Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

REVISITING DIRECT BOOTSTRAP RESAMPLING FOR INPUT MODEL UNCERTAINTY

Russell R. Barton

Department of SCIS
The Pennsylvania State University
210 Business Building
University Park, PA 16803, USA

Henry Lam

Department of IEOR
Columbia University
500 W. 120th Street
New York, NY 10027, USA

Eunhye Song

Department of IME
The Pennsylvania State University
310 Leonhard Building
University Park, PA 16802, USA

ABSTRACT

Metamodel-based bootstrap methods for characterizing input model uncertainty have disadvantages for settings where there are a large number of input distributions, or when using empirical distributions to drive the simulation. Early direct bootstrapping of empirical distributions did not take into account the distinction between intrinsic and extrinsic variations in the resampled quantities. When the intrinsic uncertainty is large, the result is overcoverage of the bootstrap percentile intervals. We explore ways of accounting for both sources in direct bootstrap characterization of input model uncertainty, and study the impact on confidence interval (CI) coverage. Four new bootstrap-based CIs for the expected simulation output under the unknown true distribution are proposed, basic shrinkage CI, percentile shrinkage CI, basic hierarchical bootstrap CI, and percentile hierarchical bootstrap CI, and their empirical performances are demonstrated using an example.

1 INTRODUCTION

When input models are fitted to data, the finiteness of the data introduces sampling error to the fitted input distributions. Hence, any simulation analysis that assumes the fitted input distributions are the "true" real-world distribution representing randomness in data fails to capture the uncertainty in the simulation output caused by such sampling error. In particular, confidence interval coverage can be severely affected by input uncertainty. Barton and Schruben (2001) show the coverage of nominal 90% confidence intervals for the expected waiting time of capacitated queues to be no better than 20% when the sample size of the real-world data is 500 and the sampling error of the fitted input distributions is ignored. Input model uncertainty analysis characterizes the impact of such sampling error on simulation output performance measure. For instance, one can compute the confidence interval (CI) for the performance measure that accounts for the sampling error in the fitted input distributions. Clearly, the sampling distribution of the fitted distributions is unknown, however, we may approximate the sampling distribution by bootstrapping hinging on the idea that bootstrapped samples resemble the statistical characteristics of real-world samples. Since first applied by Barton and Schruben (1993), bootstrapping has been a popular tool for quantifying

input model uncertainty (Cheng and Holland 1997; Barton and Schruben 2001; Barton et al. 2014; Song and Nelson 2015).

Input model uncertainty analysis via bootstrapping involves three steps: resampling the input data, computing the output (simulation run or runs) and using the accumulated outputs to quantify input model uncertainty by computing relevant statistics. Methods fall into three categories based on how the first two steps are completed: i) direct resampling of the empirical distributions, followed by simulation runs; ii) direct parametric bootstrap resampling, assuming that the input distribution parametric families are known, followed by simulation runs, and iii) parametric bootstrap resampling, evaluated via a previously fitted simulation metamodel of the response surface rather than via simulation runs. Bayesian approaches can be applied in settings ii) and iii), where the resampling is replaced by sampling from the posterior distribution given the sample data.

Approaches i) and ii) may exhibit overcoverage due to improper accounting for the "intrinsic error", namely the statistical error arising from the finiteness of the simulation effort in estimating the performance measure (Barton 2007). Metamodels used in the approach iii) greatly reduce this phenomenon, and when stochastic kriging is used with the metamodeling strategy, prediction error can be captured (Barton et al. 2014).

The metamodeling approach iii) has drawbacks as well. When the number of input model parameters is in the hundreds to thousands, fitting a metamodel for this number of variables becomes impractical for two reasons. First, the computational complexity of estimating the parameters of the metamodel increases as a function of the number of variables. Second, the number of simulation runs to fit such a model can be an order of magnitude greater than the number of variables, even in a favorable case. This means that the third approach may require more simulation runs than the direct bootstrapping used in the first two approaches. This problem is compounded if the metamodel must use as input an empirical distributions rather than parametric ones. For each input distribution, rather than one or two parameters, the metamodel must take as input the empirical cumulative distribution function (ecdf) values at tens or hundreds of points corresponding to the original data.

In this paper, we focus on a nonparametric framework for quantifying input uncertainty. This is motivated from the fact that the assumption of known parametric distribution families for input models introduces its own source of error, whose effect on simulation output can sometimes be hard to quantify. Empirical distributions for input models avoid this problem, at least in the asymptotic sense. This nonparametric framework necessitates the use of approach i). However, existing methods in i) encounter the problem of overcoverage when the intrinsic error is relatively large. To account for this intrinsic error, naive remedies would call for more demanding simulation runs, essentially to wash out the Monte Carlo noise, or else suffer from statistical inefficiency, as we will describe in the sequel.

Our main contribution is to introduce two new approaches to produce direct bootstrap CIs that account for both intrinsic and extrinsic errors in terms of correct coverage, and also with satisfactory statistical efficiency using affordable computational effort. To attain the latter properties, both of our introduced approaches hinge on properly adjusting the bootstrap samples from few simulation runs to exhibit statistical behaviors as if there were more runs. The first approach uses a shrinkage on the bootstrap samples that deflate their variances, while the second approach uses a hierarchical bootstrap that de-biases the quantile estimates used to construct the CI.

The remainder of this paper is as follows. Section 2 presents our mathematical formulation and the standard approaches to conduct direct bootstrapping. Sections 3 and 4 explain the problems with these approaches and present remedies using shrinkage and the hierarchical bootstrap respectively. Section 5 provides a limited computational comparison. Section 6 concludes with an assessment of implications and future work.

2 MATHEMATICAL FRAMEWORK

We follow the notation in Song and Nelson (2015). Suppose the objective of simulation analysis is to estimate a performance measure of the system, which can be represented as an expected value of a simulation output. The input distribution (cdf) that drives the simulation is denoted as F, which we allow to be multivariate. The simulation output from the rth replication of the simulation is

$$Y_r(F) = \eta(F) + \varepsilon_r(F), r = 1,2,\ldots,R,$$

where $\eta(F) \triangleq \mathrm{E}[Y_r(F)|F]$ and ε is a random variable representing the between-replication output differences due to the finiteness of the simulation runs with zero mean and finite variance $\sigma^2(F)$. Input model uncertainty is introduced when we use a fitted distribution $\widehat{F}$ in place of the true distribution F^c assuming such a distribution exists. The input distribution fitted to a size-n real-world sample is denoted by $\widehat{F}$. When $\widehat{F}$ is used to run simulations,

$$Y_r(\widehat{F}) = \eta(\widehat{F}) + \varepsilon_r(\widehat{F}) = \eta(F^c) + \Big(\eta(\widehat{F}) - \eta(F^c)\Big) + \varepsilon_r(\widehat{F}). \tag{1}$$

