

If there is no screening (Algorithm A.1), inequality (6) holds trivially. If there is screening (Algorithm A.2), inequality (6) follows from reasoning along the lines of the appendix of Nelson et al. (2001b): first, by construction of I, the probability of correct screening $\Pr\{[k] \in I\} = \Pr\{\forall i = 1, 2, \ldots, k, \bar{D}_{[k]i} \geq -W_{[k]i}\}$. Next, define $\sigma^2_{hi} := \mathrm{Var}[X_h - X_i]$ and $Z_i := (\bar{D}_{[k]i} - (\mu_{[k]} - \mu_i))/(\sigma_{[k]i}/\sqrt{n_0})$, which is standard normal. By symmetry of the standard normal distribution,

$$\Pr\{[k] \in I\} = \Pr\left\{\forall i = 1, \ldots, k, Z_i \leq \frac{W_{[k]i} + (\mu_{[k]} - \mu_i)}{\sigma_{[k]i}/\sqrt{n_0}}\right\}$$

$$\geq \Pr\left\{\forall i = 1, \ldots, k, Z_i \leq t_{n_0-1, 1-\alpha_I/(k-1)} \frac{S_{[k]i}}{\sigma_{[k]i}}\right\},$$

by definition of $W_{[k]i}$ and using $\mu_{[k]} - \mu_i \geq 0$. Applying the Bonferroni inequality, the probability of correct screening is at least

$$1 - \sum_{i=1}^{k} \Pr\left\{Z_i > t_{n_0-1, 1-\alpha_I/(k-1)} \frac{S_{[k]i}}{\sigma_{[k]i}}\right\}.$$

The term for $i = [k]$ is zero because $Z_{[k]} = 0$, while the other $k - 1$ terms are $\alpha_I/(k - 1)$ because Z_i and $(n_0 - 1)(S_{[k]i}/\sigma_{[k]i})^2$ are independent and their distributions are respectively standard normal and chi-squared with $n_0 - 1$ degrees of freedom. Consequently, $\Pr\{[k] \in I\} \geq 1 - \alpha_I$. □

Proposition B.3. *Under the conditions of Proposition* B.2, *the multistage procedure with early stopping (Algorithm* A.3) *makes inequalities* (2) *and* (3) *hold.*

Proof. Inequality (6) follows from a screening error decomposition via the Bonferroni inequality:

$$\Pr\{[k] \notin I(m)\} \leq \sum_{l=0}^{m-1} \sum_{i \neq [k]} \Pr\{\bar{X}_{[k]}(l) < \bar{X}_i(l) - W_{[k]i}(l)\} \leq \sum_{l=0}^{m-1} \sum_{i \neq [k]} \frac{\alpha_I}{m(k-1)} = \alpha_I.$$

The univariate inference $\Pr\{\bar{X}_{[k]}(l) < \bar{X}_i(l) - W_{[k]i}(l)\} \leq \alpha_I/(m(k - 1))$ is the same as in the proof of Proposition B.2 because the sample sizes $N(l)$ are constants.

Inequality (10) holds with $G_a(x) = G_b(x) = F_{t_{N(l^*)-q-1}}(x) - \alpha_C$ by Proposition 4 of Nelson and Staum (2006), which applies because there is a residual variance estimator (called $\hat{\sigma}_i^2$ here and $\hat{\tau}_i^2(n_0)$ there) formed from a regression using an initial sample of a fixed number of observations (called $N(l^*)$ here and n_0 there), and the final sample size N_i is set in the same way as a function of the residual variance estimator. □

Proposition B.4. *Under the conditions of Proposition* B.2, *the multistage procedure with restarting (Algorithm* A.4) *makes inequalities* (2) *and* (3) *hold.*

