Proceedings of the 2012 Winter Simulation Conference
C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, eds.

A QUICK ASSESSMENT OF INPUT UNCERTAINTY

Bruce E. Ankenman
Barry L. Nelson

Department of Industrial Engineering & Management Sciences
Northwestern University
Evanston, IL 60208 USA

ABSTRACT

"Input uncertainty" refers to the frequently unrecognized, and rarely quantified, impact of using simulation input distributions that are estimated or "fit" to a finite sample of real-world data. In this paper we present a relatively simple method for obtaining a quick assessment of the overall effect of input uncertainty on simulation output, and a somewhat more involved follow-up analysis that can identify the largest sources of input uncertainty. Numerical illustrations of both methods are provided.

1 INTRODUCTION

Input models are the fully specified probability models that drive a stochastic simulation. Interarrival-time and service-time distributions in queueing simulations; component time-to-failure distributions in reliability simulations; distributions for the values of underlying assets in financial simulations; and bed occupancy-time distributions in hospital simulations are examples of input models. Classical simulation output analysis is conditional on the input models being correct; that is, it assumes that the input models accurately represent the stochastic phenomena in the real-world system. However, if these input models are functions of real-world data then they will not perfectly represent the true characteristics, and this in turn should imply additional uncertainty about the simulation results; however, this additional uncertainty is not captured by conditional analysis.

Quantification of *input uncertainty* is an area of active research and a number of sophisticated and rigorously justified methods have been proposed; see for instance Barton et al. (2011), Cheng and Holland (1998, 2004), Chick (2001), and Zouaoui and Wilson (2003, 2004). Our goal is to provide a method for assessing the impact of input uncertainty on simulation-based performance estimates that requires relatively little additional simulation effort and that is simple enough to implement manually if necessary. We also provide a follow-up analysis that, with a bit more effort, can identify which input distributions are the largest contributors to input uncertainty.

Section 2 describes our quick assessment technique and illustrates its use on a simple queueing model. Section 3 presents the follow-up analysis that uncovers which distributions are responsible for most of the uncertainty. Section 4 applies both steps to a realistic manufacturing simulation.

2 QUICK OVERALL ASSESSMENT

To simplify the presentation, suppose initially that the simulation has a single input distribution on which we have an i.i.d. sample of real-world data $X_1, X_2, \ldots, X_m$. Denote the true but unknown input distribution by F_X, and our estimate of it by $\widehat{F}$ which is a function of the real-world data. This fitted distribution could be the empirical distribution or a parametric distribution with estimated parameters (although the empirical distribution will play a key role in our method).

978-1-4673-4781-5/12/$31.00 ©2012 IEEE

Here we only consider simulations for which we make i.i.d. replications. Represent the output of the simulation on replication j, implemented using estimated input distribution $\widehat{F}$, as

$$Y_j(\widehat{F}) = \mu(\widehat{F}) + \varepsilon_j$$

where $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are i.i.d. with mean 0 and variance σ^2 representing the natural variability from replication to replication in the simulation. The mean term, $\mu(\widehat{F})$, depends on what input model we actually used in the simulation. Our goal is to estimate $\mu \equiv \mu(F_X)$ which we estimate by

$$\bar{Y}(\widehat{F}) = \frac{1}{n}\sum_{j=1}^{n} Y_j(\widehat{F}). \tag{1}$$

This estimator, which is the one that will be used for decision making, has standard error $\sigma/\sqrt{n}$ *conditional on the input distribution* $\widehat{F}$. Our assessment of input uncertainty is in addition to this standard analysis.

As a thought experiment, suppose that rather than a single sample of real-world data, we had b independent samples $X_{i1}, X_{i2}, \ldots, X_{im}, i = 1, 2, \ldots, b$, and from the ith sample we fit an input model $\widehat{F}_i$. We could then consider simulating r replications using *each* input model, leading to the output

$$Y_j(\widehat{F}_i) = \mu(\widehat{F}_i) + \varepsilon_{ij} \tag{2}$$

for $i = 1, 2, \ldots, b$ and $j = 1, 2, \ldots, r$. For reasons that will become clear later we do not require $r = n$.