Note that $\widehat{F}$ is a realization of a random process but F^c is deterministic. In that context, (1) shows two sources of error in the simulation output. The first error $\eta(\widehat{F}) - \eta(F^c)$ has been called "bias error" by Cheng and Holland (2004), "extrinsic error" by Barton et al. (2014), and Song et al. (2014) define its variance as "input uncertainty." The second error, $\varepsilon_r(\widehat{F})$, comes from the finiteness of the simulation effort, which is referred to as "variance error" by Cheng (1994), "intrinsic error" by Barton et al. (2014), and "simulation error" by Song et al. (2014). It is reasonable in many cases to assume that $\varepsilon_r(\widehat{F}) \sim N(0, \sigma^2(\widehat{F}))$ but Barton and Schruben (2001) show that extrinsic error can be quite non-Gaussian.

2.1 Basic and Percentile Bootstraps

Our objective is to provide an asymptotically correct $1-\alpha$ confidence interval (CI) for $\eta(F^c)$ by finding $q_{\alpha/2}$ and $q_{1-\alpha/2}$ such that

$$\lim_{n\to\infty} \Pr\left(q_{\alpha/2} \le \eta(F^c) \le q_{1-\alpha/2}\right) = 1-\alpha,$$

while estimated $[q_{\alpha/2}, q_{1-\alpha/2}]$ may show overcoverage given finite n and simulation effort. We wish to characterize input model error for a limited sample, not with a large sample assumption that permits a delta method approach (Cheng and Holland 2004; Morgan et al. 2017). Further, we are interested in approaches appropriate for input models using empirical distributions, as opposed to the assumption of parametric distributions appearing in most prior work.

Barton and Schruben (2001) use the following direct bootstrap approach to estimate $q_{\alpha/2}$ and $q_{1-\alpha/2}$ nonparametrically. Suppose that the size-n sample data result in empirical distribution $\widehat{F}_0$. The bootstrap assumption is that, given $\widehat{F}_0$, the quantity $\widehat{F}_b$ constructed by resampling n values from the original data with replacement satisfies

$$\eta(\widehat{F}_b) - \eta(\widehat{F}_0) \xrightarrow{D} \eta(\widehat{F}_0) - \eta(F^c) \tag{2}$$

as $n \to \infty$. However, in the stochastic simulation context, $\eta(F)$ cannot be computed exactly and can only be estimated by $\bar{Y}_R(F) = \sum_{r=1}^{R} Y_r(\widehat{F}_0)/R$ given R replications. Therefore, instead of (2) we hope

$$\bar{Y}_R(\widehat{F}_b) - \eta(\widehat{F}_0) \xrightarrow{D} \bar{Y}_R(\widehat{F}_0) - \eta(F^c) \tag{3}$$

as $n \to \infty$. However, in practice $\eta(\widehat{F}_0)$ in the left-hand-side of (3) is also unknown. Therefore, we use $\bar{Y}_{R_0}(\widehat{F}_0)$ to approximate $\eta(\widehat{F}_0)$ in (3) for some large R_0. If we define the $\alpha/2$ and $1-\alpha/2$ sample quantiles

of $\bar{Y}_R(\widehat{F}_1) - \bar{Y}_{R_0}(\widehat{F}_0), \bar{Y}_R(\widehat{F}_2) - \bar{Y}_{R_0}(\widehat{F}_0), \ldots, \bar{Y}_R(\widehat{F}_B) - \bar{Y}_{R_0}(\widehat{F}_0)$ as $\widehat{\tau}_{\alpha/2}$ and $\widehat{\tau}_{1-\alpha/2}$, respectively, then a $1-\alpha$ confidence interval for $\eta(F^c)$ is constructed as

$$\bar{Y}_R(\widehat{F}_0) - \widehat{\tau}_{1-\alpha/2} \le \eta(F^c) \le \bar{Y}_R(\widehat{F}_0) - \widehat{\tau}_{\alpha/2}. \tag{4}$$

Note $\bar{Y}_R(\widehat{F}_0) \neq \bar{Y}_{R_0}(\widehat{F}_0)$, if $R < R_0$. This scheme is justified when R_0 is chosen big enough (more precisely, that R and the number of bootstrap iterations are also large, but these can be of smaller order than R_0). The interval (4) corresponds to the so-called basic bootstrap in the statistics literature (Davison and Hinkley 1997).

Alternately, one may attempt to directly use the bootstrap percentile interval

$$\widehat{q}_{\alpha/2} \le \eta(F^c) \le \widehat{q}_{1-\alpha/2}. \tag{5}$$

where $\widehat{q}_{\alpha/2}$ and $\widehat{q}_{1-\alpha/2}$ are the $\alpha/2$ and $1-\alpha/2$ sample quantiles of $\bar{Y}_R(\widehat{F}_b)$'s. The validity of this scheme, however, relies on choosing $R = R_0$ and the assumption that the distribution of $\bar{Y}_R(\widehat{F}_b) - \bar{Y}_{R_0}(\widehat{F}_0)$ is symmetric. However, when R is small, (5) may give an incorrect coverage.

To illustrate the last point, consider the basic bootstrap in (4). To make our discussion more concrete, we write $\widehat{\tau}_p(\xi)$ as the p sample quantile of a generic random variable ξ, so that $\widehat{\tau}_{1-\alpha/2}$ defined in (4) can be written as $\widehat{\tau}_{1-\alpha/2}(\bar{Y}_R(\widehat{F}_b) - \bar{Y}_{R_0}(\widehat{F}_0))$. Here the quantile is taken with respect to the randomness of $\widehat{F}_b$ given $\widehat{F}_0$. The lower bound in (4) is therefore

$$\bar{Y}_R(\widehat{F}_0) - \widehat{\tau}_{1-\alpha/2}(\bar{Y}_R(\widehat{F}_b) - \bar{Y}_{R_0}(\widehat{F}_0)). \tag{6}$$

If the distribution of $\bar{Y}_R(\widehat{F}_b) - \bar{Y}_{R_0}(\widehat{F}_0)$ is symmetric, then $-\widehat{\tau}_{1-\alpha/2}(\bar{Y}_R(\widehat{F}_b) - \bar{Y}_{R_0}(\widehat{F}_0)) = \widehat{\tau}_{\alpha/2}(\bar{Y}_R(\widehat{F}_b) - \bar{Y}_{R_0}(\widehat{F}_0))$, and (6) becomes

$$\bar{Y}_R(\widehat{F}_0) + \widehat{\tau}_{\alpha/2}(\bar{Y}_R(\widehat{F}_b) - \bar{Y}_{R_0}(\widehat{F}_0)). \tag{7}$$

In situations where no simulation is needed to approximate the output, $\bar{Y}_R(\widehat{F}_0)$ is the same as $\bar{Y}_{R_0}(\widehat{F}_0)$ and hence (7) is equal to $\widehat{\tau}_{\alpha/2}(\bar{Y}_R(\widehat{F}_b))$; similarly for the upper bound. This leads to the bootstrap percentile interval. However, if we have $R \neq R_0$, then we should use (7), which unfortunately does not lead to a clean bootstrap percentile formula.