Proof. Steps 6–8 of Algorithm A.4 are simply the standard algorithm (Algorithm A.1) applied with unequal initial sample sizes n_i and a set $I(M)$ of systems both of which are determined by Steps 1–5 of Algorithm A.4. We can view this as a randomly generated simulation problem, where restarting makes the random variates used in Steps 6–8 independent of the mechanism in Steps 1–5 that randomly generates the problem. We compensate for the unequal sample sizes by using $n := \min_{i \in I(M)} n_i$ in setting the degrees of freedom while computing the scaling constant c. Decreasing the degrees of freedom increases the final sample size and thus also increases the probability that the confidence

interval contains the largest mean $\mu_{[k]}$. Applying Proposition 4 of Nelson and Staum (2006) to the randomly generated problem shows that inequality (10) holds with $G_a(x) = G_b(x) = F_{t_{n-q-1}}(x) - \alpha_C$ for each $i \in I(M)$. Because there is no screening in Steps 6–8, $I = I(M)$ and Proposition 3.1 implies

$$\Pr\left\{\max_{i \in I} \mu_i \geq \max_{i \in I} \hat{\mu}_i - a\right\} \geq 1 - \alpha_a \qquad \text{and} \qquad \Pr\left\{\max_{i \in I} \mu_i \leq \max_{i \in I} \hat{\mu}_i + b\right\} \geq 1 - \alpha_b + \alpha_I.$$

The reason that the upper bound for the probability of a violation of the upper confidence limit is $\alpha_b - \alpha_I$ is that Step 6 of Algorithm A.4 sets $\alpha_b'' \leftarrow \alpha_b - \alpha_I - \alpha_C$ while the corresponding Step 4 of Algorithm A.1 sets $\alpha_b'' \leftarrow \alpha_b - \alpha_C$ because no screening takes place in the standard algorithm.

Consider the lower confidence limit and notice that $\mu_{[k]} := \max_{i=1,2,\ldots,k} \mu_i \geq \max_{i \in I} \mu_i$, whatever the subset $I \subseteq \{1, 2, \ldots, k\}$ generated by Steps 1–5 of Algorithm A.4 may be. Consequently,

$$\Pr\left\{\mu_{[k]} \geq \max_{i \in I} \hat{\mu}_i - a\right\} \geq \Pr\left\{\max_{i \in I} \mu_i \geq \max_{i \in I} \hat{\mu}_i - a\right\} \geq 1 - \alpha_a,$$

which verifies inequality (2). Next consider the upper confidence limit and notice that if $[k] \in I$, then $\mu_{[k]} := \max_{i=1,2,\ldots,k} = \max_{i \in I} \mu_i$. Consequently,

$$\Pr\left\{\mu_{[k]} \leq \max_{i \in I} \hat{\mu}_i + b\right\} \geq \Pr\left\{[k] \in I, \mu_{[k]} \leq \max_{i \in I} \hat{\mu}_i + b\right\}$$

$$= \Pr\left\{[k] \in I, \max_{i \in I} \mu_i \leq \max_{i \in I} \hat{\mu}_i + b\right\}$$

$$\geq 1 - \Pr\{[k] \notin I\} - \Pr\left\{\max_{i \in I} \mu_i > \max_{i \in I} \hat{\mu}_i + b\right\}.$$

From the result of Proposition 3.1, we found $\Pr\{\max_{i \in I} \mu_i > \max_{i \in I} \hat{\mu}_i + b\} \leq \alpha_b - \alpha_I$. Because Steps 3 to 5 of Algorithms A.3 and A.4, which perform screening, are the same, the proof of Proposition B.3 applies here and shows that inequality (6) holds: $\Pr\{[k] \notin I\} \leq \alpha_I$. The result is $\Pr\{\mu_{[k]} \leq \max_{i \in I} \hat{\mu}_i + b\} \geq 1 - \alpha_I - (\alpha_b - \alpha_I) = 1 - \alpha_b$, which verifies inequality (3). $\square$

Appendix C. Variants
This appendix discusses possible variants of the procedures discussed in the text.