Model (2) is known as a *random-effects model* (see, for instance, Dean and Voss (1999), or Montgomery (2009)), because the mean term $\mu(\widehat{F}_i)$ is random, in this case depending on the particular sample $X_{i1}, X_{i2}, \ldots, X_{im}$ that we happened to use to estimate F_X. This model has also been suggested by Freimer and Schruben (2002), Sun et al. (2011) and Zouaoui and Wilson (2003). One way to characterize input uncertainty is $\sigma_I^2 = \mathrm{Var}[\mu(\widehat{F}_i)]$, the variability due to the possibility of having observed a different real-world sample of data from F_X. A simplifying approximation we will make is that σ^2, the variance of the simulation output, does not depend on which $\widehat{F}_i$ we have; this will rarely be precisely true, but will be acceptable to get an approximation of the effect of input uncertainty.

We will focus on the ratio $\gamma = \sqrt{n}\sigma_I/\sigma$ which is a measure of how large input uncertainty is relative to the standard error of $\bar{Y}(\widehat{F})$. In particular, we desire a confidence interval (CI) for γ. Under the additional assumptions that $\mu(\widehat{F}_i) \sim \mathrm{N}(\mu', \sigma_I^2)$ and $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$ there is a standard $(1-\alpha)100\%$ CI for σ_I^2/σ^2 (see for instance Dean and Voss 1999 §17.3.5):

$$\frac{1}{r}\left[\frac{\mathrm{MST}}{\mathrm{MSE}\cdot F_{b-1,b(r-1),1-\alpha/2}} - 1\right] \le \frac{\sigma_I^2}{\sigma^2} \le \frac{1}{r}\left[\frac{\mathrm{MST}}{\mathrm{MSE}\cdot F_{b-1,b(r-1),\alpha/2}} - 1\right] \tag{3}$$

where $F_{\nu_1,\nu_2,\beta}$ is the β-quantile of the F distribution with (ν_1, ν_2) degrees of freedom, and

$$\begin{aligned}
\mathrm{MSE} &= \frac{1}{b(r-1)}\sum_{i=1}^{b}\sum_{j=1}^{r}\left(Y_j(\widehat{F}_i) - \bar{Y}(\widehat{F}_i)\right)^2 \\
\mathrm{MST} &= \frac{r}{b-1}\sum_{i=1}^{b}\left(\bar{Y}(\widehat{F}_i) - \bar{Y}\right)^2
\end{aligned}$$

with

$$\bar{Y} = \frac{1}{br}\sum_{i=1}^{b}\sum_{j=1}^{r} Y_j(\widehat{F}_i).$$

This CI for σ_I^2/σ^2 can be adapted in the obvious way to obtain a CI for γ: multiply through by n and take the square root. Dean and Voss (1999, §17.4) also provide an unbiased estimator for σ_I^2/σ^2, which can be adapted to give the following point estimator of γ:

$$\widehat{\gamma} = \sqrt{\frac{n}{r}\left[\left(\frac{b(r-1)-2}{b(r-1)}\right)\frac{\text{MST}}{\text{MSE}} - 1\right]}.$$

Notice that we do not assume $\mu' = \mu$ because even if $\widehat{F}$ is an unbiased estimator for F_X this does not guarantee $\text{E}[\mu(\widehat{F})] = \mu(F_X)$.

The problem with this proposal is the need for more than one real-world sample. Therefore, we use the statistical technique of bootstrapping to imitate the effect of having multiple real-world samples, even though we only have one. Again let $\{X_1, X_2, \ldots, X_m\}$ be the one real-world sample of data. Instead of multiple real-world samples, we generate b bootstrap samples, each of size m, by sampling m times with replacement from $\{X_1, X_2, \ldots, X_m\}$. As is common in the bootstrapping literature, we denote the ith bootstrap sample by $X_{i1}^\star, X_{i2}^\star, \ldots, X_{im}^\star$. We then fit an input distribution $\widehat{F}_i^\star$ to this bootstrap sample and let $\widehat{F}_i^\star$ can stand in for $\widehat{F}_i$. In a problem with more than one input model we do bootstrapping for each distribution separately. Here is the procedure in algorithm form:

Algorithm Quick

1. Given real-world data $\{X_1, X_2, \ldots, X_m\}$, do the following:
2. For i from 1 to b
   (a) Generate bootstrap sample $X_{i1}^\star, X_{i2}^\star, \ldots, X_{im}^\star$ by sampling m times with replacement from $\{X_1, X_2, \ldots, X_m\}$.
   (b) Fit $\widehat{F}_i^\star$ to $X_{i1}^\star, X_{i2}^\star, \ldots, X_{im}^\star$.
   (If there is more than one input model, do Steps 2a and 2b for each one.)
   (c) Simulate r replications $Y_j(\widehat{F}_i^\star)$, $j = 1, 2, \ldots, r$ using input model(s) $\widehat{F}_i^\star$.
3. Report the point estimate and CI for $\gamma = \sqrt{n}\sigma_I/\sigma$.