Note that both the percentile and basic bootstrap intervals can have problems with coverage if the bootstrap distribution is highly asymmetric. See Davison and Hinkley (1997) for alternatives for this case.

2.2 Asymptotic Normality Based Interval

The bootstrap intervals discussed above can be compared to methods based on asymptotic normality directly. Cheng and Holland (1997) suggest to use the latter to construct CIs focusing on parametric $\widehat{F}$, where the standard error can be estimated using the bootstrap. More specifically, they use

$$\bar{Y}_{R_0}(\widehat{F}_0) - z_{1-\alpha/2}\sqrt{V^2 + \frac{\sigma^2}{R_0}} \le \eta(F^c) \le \bar{Y}_{R_0}(\widehat{F}_0) + z_{1-\alpha/2}\sqrt{V^2 + \frac{\sigma^2}{R_0}} \tag{8}$$

as a $1-\alpha$ confidence interval, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution. The quantity V^2 is the variance contributed from the input noise, namely $\mathrm{Var}(\eta(\widehat{F}_0))$, and σ^2 is the variance from the simulation noise that, at least asymptotically, is given by $\sigma^2(F)$ defined previously.

The variance V^2 can be approximated by the bootstrap input variance, $\mathrm{Var}(\eta(\widehat{F}_b))$, where $\mathrm{Var}(\cdot)$ is with respect to the sampling distribution of $\widehat{F}_b$ given $\widehat{F}_0$. Introducing the notation $Y_{b,r} \equiv Y_r(\widehat{F}_b)$ for convenience, we define

$$SS_W = \sum_{b=1}^{B}\sum_{r=1}^{R}\left(Y_{b,r} - \bar{Y}_R(\widehat{F}_b)\right)^2 \text{ and } SS_B = \sum_{b=1}^{B}\left(\bar{Y}_R(\widehat{F}_b) - \frac{1}{B}\sum_{i=1}^{B}\bar{Y}_R(\widehat{F}_i)\right)^2. \tag{9}$$

Then an estimate of V^2 is

$$\frac{SS_B}{B-1} - \frac{SS_W}{BR(R-1)}. \tag{10}$$

This formula is an unbiased estimator for $\mathrm{Var}(\eta(\widehat{F}_b))$, which consists of $\mathrm{Var}(\bar{Y}_R(\widehat{F}_b))$ estimated by the first term in (10), and adjusted for the bias introduced by $\mathrm{E}[\sigma^2(\widehat{F}_b)/R]$ that is estimated by the second term in (10). When B and R are small, (10) may result in a negative value in which case $\widehat{V}^2 = 0$ (Ankenman and Nelson 2012). On the other hand, the simulation variance σ^2 typically can be estimated more straightforwardly, for instance by taking the sample variance from all the simulation runs. These quantities have been studied in Ankenman and Nelson (2012), who focus on the ratio $\mathrm{Var}(\eta(\widehat{F}_b))/\mathrm{E}[\sigma^2(\widehat{F}_b)/R]$ under the assumption of homoscedastic simulation error, i.e., $\sigma^2(\widehat{F}_b) = \sigma^2$ for all b.

3 SHRINKAGE ADJUSTMENT

To motivate our investigation, let us first compare the performance of the basic bootstrap (4) and the asymptotic-normality-based interval (8). Both of them require R_0 simulation runs to estimate a point estimate, and B bootstrap iterations, each with R simulation runs, to estimate the quantile or standard error. The basic bootstrap requires in addition a separate simulation of size R to estimate another point estimate, namely the first term in (4). For simplicity, let us consider the case where R_0 is very big (in the sense that the Monte Carlo error of the corresponding point estimate is outwashed). Then, given a large number of bootstrap samples, the overall computational efforts for the basic bootstrap and the asymptotic-normality-based approach are roughly comparable.

Under the above presumptions, the asymptotic-normality-based interval is approximately

$$\bar{Y}_{R_0}(\widehat{F}_0) - z_{1-\alpha/2}V \le \eta(F^c) \le \bar{Y}_{R_0}(\widehat{F}_0) + z_{1-\alpha/2}V,$$

which has a half-width $z_{1-\alpha/2}V$. On the other hand, the half-width of the basic bootstrap interval in (4) depends on the standard error of $\bar{Y}_R(\widehat{F}_b) - \bar{Y}_{R_0}(\widehat{F}_0)$, which is approximately $z_{1-\alpha/2}\sqrt{V^2 + \sigma^2/R}$ (recall that R_0 is assumed big). This half-width is longer than the asymptotic-normality-based approach in general, thus making the basic bootstrap less efficient. Our main investigation is to lift the construction of a quantile-based interval, using the basic or the percentile bootstrap, to the same level of efficiency as the asymptotic-normality-based approach. This is motivated from documented studies that quantile-based approach has better finite-sample coverage properties.

We consider two approaches to address the above efficiency issue. The first is a *shrinkage* operation described in Davison and Hinkley (1997) that aims to remove the "excess variation" in $\bar{Y}_R(\widehat{F}_b)$'s. To explain, recall how we have obtained (4). We started with a point estimate $\bar{Y}_R(\widehat{F}_0)$, and we approximated the distribution of $\bar{Y}_R(\widehat{F}_0) - \eta(F^c)$ with a bootstrapped version $\bar{Y}_R(\widehat{F}_b) - \eta(\widehat{F}_0)$. This ends up giving us a half-width of approximately $z_{1-\alpha/2}\sqrt{V^2 + \sigma^2/R}$. Suppose that we could use $\bar{Y}_{R_0}(\widehat{F}_0)$ as our point estimate, and that we could efficiently bootstrap the corresponding distribution of $\bar{Y}_{R_0}(\widehat{F}_b) - \eta(\widehat{F}_0)$, then we could shrink the half-width to $z_{1-\alpha/2}V$ (again, assuming R_0 is a big number). But bootstrapping $\bar{Y}_{R_0}(\widehat{F}_b) - \eta(\widehat{F}_0)$ directly is computationally intense because R_0 is big (this in particular also means placing more computational effort than the asymptotic-normality-based approach). Instead, we adjust the samples of $\bar{Y}_R(\widehat{F}_b) - \eta(\widehat{F}_0)$ so that they exhibit the same variance as $\bar{Y}_{R_0}(\widehat{F}_b) - \eta(\widehat{F}_0)$. To do so, we define