C.1. Common Random Numbers

C.1.1. Grouping. Common random numbers are intended to induce positive correlation between systems, reducing the variances of the differences of their sample means, and thus facilitating screening. However, common random numbers may instead induce negative correlation between some pairs of systems, which inflates the variance of the difference of their sample means. If this were known in advance, it would be possible to divide the systems into groups such that no group contains a pair of systems with negative correlation under common random numbers. Then one would give each group its own set of common random numbers, independent of those belonging to all other groups. This approach ensures that all systems have nonnegative correlation, so that common random numbers cannot hurt screening. Moreover, this approach delivers a multiplicative error decomposition, as explained in §C.2.1. However, we found that this was not helpful for the examples we considered. To screen out an inferior system i quickly requires that there be *some* superior system h such that the expectation of the difference $\bar{X}_h - \bar{X}_i$ is large relative to its standard deviation. We found that typically a system has negative correlation only with a few of the superior systems, not all of them, and that the negative correlations are small in magnitude. Consequently, negative correlations have a very small effect on screening. The multiplicative error decomposition discussed in §C.2.1 also has only a very small effect on simulation efficiency. Thus, grouping systems to avoid negative correlation has only very slight benefits. These benefits are less important than the drawback that some pairs of systems with positive correlation are split between different groups, because one member of the pair has negative correlation with a third system, and thus the benefits of common random numbers for this pair are lost. In conclusion, we recommend not dividing systems into groups that are simulated independently.

C.1.2. Multistage Procedures without CRN. The sample size during screening should be the same for all systems when using CRN. Suppose instead that screening featured comparisons of averages over samples of unequal size, $\sum_{j=1}^{n_h} X_{hj}/n_h$ and $\sum_{j=1}^{n_i} X_{ij}/n_i$, where $n_h < n_i$. The variance of the difference between these averages is $\sigma_i^2/n_i - 2\sigma_{hi}^2/n_i + \sigma_h^2/n_h$, where $\sigma_{hi}^2 = \mathrm{Cov}(X_{hj}, X_{ij})$. Using only n_h replications to form both sample averages, $\mathrm{Var}[\sum_{j=1}^{n_h}(X_{ij} - X_{hj})/n_h] = (\sigma_i^2 - 2\sigma_{hi}^2 + \sigma_h^2)/n_h$. The change in variance due to using the extra replications X_{ij} for $j = n_h+1, n_h+2, \dots, n_i$ is $(1/n_i - 1/n_h)(\sigma_i^2 - 2\sigma_{hi}^2) = (1/n_h - 1/n_i)(2\rho_{hi}\sigma_h - \sigma_i)\sigma_i$, where ρ_{hi} is the correlation that common random numbers induce between X_h and X_i. When $\rho_{hi} > \sigma_i/(2\sigma_h)$, this change is positive, meaning that the inclusion of extra replications of X_i actually *increases* the variance of the difference used in screening, making screening less effective. Thus, when common random numbers are effective in inducing high correlation, the use of unequal sample sizes during screening is a mistake.

In the absence of common random numbers, it would be possible to allow different systems to have different sample sizes during screening, and to replace sample variances of differences S_{hi}^2 with sums of sample variances S_h^2 and S_i^2. Lesnevski et al. (2004) describe a scheme for choosing different sample sizes during screening and present numerical results for the basket put example described in §2.1. However, the presence of unequal sample sizes in screening complicates matters. The screening threshold

$$W_{hi} = t_{n_0-1,\, 1-\alpha_I/(m(k-1))}\sqrt{\frac{S_h^2}{N_h(l)} + \frac{S_i^2}{N_i(l)}}$$

in Lesnevski et al. (2004) can only be proved to deliver $\Pr\{[k] \notin I\} \leq 2\alpha_I$: see the appendix of Nelson et al. (2001b). However, $\Pr\{[k] \notin I\} \leq \alpha_I$ holds in limiting cases and held reliably in extensive simulation experiments (Nelson et al. 2001a). This issue does not affect our procedures with common random numbers.

C.2. Error Spending

C.2.1. Multiplicative Decomposition. In Lesnevski et al. (2004), we used a multiplicative decomposition $1 - \alpha_b = (1 - \alpha_I)(1 - \alpha_b')$. This is frequently possible in settings such as inequality (5); see also Wilson (2001). However, we found that multiplicative decomposition provided negligible efficiency gains over additive decomposition. Furthermore, in the presence of common random numbers, discussed in §4.3, it is easier to establish coverage bounds given an additive decomposition.