The reason that we focus on the parameter $\gamma = \sqrt{n}\sigma_I/\sigma$ is that the point estimator of μ is $\bar{Y}(\widehat{F})$ (see (1)) and it is based on n simulation replications. However, the experiment to assess input uncertainty in Algorithm Quick need not be tied to n simulation replications. Instead, we can choose how many replications, say N, to spend on the input uncertainty experiment, and try to choose $br = N$ optimally. In the Appendix we show that the optimal or near-optimal choice of b to minimize the width of the CI for γ given a fixed budget N is $b^\star = 10$ bootstraps, meaning we use $r = N/10$ replications on each bootstrap experiment.

To illustrate the approach we start with a mathematically tractable example from Barton et al. (2011). For an $M/M/\infty$ queue with arrival rate λ and mean service time τ the steady-state number of customers in the system is Poisson with mean $\mu = \lambda\tau$. Suppose λ and τ are not known but we observe m i.i.d. interarrival times, $A_1, A_2, \ldots, A_m$, and m i.i.d. service times $X_1, X_2, \ldots, X_m$ from the real world and use them to fit input models. Specifically, we estimate λ by

$$\widehat{\lambda} = \left(\frac{1}{m}\sum_{i=1}^{m} A_i\right)^{-1}$$

and τ by

$$\widehat{\tau} = \frac{1}{m}\sum_{i=1}^{m} X_i.$$

We assume that we are confident that the distributions of A and X are exponential, but need to estimate the values of the parameters. Thus, our input model $\widehat{F}$ consists of two exponential distributions with parameters $\widehat{\lambda}$ and $1/\widehat{\tau}$. For the purpose of this illustration, we have $m = 100$ real-world observations of both the interarrival and service times (which we obtained by generating them from exponential distributions with $\lambda = 5$ and $\tau = 1$). Bootstrapping therefore consists of resampling this data and re-estimating λ and τ.

To allow for a mathematical analysis, suppose that when we simulate this queue we record a single observation, the number in the queue in steady state, on each of n replications. Then we can show that

$$\mathrm{Var}[\bar{Y}(\widehat{F})] \approx \frac{2(\lambda\tau)^2}{m} + \frac{\lambda\tau}{n} = \sigma_I^2 + \frac{\sigma^2}{n}.$$

Thus,

$$\gamma = \sqrt{\frac{n\lambda\tau}{m}}.$$

Notice that γ can only be decreased by having more real-world data m.

Suppose that our conditional experiment design was to make $n = 200$ replications, and we are willing to spend another $N = 200$ replications to assess input uncertainty. Therefore, we apply Algorithm Quick with $b^\star = 10$ bootstrap experiments, each of $r = 20$ replications. This experiment yields $\widehat{\gamma} = 3.8$ with 95% CI of $[1.1, 8.5]$. Thus, input uncertainty is from 1 to 8 times as large as estimator uncertainty. Notice that the true value of $\gamma = \sqrt{200 \cdot 5 \cdot 1/100} \approx 3.2$.

In this example we had parametric input distributions (exponential) with unknown parameters, so the bootstrap resampled data produced different possible parameter values. Other approaches are possible:

- Rather than generating new parameter values by resampling the real-world data and reestimating the parameters, we could resample the parameters directly from their sampling distributions, if known. For instance, maximum likelihood estimators are known to be asymptotically normal, and this could provide a distribution for resampling parameter values.
- Rather than refitting distributions at all, we could drive the simulations by the empirical distribution; i.e.,

  $$\widehat{F}_i^\star(x) = \frac{1}{m}\sum_{\ell=1}^{m} I(X_{i\ell}^\star \le x).$$

  This avoids the potentially slow process of refitting parametric distributions.