$$\widehat{Y}_R(\widehat{F}_b) = c\bar{\bar{Y}}_R + (1-c)\,\bar{Y}_R(\widehat{F}_b), 1 \le b \le B, \tag{11}$$

where $\bar{\bar{Y}}_R$ is the grand mean given by $\frac{1}{B}\sum_{i=1}^{B}\bar{Y}_R(\widehat{F}_i)$, and c satisfies

$$(1-c)^2 = \frac{B}{B-1}\frac{\mathrm{Var}(\eta(\widehat{F}_b))}{\mathrm{Var}(\eta(\widehat{F}_b)) + \sigma^2/R}. \tag{12}$$

To verify our claim, we rewrite the variance of (11) (under $\widehat{F}_0$) as follows. Defining $\bar{\bar{Y}}_{R,-b}$ as the sample mean of $\bar{Y}_R(\widehat{F}_i), \forall i \neq b$,

$$\begin{aligned}
\mathrm{Var}(\widehat{Y}_R(\widehat{F}_b)) &= \mathrm{Var}\left(c\bar{\bar{Y}}_R + (1-c)\,\bar{Y}_R(\widehat{F}_b)\right) \\
&= \mathrm{Var}\left(\frac{c}{B}((B-1)\bar{\bar{Y}}_{R,-b} + \bar{Y}_R(\widehat{F}_b)) + (1-c)\,\bar{Y}_R(\widehat{F}_b)\right) \\
&= \mathrm{Var}\left(\frac{c}{B}(B-1)\bar{\bar{Y}}_{R,-b} + \left(\frac{c}{B} + (1-c)\right)\bar{Y}_R(\widehat{F}_b)\right) \\
&= \frac{c^2(B-1)^2}{B^2}\mathrm{Var}\left(\bar{\bar{Y}}_{R,-b}\right) + \left(\frac{c}{B} + (1-c)\right)^2 \mathrm{Var}(\bar{Y}_R(\widehat{F}_b)) \\
&= \frac{c^2(B-1)^2}{B^2}\frac{1}{B-1}\left(\mathrm{Var}(\eta(\widehat{F}_b)) + \frac{\sigma^2}{R}\right) + \left(\frac{c}{B} + (1-c)\right)^2\left(\mathrm{Var}(\eta(\widehat{F}_b)) + \frac{\sigma^2}{R}\right) \\
&= \left(\frac{c^2}{B} - \frac{c^2}{B^2} + \frac{c^2}{B^2} + \frac{2c(1-c)}{B} + (1-c)^2\right)\left(\mathrm{Var}(\eta(\widehat{F}_b)) + \frac{\sigma^2}{R}\right) \\
&= \left((1-c)^2\frac{B-1}{B} + \frac{1}{B}\right)\left(\mathrm{Var}(\eta(\widehat{F}_b)) + \frac{\sigma^2}{R}\right).
\end{aligned}$$

Thus, if we choose c satisfying

$$\left((1-c)^2\frac{B-1}{B} + \frac{1}{B}\right)\left(\mathrm{Var}(\eta(\widehat{F}_b)) + \frac{\sigma^2}{R}\right) = \mathrm{Var}(\eta(\widehat{F}_b)), \tag{13}$$

then we match the variance of $\widehat{Y}_R(\widehat{F}_b)$ with $\bar{Y}_{R_0}(\widehat{F}_b)$ for large R_0. The relation (13) gives

$$(1-c)^2 = \frac{B}{B-1}\frac{\mathrm{Var}(\eta(\widehat{F}_b))}{\mathrm{Var}(\eta(\widehat{F}_b)) + \sigma^2/R} - \frac{1}{B-1}. \tag{14}$$

We have ignored the small second term in (14) when choosing the c in (12). The above derivation follows the idea in Davison and Hinkley (1997), but there they assume a linear model and in that case (12) can be shown to give a variance that is unbiased under the true distribution. Here, since we have a nonlinear simulation model to begin with, we are contended that the variance of $\widehat{Y}_R(\widehat{F}_b)$ matches that of the bootstrap variance $\mathrm{Var}(\eta(\widehat{F}_b))$.

Note that the choice of c depends on $\mathrm{Var}(\eta(\widehat{F}_b))$ and σ^2, which are unknown in general. We plug in estimates for them using ideas similar to (10), giving

$$1 - \widehat{c} = \sqrt{\max\left\{0, \frac{B}{B-1} - \frac{SS_W}{R(R-1)\,SS_B}\right\}}. \tag{15}$$

See (9) for the expressions for SS_W and SS_B.

To sum up, in the shrinkage basic bootstrap, we first obtain a point estimate $\bar{Y}_{R_0}(\widehat{F}_0)$ by running R_0 simulation runs under the empirical distribution $\widehat{F}_0$. We then obtain each $\bar{Y}_R(\widehat{F}_b)$ by resampling from $\widehat{F}_0$ and running R simulation runs, like in the basic bootstrap before. Then we form the new bootstrap output sample $\widehat{Y}_R(\widehat{F}_b)$ using (11) with c defined in (15). The $1-\alpha$ CI for $\eta(F^c)$ is constructed as

$$\bar{Y}_{R_0}(\widehat{F}_0) - \tilde{\tau}_{1-\alpha/2} \leq \eta(F^c) \leq \bar{Y}_{R_0}(\widehat{F}_0) - \tilde{\tau}_{\alpha/2}, \tag{16}$$

where $\tilde{\tau}_{1-\alpha/2}$ and $\tilde{\tau}_{\alpha/2}$ are now the $1-\alpha/2$ and $\alpha/2$ sample quantiles of $\widehat{Y}_R(\widehat{F}_b) - \bar{Y}_{R_0}(\widehat{F}_0)$.