In the case of independent sampling of the X_i, or by means of Slepian's inequality (Hochberg and Tamhane 1987, Theorem A2.2.1) in the case when common random numbers induce nonnegative correlation among all systems ($\mathrm{Corr}[X_i, X_j] \geq 0$ for all i, j), one may use a multiplicative decomposition in inequality (8) instead of an additive decomposition. That is, instead of α_a/k in inequality (9), we would have $1 - (1 - \alpha_a)^{1/k}$. The result is a reduction in the required sample sizes to attain the fixed confidence interval width, but we found that this effect was negligible in practice.

C.2.2. Unequal Allocations. In inequality (8), we could allocate error unequally across systems as long as the individual error probabilities sum to α_a. If we could guess in advance some information about the systems, we might allocate less error to those systems that are more likely to be screened out or have lower variances.

When systems are simulated independently, it is possible to give unequal allocations of error in constructing the various thresholds $W_{hi}(l)$. While it would require good advance guesses about the problem's structure to motivate unequal allocations across systems, the m screening stages are different because some come before others, and the earlier ones have higher variances associated with the sample averages. Therefore, it might make sense to allocate more error to earlier stages so as to screen out systems more quickly at first, but we do not explore this possibility here.

C.3. A Different Approach to Estimating the Largest Mean
Instead of using the procedure of Chen and Dudewicz (1976), one might estimate the largest mean by first employing a selection-of-the-best procedure, and then generating independent observations to estimate the mean of the selected system. In the existing literature, selection procedures use an indifference-zone approach: they find a system whose mean is within δ of the best with a given probability. To be valid, the combined procedure must divide the fixed width L and the error probability α between selection and estimation. For example, half of L and half of the error could be allocated for selection. (Selection procedures require an indifference-zone parameter δ and only guarantee to select

Table EC.1   Efficiency of Our Multistage Procedure with Restarting Relative to a Selection-and-Estimation Procedure at 99% Confidence

	Example					
	Options portfolio			Basket put		
	Precision $= L$					
Allocated to selection	0.3%	1%	5%	0.3%	1%	5%
$\delta = L/4$	7.8	20	55	6.4	13	4.2
$\delta = L/2$	5.9	10	18	5.2	5.9	1.5
$\delta = 3L/4$	17	18	17	17	13	2.3

the best if it is δ better than all other systems. It is necessary to choose $\delta < L$ since it is impossible to recover from an incorrect selection in the estimation phase.)

The major disadvantage of such a procedure is that it should be significantly less efficient than our multistage procedures. The reason is that the total sample size required for estimation is inversely proportional to L^2 (see Equations (13) and (14)): halving L by allocating half the width to selection causes the required sample size to increase fourfold. On the other hand, for all available selection procedures, the expected number of observations required to select a single system also grows very rapidly as the indifference-zone parameter δ shrinks. For example, in the fully sequential ("KN") procedure of Kim and Nelson (2001), the distance to the screening boundary is proportional to $1/\delta^2$. However L is split between selection and estimation, the expected number of observations required for the combined procedure will be large.

To illustrate this, let us consider a combined procedure where we use the KN procedure as the selection method. Using the basket put and options portfolio examples, we performed a number of experiments allocating one quarter, one half, or three quarters of L to selection. We set L, α_a, α_b and the confidence level the same way we did in our experiments in §5, and we set the error allocated for selection here equal to the error allocated for screening in §5. Table EC.1 summarizes the results, which confirm that our multistage procedure is more efficient than a simpler procedure combining separate phases of selection and estimation. The numbers reported are ratios of the total sample size required by a procedure which follows KN by estimation to the sample size required by our multistage procedure with restarting.