3 FOLLOW-UP ANALYSIS

If the quick analysis of input uncertainty above reveals that it is a concern, as it often will, then it may be worthwhile to identify which input models contribute "more than their fair share" to the overall input uncertainty. In this section we present a follow-up analysis to do that.

Suppose that there are L independent, univariate input distributions, $F_{X_1}, F_{X_2}, \ldots, F_{X_L}$, and for the ℓth input distribution we have m_ℓ i.i.d. real-world observations $X_{\ell,1}, X_{\ell,2}, \ldots, X_{\ell,m_\ell}$. For instance, in the $M/M/\infty$ example there were $L = 2$ input distributions (both believed to be exponential but with unknown parameters) and we had $m_1 = m_2 = 100$ real-world observations from each.

Let $\widehat{F}_\ell, \ell = 1, 2, \ldots, L$ denote the input distribution fitted directly to these data. In other words, these are the distributions that would be used in the standard conditional analysis; therefore, $\widehat{F} = \{\widehat{F}_1, \widehat{F}_2, \ldots, \widehat{F}_L\}$. Further, let $\widehat{F}_{\ell,i}^\star, i = 1, 2, \ldots, b$ denote the b bootstrap distributions created for the ℓth input model in the overall analysis above. We assume that these distributions are retained, but they could also be regenerated as needed if that is more convenient.

To facilitate our follow-up analysis we make the following additional modeling approximation:

$$\mathrm{Var}[Y_j(\widehat{F})] = \sigma_I^2 + \sigma^2 = \sum_{\ell=1}^{L} \sigma_\ell^2 + \sigma^2 \tag{4}$$

where σ_ℓ^2 is the input uncertainty variance contributed by the ℓth input model. We consider a "fair share" of the uncertainty contributed by the ℓth input model to be

$$\frac{\gamma^2}{L} = \frac{n\sigma_I^2}{L\sigma^2}$$

and we want to identify which input distributions contribute more than that. We will do so using a method that is motivated by a factor screening approach called sequential bifurcation (see, for instance, Wan, Ankenman and Nelson 2006). The idea is to efficiently eliminate *groups* of input distributions that contribute little to input uncertainty so that we can focus on ones that contribute the most.

Here is how we test the contribution of a group of distributions: Let $\mathcal{G} \subset \{1,2,\ldots,L\}$ be a subset of consecutively numbered factors; e.g., if $L = 12$ distributions then we could have $\mathcal{G} = \{4,5,6,7\}$. We are interested in whether

$$\gamma^2(\mathcal{G}) = \frac{n\sum_{\ell\in\mathcal{G}}\sigma_\ell^2}{\sigma^2}$$

is large relative to $|\mathcal{G}|\gamma^2/L$; if it is then the group contains at least one input distribution that is contributing more than its fair share. We do this by exploiting three ideas:

- We have an estimate $\widehat{\gamma}^2$ of γ^2 from the overall assessment, so we use that to define the standard.
- Under model (4) we can construct a CI for $\gamma^2(\mathcal{G})$ by running b bootstrap simulation experiments where we use the input distributions $\widehat{F}^\star_{\ell,i}, i = 1,2,\ldots,b$ for $\ell \in \mathcal{G}$ but use $\widehat{F}_\ell$ for $\ell \notin \mathcal{G}$. This is like treating the input distributions that are not in $\mathcal{G}$ as being known and fixed, and using bootstrapping to assess the overall input uncertainty due to the distributions in $\mathcal{G}$.
- If the CI for $\gamma^2(\mathcal{G})$ is completely to the right of $|\mathcal{G}|\gamma^2/L$, then we can claim with high confidence that some distributions in the group are big contributors to input uncertainty. Therefore, we split the group and then evaluate the subgroups in the same way.

We illustrate the quick assessment and follow-up analysis for a realistic example in the next section.

4 APPLICATION

The example to which we will apply our method has been used in several papers by Jack Kleijnen, Fredrik Persson and various co-authors (see e.g., Kleijnen, Bettonvil and Persson (2003) and Persson and Olhager (2002)). The simulation model and a student version of the Taylor II software for running the model were generously provided by Fredrik Persson.