The same idea could be applied to construct a percentile interval. Starting from the shrinkage basic bootstrap above, we have, using the notations in (6),

$$\bar{Y}_{R_0}(\widehat{F}_0) - \widehat{\tau}_{1-\alpha/2}(\widehat{Y}_R(\widehat{F}_b) - \bar{Y}_{R_0}(\widehat{F}_0)) \le \eta(F^c) \le \bar{Y}_{R_0}(\widehat{F}_0) - \widehat{\tau}_{\alpha/2}(\widehat{Y}_R(\widehat{F}_b) - \bar{Y}_{R_0}(\widehat{F}_0))$$

being a valid CI that has a half-width roughly $z_{1-\alpha/2}V$. Now, if $\bar{Y}_R(\widehat{F}_b) - \bar{Y}_{R_0}(\widehat{F}_0)$ is approximately symmetric, so is $\widehat{Y}_R(\widehat{F}_b) - \bar{Y}_{R_0}(\widehat{F}_0)$, and thus we have

$$\bar{Y}_{R_0}(\widehat{F}_0) - \widehat{\tau}_{1-\alpha/2}(\widehat{Y}_R(\widehat{F}_b) - \bar{Y}_{R_0}(\widehat{F}_0)) \approx \bar{Y}_{R_0}(\widehat{F}_0) + \widehat{\tau}_{\alpha/2}(\widehat{Y}_R(\widehat{F}_b) - \bar{Y}_{R_0}(\widehat{F}_0)) = \widehat{\tau}_{\alpha/2}(\widehat{Y}_R(\widehat{F}_b)).$$

The upper bound can be derived similarly. Hence, in the shrinkage percentile bootstrap, the CI is

$$\tilde{q}_{\alpha/2} \le \eta(F^c) \le \tilde{q}_{1-\alpha/2}, \tag{17}$$

where $\tilde{q}_{\alpha/2}$ and $\tilde{q}_{1-\alpha/2}$ are the $\alpha/2$ and $1-\alpha/2$ sample quantiles of $\widehat{Y}_R(\widehat{F}_b)$. Note that, in either the basic or the percentile bootstrap, computing $\widehat{c}$ and constructing the shrinkage interval requires no additional simulation runs.

Below, we present the complete algorithm to construct the shrinkage basic and percentile CIs.

Algorithm: Shrinkage CIs

1. Run R_0 replications of the simulator using $\widehat{F}_0$ as the input model to obtain $\bar{Y}_{R_0}(\widehat{F}_0) = \sum_{r=1}^{R_0} Y_r(\widehat{F}_0)/R_0$.
2. For $b = 1,2,\ldots,B$
   (a) Generate $\widehat{F}_b$ by resampling $\widehat{F}_0$ n times.
   (b) Using $\widehat{F}_b$, run R replications, $Y_1(\widehat{F}_b), Y_2(\widehat{F}_b), \ldots, Y_R(\widehat{F}_b)$, and compute $\bar{Y}_R(\widehat{F}_b) = \sum_{r=1}^{R} Y_r(\widehat{F}_b)/R$.
3. Using $Y_r(\widehat{F}_b), b = 1,2\ldots,B, r = 1,2,\ldots,R$, calculate $\widehat{c}$ from (15).
4. For $b = 1,\ldots,B$, compute $\widehat{Y}_R(\widehat{F}_b) = \widehat{c}\bar{\bar{Y}}_R + (1-\widehat{c})\bar{Y}_R(\widehat{F}_b)$, where $\bar{\bar{Y}} = \sum_{b=1}^{B} Y_R(\widehat{F}_b)/B$.
5. CI construction:
   (a) (**Basic CI**) Find the empirical $\alpha/2$ and $1-\alpha/2$ quantiles of $\widehat{Y}_R(\widehat{F}_1) - \bar{Y}_{R_0}(\widehat{F}_0)$, $\widehat{Y}_R(\widehat{F}_2) - \bar{Y}_{R_0}(\widehat{F}_0), \ldots, \widehat{Y}_R(\widehat{F}_B) - \bar{Y}_{R_0}(\widehat{F}_0)$, $\tilde{\tau}_{\alpha/2}$ and $\tilde{\tau}_{1-\alpha/2}$, respectively, and construct the basic shrinkage CI in (16).
   (b) (**Percentile CI**) Find the empirical $\alpha/2$ and $1-\alpha/2$ quantiles of $\widehat{Y}_R(\widehat{F}_1), \widehat{Y}_R(\widehat{F}_2), \ldots, \widehat{Y}_R(\widehat{F}_B)$, $\tilde{q}_{\alpha/2}$ and $\tilde{q}_{1-\alpha/2}$, respectively, and construct the percentile shrinkage CI in (17).

4 HIERARCHICAL BOOTSTRAP

Next we discuss a hierarchical bootstrap approach that accounts for both the extrinsic and intrinsic errors in the basic and the percentile bootstraps that, like the shrinkage, allows a comparable computational effort and statistical efficiency as the normality-based approach. We start by discussing the basic bootstrap. Again, let us assume that R_0 is large. Consider the scheme

$$\bar{Y}_{R_0}(\widehat{F}_0) - \widehat{\tau}_{1-\alpha/2}(\bar{Y}_{R_0}(\widehat{F}_b) - \eta(\widehat{F}_0)) \le \eta(F^c) \le \bar{Y}_{R_0}(\widehat{F}_0) - \widehat{\tau}_{\alpha/2}(\bar{Y}_{R_0}(\widehat{F}_b) - \eta(\widehat{F}_0)).$$

As discussed before, though this is valid, it requires huge computational effort since approximating $\widehat{\tau}_{1-\alpha/2}(\bar{Y}_{R_0}(\widehat{F}_b) - \eta(\widehat{F}_0))$ and $\widehat{\tau}_{\alpha/2}(\bar{Y}_{R_0}(\widehat{F}_b) - \eta(\widehat{F}_0))$ requires many iterations each with R_0 simulation runs. Our hierarchical bootstrap approach aims to replace these quantities with $\widehat{\tau}_{1-\alpha/2}(\bar{Y}_R(\widehat{F}_b) - \eta(\widehat{F}_0))$ and $\widehat{\tau}_{\alpha/2}(\bar{Y}_R(\widehat{F}_b) - \eta(\widehat{F}_0))$, and subsequently de-bias them back to $\widehat{\tau}_{1-\alpha/2}(\bar{Y}_{R_0}(\widehat{F}_b) - \eta(\widehat{F}_0))$ and $\widehat{\tau}_{\alpha/2}(\bar{Y}_{R_0}(\widehat{F}_b) - \eta(\widehat{F}_0))$. In other words, we want to find

$$\widehat{\tau}_{1-\alpha/2}(\bar{Y}_R(\widehat{F}_b) - \eta(\widehat{F}_0)) - \widehat{\tau}_{1-\alpha/2}(\bar{Y}_{R_0}(\widehat{F}_b) - \eta(\widehat{F}_0)) = \widehat{\tau}_{1-\alpha/2}(\bar{Y}_R(\widehat{F}_b)) - \widehat{\tau}_{1-\alpha/2}(\bar{Y}_{R_0}(\widehat{F}_b)) \tag{18}$$

and

$$\widehat{\tau}_{\alpha/2}(\bar{Y}_R(\widehat{F}_b)-\eta(\widehat{F}_0))-\widehat{\tau}_{\alpha/2}(\bar{Y}_{R_0}(\widehat{F}_b)-\eta(\widehat{F}_0))=\widehat{\tau}_{\alpha/2}(\bar{Y}_R(\widehat{F}_b))-\widehat{\tau}_{\alpha/2}(\bar{Y}_{R_0}(\widehat{F}_b)). \tag{19}$$