Appendix D. Control Variate Estimators

This appendix provides definitions and notation for the paper's use of control variates; it is based on Nelson and Staum (2006). We assume that X_{ij}, the jth output from the simulation of system i, can be represented as

$$X_{ij} = \mu_i + (C_{ij} - \xi_i)'\beta_i + \eta_{ij}.$$

For each system $i = 1, 2, \ldots, k$ and any sample size n, $\{\eta_{ij}, j = 1, 2, \ldots, n\}$ are i.i.d. $\mathcal{N}(0, \tau_i^2)$ random variables. The $q_i \times 1$ vector C_{ij} is called the *control variate*; for fixed i and $j = 1, 2, \ldots, n$ the control variates are also i.i.d., are independent of η_{ij}, and have known expected value ξ_i. The multiplier β_i is a $q_i \times 1$ vector of unknown constants that captures the relationship between the output X_{ij} and the control C_{ij}, while η_{ij} represents that part of the variability in X_{ij} that is not explained by the controls. We define the CV estimator; the development is based on Nelson (1990).

Let

$$X_i(n) = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{in} \end{pmatrix} \quad \text{and} \quad C_i(n) = \begin{pmatrix} C_{i1} \\ C_{i2} \\ \vdots \\ C_{in} \end{pmatrix}$$

be vectors of the output and controls across all n replications from system i. Define the sample mean of the outputs and controls as $\bar{X}_i(n) := \sum_{j=1}^{n} X_{ij}/n$ and $\bar{C}_i(n) := \sum_{j=1}^{n} C_{ij}/n$. In this appendix, for clarity we append "(n)" to represent quantities defined across n replications.

To define the CV point estimator, let

$$L_i'(n) := [(C_{i1} - \bar{C}_i(n)), (C_{i2} - \bar{C}_i(n)), \dots, (C_{in} - \bar{C}_i(n))].$$

Then the CV estimator of μ_i is

$$\hat{\mu}_i(n) = \left[\frac{1}{n}\mathbf{1}_{n \times 1}' - (\bar{C}_i(n) - \xi_i)'(L_i'(n)L_i(n))^{-1}L_i'(n)\right] X_i(n)$$

$$= \bar{X}_i(n) - (\bar{C}_i(n) - \xi_i)'\hat{\beta}_i,$$

where $\mathbf{1}_{n \times 1}$ is a column n-vector whose entries all equal one, and $\hat{\beta}_i$, defined by the equations immediately above, is the usual least-squares regression slope coefficient (Nelson 1990). Also define

$$\hat{\tau}_i^2(n) := \frac{1}{n - q_i - 1} \sum_{j=1}^{n} [X_{ij} - \hat{\mu}_i(n) - (C_{ij} - \xi_i)'\hat{\beta}_i(n)]^2$$

as the residual variance estimator.

In Nelson and Staum (2006) we show that if the assumptions made in this appendix hold and C_{ij} has a multivariate normal distribution, then $\hat{\mu}_i(N_i) - \mu_i$ satisfies inequality (10) with $G_a(x) = G_b(x) = F'_{t_{n_0-q-1}}(x) - \alpha_C$.

References

See references list in the main paper.

Hochberg, Y., A. C. Tamhane. 1987. *Multiple Comparison Procedures*. John Wiley & Sons, New York.

Kim, S.-H., B. L. Nelson. 2001. A fully sequential procedure for indifference-zone selection in simulation. *ACM Trans. Model. Comput. Simulation* **11** 251–273.

Nelson, B. L. 1990. Control variate remedies. *Oper. Res.* **38** 974–992.

Nelson, B. L., J. Swann, D. Goldsman, W. Song. 2001a. Online companion for "Simple procedures for selecting the best simulated system when the number of alternatives is large." Available at http://or.pubs.informs.org/Media/Nelson.pdf.

Nelson, B. L., J. Swann, D. Goldsman, W. Song. 2001b. Simple procedures for selecting the best simulated system when the number of alternatives is large. *Oper. Res.* **49** 950–963.

Wilson, J. R. 2001. A multiplicative decomposition property of the screening-and-selection procedures of Nelson et al. *Oper. Res.* **49** 964–966.