The simulation is a model of a manufacturing system that produced mobile phones for the Ericsson Company. The manufacturing system is described in detail in Persson and Olhager (2002) and we used the model that they described as the "old" structure of the production line. For our purposes, it is sufficient to understand that the production line consists of conveyors that lead the product through a series of machines that perform functions common to the production of circuit boards and consumer devices, such as soldering, testing and assembly. There are many outputs of the simulation model including overall lead time, quality and cost. In addition, there are outputs of quality, lead time and utilization for each machine. We begin by focusing on the same response that was used in Kleijnen, Bettonvil and Persson (2003) which is the total weekly cost of production.

Each of the machines in the simulation has a random processing time that is modeled by a lognormal distribution with two specified parameters: mean and standard deviation. We presume that the mean and standard deviation that are used in the simulation model are based on the observed mean and standard deviation from some actual machine that is similar to the one being modeled in the simulation. Of course, the observed mean and standard deviation will not be the true mean and standard deviation of that machine. The goal of our method is to quantify the amount of input uncertainty, which is the variability that is caused by misspecification of these parameter values. Additionally, if there is substantial input uncertainty, then

Table 1: The baseline values of the processing-time parameters for the 23 machines under study.

Machine	Mean	StDev	Machine	Mean	StDev
Mach 3	2.5	1	Mach 74	10	1
Mach 7	0.9	1	Mach 143	14	1
Mach 16	1.4	1	Mach 145	23	1
Mach 29	10	1	Mach 146	14	1
Mach 33	4	1	Mach 147	27	1
Mach 37	9	1	Mach 110	2.5	1
Mach 68	10	1	Mach 113	2.8	1
Mach 69	10	1	Mach 114	9.5	1
Mach 70	10	1	Mach 115	9.5	1
Mach 71	10	1	Mach 139	14.2	1
Mach 72	10	1	Mach 140	23.7	1
Mach 73	10	1			

we would like to deduce which machines are primarily responsible for that uncertainty. This knowledge would allow us to collect additional information on the parameter values for those machines to get more precise parameter estimates and thereby reduce the effect of input uncertainty.

We have chosen 23 machines from the simulation model to study. Table 1 displays the 23 machines and the parameter settings for each of their lognormal processing-time distributions.

We do not actually have the original real-world data, so to apply our method as presented in Section 2, we started by generating a set of 500 random processing times from each of the lognormal processing-time distributions in Table 1. We treated these 23 sets of 500 observations as the "real world" sample from each machine. We then followed the Quick Algorithm in Section 2 to create $b = 10$ bootstrap samples of size $m = 500$ for each of the 23 machines. For each bootstrap sample, we calculated the sample mean and sample standard deviation, which led to 10 sets of empirically estimated parameter values for each machine.

For example, the parameters from the 10 bootstrap samples from Machine 3 are shown below:

Bootstrap #	Mean	StDev
1	2.482124	1.048387
2	2.491728	1.023643
3	2.492922	1.024837
4	2.388243	0.897349
5	2.414130	0.974626
6	2.433462	1.025614
7	2.429213	0.923263
8	2.426516	0.947910
9	2.434552	1.001201
10	2.436321	0.949222

The variability of parameter values represents the input variation from the real-world sample taken for Machine 3. To quantify the overall input uncertainty, we ran $r = 10$ replications of the simulation using the empirical parameter values from the first bootstrap sample of each of the 23 machines and then repeated this for each of the other nine bootstrap samples until we had 100 total observations. We call these 100 runs of the simulation Data Set 1.

Using Data Set 1 and following the derivations in Section 2, we found that for a simulation experiment specifying $n = 10$ replications and output total weekly cost, the ratio of input uncertainty to standard error of the simulation has a point estimate of $\widehat{\gamma} = 0.84$ and a 95% confidence interval of (0, 2.23). The lower

Table 2: The sequential bifurcation data sets.