Since the differences in (18) and (19) stem from the difference in simulation size, we propose the following scheme. Consider the existing bootstrapped outputs $Y_b^*=\bar{Y}_R(\widehat{F}_b), b=1,\ldots,B$, where the simulation runs from each resample are averaged. For each Y_b^*, compute $Y_{b,r}^*=Y_b^*+\varepsilon_{b,r}^*, r=1,2,\ldots,R$, where $\varepsilon_{b,r}^*$ is sampled from $\varepsilon_{b,r}^*\sim N(0,S^2(\widehat{F}_b))$ and $S^2(\widehat{F}_b)$ is an estimator of $\sigma^2(\widehat{F}_b)$. Then take the average of $Y_{b,r}^*$ across r, for each b, to obtain Y_b^{**}. Use these Y_b^{**} to form the quantiles, say $\widehat{q}_{1-\alpha/2}^{**}$ and $\widehat{q}_{\alpha/2}^{**}$. Correspondingly, let $\widehat{q}_{1-\alpha/2}^{*}$ and $\widehat{q}_{\alpha/2}^{*}$ be the quantiles of Y_b^*. We then use $\widehat{q}_{1-\alpha/2}^{**}-\widehat{q}_{1-\alpha/2}^{*}$ and $\widehat{q}_{\alpha/2}^{**}-\widehat{q}_{\alpha/2}^{*}$ as estimates of (18) and (19). In this way, we approximate $\widehat{\tau}_{1-\alpha/2}(\bar{Y}_{R_0}(\widehat{F}_b)-\eta(\widehat{F}_0))$ with

$$\widehat{\tau}_{1-\alpha/2}(\bar{Y}_R(\widehat{F}_b)-\bar{Y}_{R_0}(\widehat{F}_0))-(\widehat{q}_{1-\alpha/2}^{**}-\widehat{q}_{1-\alpha/2}^{*})$$

and $\widehat{\tau}_{\alpha/2}(\bar{Y}_{R_0}(\widehat{F}_b)-\eta(\widehat{F}_0))$ with

$$\widehat{\tau}_{\alpha/2}(\bar{Y}_R(\widehat{F}_b)-\bar{Y}_{R_0}(\widehat{F}_0))-(\widehat{q}_{\alpha/2}^{**}-\widehat{q}_{\alpha/2}^{*}).$$

A $1-\alpha$ basic hierarchical bootstrap CI is

$$\begin{aligned}\bar{Y}_{R_0}(\widehat{F}_0)-\widehat{\tau}_{1-\alpha/2}(\bar{Y}_R(\widehat{F}_b)-\bar{Y}_{R_0}(\widehat{F}_0))+(\widehat{q}_{1-\alpha/2}^{**}-\widehat{q}_{1-\alpha/2}^{*})&\le\eta(F^c)\\ &\le\bar{Y}_{R_0}(\widehat{F}_0)-\widehat{\tau}_{\alpha/2}(\bar{Y}_R(\widehat{F}_b)-\bar{Y}_{R_0}(\widehat{F}_0))+(\widehat{q}_{\alpha/2}^{**}-\widehat{q}_{\alpha/2}^{*}).\end{aligned}$$

Equivalently, this is

$$2\bar{Y}_{R_0}(\widehat{F}_0)-2\widehat{q}_{1-\alpha/2}^{*}+\widehat{q}_{1-\alpha/2}^{**}\le\eta(F^c)\le 2\bar{Y}_{R_0}(\widehat{F}_0)-2\widehat{q}_{\alpha/2}^{*}+\widehat{q}_{\alpha/2}^{**}. \tag{20}$$

Roughly speaking, (20) uses $\widehat{q}_{1-\alpha/2}^{**}-\widehat{q}_{1-\alpha/2}^{*}$ and $\widehat{q}_{\alpha/2}^{**}-\widehat{q}_{\alpha/2}^{*}$ to capture the inflation in the quantiles due to the noise from R simulation runs, using a normal approximation for the simulation noise.

Similar idea applies to the percentile bootstrap. In this case, we need to approximate $\widehat{\tau}_{1-\alpha/2}(\bar{Y}_{R_0}(\widehat{F}_b))$ and $\widehat{\tau}_{\alpha/2}(\bar{Y}_{R_0}(\widehat{F}_b))$. Using $\widehat{q}_{1-\alpha/2}^{**}-\widehat{q}_{1-\alpha/2}^{*}$ and $\widehat{q}_{\alpha/2}^{**}-\widehat{q}_{\alpha/2}^{*}$ as estimates of $\widehat{\tau}_{1-\alpha/2}(\bar{Y}_R(\widehat{F}_b))-\widehat{\tau}_{1-\alpha/2}(\bar{Y}_{R_0}(\widehat{F}_b))$ and $\widehat{\tau}_{\alpha/2}(\bar{Y}_R(\widehat{F}_b))-\widehat{\tau}_{\alpha/2}(\bar{Y}_{R_0}(\widehat{F}_b))$, as for (18) and (19), and $\widehat{q}_{1-\alpha/2}^{*}$ and $\widehat{q}_{\alpha/2}^{*}$ for $\widehat{\tau}_{1-\alpha/2}(\bar{Y}_R(\widehat{F}_b))$ and $\widehat{\tau}_{\alpha/2}(\bar{Y}_R(\widehat{F}_b))$, we approximate $\widehat{\tau}_{1-\alpha/2}(\bar{Y}_{R_0}(\widehat{F}_b))$ with $\widehat{q}_{1-\alpha/2}^{*}-(\widehat{q}_{1-\alpha/2}^{**}-\widehat{q}_{1-\alpha/2}^{*})=2\widehat{q}_{1-\alpha/2}^{*}-\widehat{q}_{1-\alpha/2}^{**}$, and similarly $\widehat{\tau}_{\alpha/2}(\bar{Y}_{R_0}(\widehat{F}_b))$ with $2\widehat{q}_{\alpha/2}^{*}-\widehat{q}_{\alpha/2}^{**}$. A $1-\alpha$ hierarchical bootstrap percentile CI is

$$2\widehat{q}_{\alpha/2}^{*}-\widehat{q}_{\alpha/2}^{**}\le\eta(F^c)\le 2\widehat{q}_{1-\alpha/2}^{*}-\widehat{q}_{1-\alpha/2}^{**}. \tag{21}$$

Below, we present the complete algorithm to construct the hierarchical basic and percentile CIs.