	Data Set 1	Data Set 2	Data Set 3	Data Set 4	Data Set 5
Mach 3	**Varying**	**Varying**	Not Varying	Not Varying	Not Varying
Mach 7	**Varying**	**Varying**	Not Varying	Not Varying	**Varying**
Mach 16	**Varying**	**Varying**	Not Varying	Not Varying	Not Varying
Mach 29	**Varying**	**Varying**	Not Varying	Not Varying	Not Varying
Mach 33	**Varying**	**Varying**	Not Varying	Not Varying	Not Varying
Mach 37	**Varying**	**Varying**	Not Varying	Not Varying	Not Varying
Mach 68	**Varying**	**Varying**	Not Varying	Not Varying	Not Varying
Mach 69	**Varying**	**Varying**	Not Varying	Not Varying	Not Varying
Mach 70	**Varying**	**Varying**	Not Varying	Not Varying	Not Varying
Mach 71	**Varying**	**Varying**	Not Varying	Not Varying	Not Varying
Mach 72	**Varying**	**Varying**	Not Varying	Not Varying	Not Varying
Mach 73	**Varying**	**Varying**	Not Varying	Not Varying	Not Varying
Mach 74	**Varying**	**Varying**	Not Varying	Not Varying	Not Varying
Mach 143	**Varying**	Not Varying	**Varying**	Not Varying	Not Varying
Mach 145	**Varying**	Not Varying	**Varying**	Not Varying	Not Varying
Mach 146	**Varying**	Not Varying	**Varying**	Not Varying	Not Varying
Mach 147	**Varying**	Not Varying	**Varying**	Not Varying	Not Varying
Mach 110	**Varying**	Not Varying	**Varying**	Not Varying	Not Varying
Mach 113	**Varying**	Not Varying	**Varying**	Not Varying	Not Varying
Mach 114	**Varying**	Not Varying	Not Varying	**Varying**	Not Varying
Mach 115	**Varying**	Not Varying	Not Varying	**Varying**	Not Varying
Mach 139	**Varying**	Not Varying	Not Varying	**Varying**	Not Varying
Mach 140	**Varying**	Not Varying	Not Varying	**Varying**	Not Varying

confidence level was actually negative. Since that is impossible, we cut off the confidence interval at 0. This confidence interval indicates that, with 95% confidence, the input uncertainty is no more than 2.23 times the size of the standard error and we estimate it to be less than the standard error. In our experience, this indicates that input uncertainty for this particular response is not a serious problem and thus a follow-up analysis on this response does not seem necessary. Our conclusion is that the amount of data collected (500 processing times for each machine) appears to be adequate for reducing the input uncertainty to a manageable level for this response, mean total weekly cost.

To demonstrate the follow-up analysis discussed in Section 3, we will focus on a different output from the simulation, namely the utilization of Machine 7. For this response, we know that the input variation of the processing time of Machine 7 is the most likely cause of any input uncertainty. However, we will use the follow-up analysis to confirm this and thereby demonstrate the effectiveness of the method.

We begin by quantifying the total amount of input uncertainty using Data Set1, but now using the Utilization of Machine 7 as the response. The ratio of input uncertainty to standard error of the simulation has a point estimate of 11.32 and a 95% confidence interval of (7.58, 21.21). This shows that, with 95% confidence, the input uncertainty is 7 to 21 times as large as the simulation standard error for this response. Clearly if this response is important, the input uncertainty is of great concern. The next step would be to implement a sequential bifurcation procedure to determine which of the 23 machines is contributing to the input uncertainty. Table 2 shows the collection of four more data sets of 100 observations which vary the parameters of selected machine processing times, but leave others constant at the value estimated from the real world-data set. Data Sets 2, 3 and 4 represent the sequential bifurcation process. Our point estimate

Table 3: The point estimates and 95% confidence intervals for the ratio of input uncertainty to standard error of the simulation for each of the sequential bifurcation data sets. "N/A" indicates that the procedure returned a negative value which can reasonably be interpreted as 0.

Data Set (Varying)	Lower Bound	Point Estimate	Upper Bound
1 (All)	7.58	11.32	21.21
2 (Mach 3-74)	6.93	10.36	19.42
3 (Mach 143-113)	0.65	1.47	3.16
4 (Mach 114-140)	N/A	N/A	0.65
5 (Mach 7)	6.95	10.39	19.47

for γ^2, the squared ratio of input uncertainty variance to simulation variance, is $(11.32)^2 = 128.1$ from Data Set 1.