Algorithm: Hierarchical Bootstrap CIs

1. Run R_0 replications of the simulator using $\widehat{F}_0$ as the input model to obtain $\bar{Y}_{R_0}(\widehat{F}_0)=\sum_{r=1}^{R_0}Y_r(\widehat{F}_0)/R_0$.
2. For $b=1,2,\ldots,B$
   (a) Generate $\widehat{F}_b$ by resampling $\widehat{F}_0$ n times.
   (b) Using $\widehat{F}_b$, run R replications, $Y_1(\widehat{F}_b),Y_2(\widehat{F}_b),\ldots,Y_R(\widehat{F}_b)$, and compute $\bar{Y}_R(\widehat{F}_b)=\sum_{r=1}^{R}Y_r(\widehat{F}_b)/R$ and $S^2(\widehat{F}_b)=\sum_{r=1}^{R}(Y_r(\widehat{F}_b)-\bar{Y}_R(\widehat{F}_b))^2/(R-1)$.
   (c) For $r=1,2,\ldots,R$, generate $\varepsilon_{br}^*\sim N(0,S^2(\widehat{F}_b))$ and compute $Y_{br}^*=\bar{Y}_R(\widehat{F}_b)+\varepsilon_{br}^*$.
   (d) Let $Y_b^{**}=\sum_{r=1}^{R}Y_{br}^*/R$.
3. Find the empirical $\alpha/2$ and $1-\alpha/2$ quantiles of $\bar{Y}_R(\widehat{F}_b)$, $\widehat{q}_{\alpha/2}^{*}$ and $\widehat{q}_{1-\alpha/2}^{*}$, respectively.

4. Find the empirical $\alpha/2$ and $1-\alpha/2$ quantiles of Y_b^{**}, $\widehat{q}_{\alpha/2}^{**}$ and $\widehat{q}_{1-\alpha/2}^{**}$, respectively.
5. CI construction:
   (a) (**Basic CI**) Construct the basic hierarchical boostrap CI in (20).
   (b) (**Percentile CI**) Construct the percentile hierarchical boostrap CI in (21).

Note that, like the shrinkage procedure, the hierarchical approach for both the basic and the percentile bootstraps does not require additional simulation runs.

5 EMPIRICAL COMPARISON OF METHODS

In this section, we compare the empirical coverage probabilities of the CIs we discussed: basic bootstrap CI (4), percentile bootstrap CI (5), basic shrinkage CI (16), percentile shrinkage CI (17), basic hierarchical bootstrap CI (20), and percentile hierarchical bootstrap CI (21). We also present the results of asymptotic normality-based CI (8) and nominal t-interval that only accounts for intrinsic error for comparison.

We simulate an $M/M/1/10$ system whose inter-arrival and service time distributions have rates 0.8 and 1, respectively. We assume these distributions are unknown, but n observations from each distribution are available. The performance measure of interest is the steady-state expected time in system. For each simulation replication, we average 200 waiting times while deleting the initial 400 observations for warm-up.

We tested two different real-world sample sizes, $n = 100$ and $n = 1{,}000$. For the number of replications at each bootstrap sample and the number of bootstrap samples, we tested $R = 2$ and $R = 40$, and $B = 100$ and $B = 1{,}000$, respectively. Table 1 show these eight combinations of (n,R,B). All cases are repeated 1,000 times and the target coverage was set to $1-\alpha = 0.9$.

Figure 1 shows empirical coverage probability and the average width for each CI for each test case. Notice that the nominal t-interval shows poor coverage when R is large as it fails to account for input model uncertainty. The percentile bootstrap CI has a better coverage than the basic bootstrap CI in all cases. The former shows overcoverage when R is small (Case 1, 2, 5, and 6), while the latter shows undercoverage when n is small (Case 1, 2, 3, and 4). Notice that the basic and percentile bootstrap CIs are wider than other CIs (except for the t-intervals) as it does not compensate for the inflation due to finite R. The undercoverage of the basic bootstrap CIs, yet with long interval widths, hints that the positions of the CIs highly vary. When R and B are small, asymptotic normal CIs undercover on average as input uncertainty is poorly estimated by (10). This becomes more severe when n is large as input uncertainty becomes small relative to intrinsic error variance, which may make (10) negative. As mentioned earlier, we set $\widehat{V}^2 = 0$ in this case, i.e., the resulting asymptotic normal CI accounts only for the intrinsic noise. Similar observations can be made for the basic and percentile shrinkage CIs. Small B and R may result in $\widehat{c} = 1$ from (12), which makes all $\widehat{Y}_R(\widehat{F}_b) = \bar{\bar{Y}}_R$. Then, both basic and percentile shrinkage CIs become degenerate. For both $n = 100$ and $n = 1{,}000$, the percentile shrinkage CIs provide good coverage when B and R are large, however, the basic shrinkage CIs undercover when $n = 100$. Hierarchical boostrap CIs are the most robust under small R and B. In particular, when R is small, the hierarchical bootstrap inflates the CIs enough to provide the correct coverage, but not as much as the basic and percentile bootstrap CIs do.

6 CONCLUSION

Basic and percentile bootstrap CIs tend to overcover or exhibit high positioning variabilities because of intrinsic noise that inflates the CIs given any finite R. Asymptotic normality-based CIs, which is typically

Table 1: Combinations of (n,R,B) used for testing the CIs in Figure 1.

Case	1	2	3	4	5	6	7	8
n	100	100	100	100	1,000	1,000	1,000	1,000
R	2	2	40	40	2	2	40	40
B	100	1,000	100	1,000	100	1,000	100	1,000

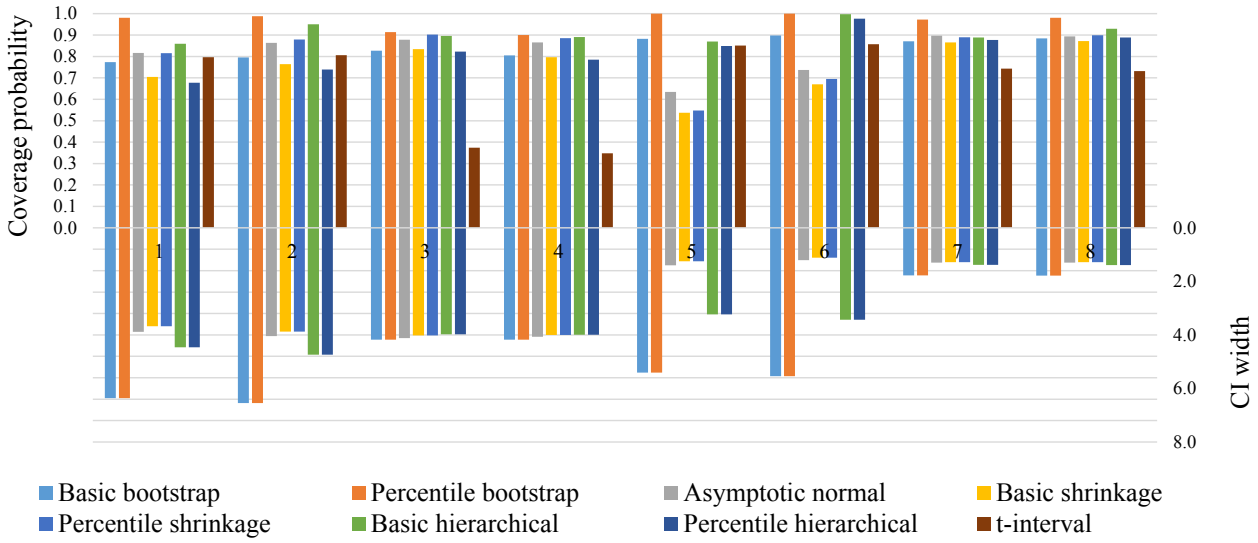

Figure 1: Empirical coverage probability and the average width for the eight CIs in comparison computed from 1,000 repeats. The target coverage is $1-\alpha = 0.9$.

used in the literature to correct for the overcoverage, relies on that $\eta(\widehat{F})$ becomes asymptotically normally distributed when real-world sample size n increases. Although it appears to work well for the example in Section 5, bootstrap-based CIs tend to be more robust to non-normality of $\eta(\widehat{F})$. We propose four new bootstrap-based CIs for $\eta(F^c)$. The percentile shrinkage CI perform well when n, R, and B are large while the basic hierarchical bootstrap CI show good coverage across all combinations of (n,R,B).

For shrinkage CIs, we assumed homoscedasticity of intrinsic error for any $\widehat{F}_b$. Relaxing this assumption could increase their coverage probability. From the experiment, the basic hierarchical bootstrap CI improves the basic bootstrap CI even for small n, while the percentile hierarchical bootstrap does not. Further analysis is required to understand their asymptotic properties.

REFERENCES

Ankenman, B. E., and B. L. Nelson. 2012. "A Quick Assessment of Input Uncertainty". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque et al., 1–10. Piscataway, New Jersey: IEEE.

Barton, R. R., and L. W. Schruben. 1993. "Uniform and Bootstrap Resampling of Empirical Distributions". In *Proceedings of the 1993 Winter Simulation Conference*, edited by G. W. Evans et al., 503–508. Piscataway, New Jersey: IEEE.

Barton, R. R., and L. W. Schruben. 2001. "Resampling Methods for Input Modeling". In *Proceedings of the 2001 Winter Simulation Conference*, edited by B. A. Peters et al., 372–378. Piscataway, New Jersey: IEEE.

Barton, R. R. 2007. "Presenting a More Complete Characterization of Uncertainty: Can it be done?". In *Proceedings of the 2007 INFORMS Simulation Society Research Workshop*, edited by S. C. et al.

Barton, R. R., B. L. Nelson, and W. Xie. 2014. "Quantifying input uncertainty via simulation confidence intervals". *INFORMS Journal on Computing* 26(1):74–87.

Cheng, R. C. H. 1994. "Selecting Input Models". In *Proceedings of the 1994 Winter Simulation Conference*, edited by J. D. Tew et al., 184–191. Piscataway, New Jersey: IEEE.

Cheng, R. C. H., and W. Holland. 1997. "Sensitivity of Computer Simulation Experiments to Errors in Input Data". *Journal of Statistical Computation and Simulation* 57(1-4):219–241.

Cheng, R. C. H., and W. Holland. 2004. "Calculation of Confidence Intervals for Simulation Output". *ACM Transactions on Modeling and Computer Simulation* 14:344–362.

Davison, A. C., and D. V. Hinkley. 1997. *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Morgan, L. E., B. L. Nelson, A. C. Titman, and D. J. Worthington. 2017. "Detecting Bias due to Input Modelling in Computer Simulation". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan et al., 1974–1985. Piscataway, New Jersey: IEEE.

Song, E., B. L. Nelson, and C. D. Pegden. 2014. "Advanced Tutorial: Input Uncertainty Quantification". In *Proceedings of the 2017 Winter Simulation Conference*, edited by A. Tolk et al., 162–176. Piscataway, New Jersey: IEEE.

Song, E., and B. L. Nelson. 2015. "Quickly Assessing Contributions to Input Uncertainty". *IIE Transactions* 47(9):893–909.

ACKNOWLEDGEMENT

Computations for this research were performed on the Penn State University's Institute for CyberScience Advanced CyberInfrastructure (ICS-ACI). We gratefully acknowledge support from the National Science Foundation under grants CMMI-1542020 and CAREER CMMI-1653339/1834710.

AUTHOR BIOGRAPHIES

RUSSELL R. BARTON is Distinguished Professor of Supply Chain and Information systems and Professor of Industrial Engineering at the Pennsylvania State University. He received a B.S. degree in electrical engineering from Princeton University and M.S. and Ph.D. degrees in operations research from Cornell University. He serves as INFORMS Vice President for Sections and Societies. He is a fellow of IIE and a Certified Analytics Professional. His research interests include applications of statistical and simulation methods to system design and to product design, manufacturing and delivery. His email address is rbarton@psu.edu.

HENRY LAM is an Associate Professor in the Department of Industrial Engineering and Operations Research at Columbia University. He received his Ph.D. degree in statistics at Harvard University in 2011, and was on the faculty of Boston University and the University of Michigan before joining Columbia in 2017. His research focuses on Monte Carlo simulation, risk and uncertainty quantification, and stochastic optimization. His email address is khl2114@columbia.edu.

EUNHYE SONG is Harold and Inge Marcus Early Career Assistant Professor in the Department of Industrial and Manufacturing Engineering at the Penn State University. Her research interests include simulation design of experiments, simulation uncertainty and risk quantification, optimization via simulation under input model risk and large-scale discrete optimization via simulation. Her email address is eus358@psu.edu.