The point estimates and confidence intervals in Table 3 from Data Set 4 indicate that input variability for the last four machines (114, 115, 139, and 140) have no significant effect on the input uncertainty for utilization at Machine 7 since the upper limit of the confidence interval for the variance ratio is lower than 4/23 of the point estimator: $0.42 = 0.65^2 < (4/23)(11.34)^2 = 22.36$. Similarly, machines 143–113 have no significant effect on the input uncertainty as $9.98 = 3.16^2 < (6/23)(11.34)^2 = 33.55$. The results from Data Set 2 indicate that Machines 3–74 have the most significant effect. Further sequential bifurcation could be done, but in this case we know that machine 7 is the most likely significant factor so we will skip to the last stage of the sequential bifurcation where only the parameters of the processing time of Machine 7 are varying. The results for Data Set 5 show that indeed this machine contributes essentially all of the input uncertainty and confirms the method for isolating the important contributors.

5 CONCLUSIONS

We have presented and illustrated a simple method for quantifying the impact of input uncertainty relative to simulation uncertainty as measured by the standard error of the simulation point estimator; that is, we estimate $\gamma = \sqrt{n}\sigma_I/\sigma$. We also provided a follow-up analysis to help identify which input distributions contribute the most uncertainty. Our focus here was on independent, univariate input distributions, so extension to multivariate input distributions is the next logical step.

One technical problem with our approach is that even though our point estimator of γ^2 is unbiased, and our CI is valid under normal-distribution assumptions, it is possible for the point estimator and CI end points to be negative even though $\gamma^2 < 0$ is impossible. This problem is not unique to our setting; it arises in all random-effects analysis. We believe that it is reasonable, if not totally satisfying, to interpret negative values as 0 as we have done here.

APPENDIX

Here we address the choice of number of bootstrap samples, b, to use for forming the CI.

Multiplying the CI (3) through by n gives a CI for γ^2 of

$$\frac{n}{r}\left[\frac{\text{MST}}{\text{MSE}\cdot F_{b-1,b(r-1),1-\alpha/2}}-1\right] \le \gamma^2 \le \frac{n}{r}\left[\frac{\text{MST}}{\text{MSE}\cdot F_{b-1,b(r-1),\alpha/2}}-1\right].$$

Given a fixed budget of N replications, our goal is to choose b (and thus $r = N/b$) to minimize the expected length of this CI. Using simple algebra and the relationship $br = N$ the expected length is

$$\text{E}\left(\frac{\text{MST}}{\text{MSE}}\right)\left(\frac{n}{N}\right) b\left(\frac{1}{F_{b-1,N-b,1-\alpha/2}}-\frac{1}{F_{b-1,N-b,\alpha/2}}\right).$$

Now n/N is fixed, and if we treat E(MST/MSE) as approximately constant as a function of b, then the expected length only depends on

$$b\left(\frac{1}{F_{b-1,N-b,1-\alpha/2}}-\frac{1}{F_{b-1,N-b,\alpha/2}}\right). \tag{5}$$

As $N\to\infty$ this becomes

$$b\left(\frac{b-1}{\chi^2_{b-1,1-\alpha/2}}-\frac{b-1}{\chi^2_{b-1,\alpha/2}}\right)$$

which (for fixed α) is only a function of b. For the standard values of α, this expression is minimized at $b=11$, and the function is flat near 11. Evaluating (5) for specific values of N, we find that the optimal b decreases slightly when N is as small as 60 to $b^\star=10$, which is the value we recommend.

REFERENCES

Barton, R. R., B. L. Nelson and W. Xie. 2011. Quantifying input uncertainty via simulation confidence intervals. *INFORMS Journal on Computing*, forthcoming.

Cheng, R. C. H. and W. Holland. 1998. Two-point methods for assessing variability in simulation output. *Journal of Statistical Computation and Simulation* **60**: 183-175.

Cheng, R. C. H. and W. Holland. 2004. Calculation of confidence intervals for simulation output. *ACM Transactions on Modeling and Computer Simulation* **14**: 344-172.

Chick, S. E. 2001. Input distribution selection for simulation experiments: Accounting for input uncertainty. *Operations Research* **49**: 744-178.

Dean, A. and D. Voss. 1999. *Design and Analysis of Experiments.* New York: Springer.

Freimer, M. and L. W. Schruben. 2002. Collecting data and estimating parameters for input distributions. In *Proceedings of the 2002 Winter Simulation Conference*, edited by E. Yücesan, C. Chen, J. L. Snowdon, and J. M. Charnes, 392-179. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kleijnen, J. P. C., B. Bettonvil and F. Persson. 2003. Finding the important factors in large discrete-event simulation: sequential bifurcation and its applications, In A. M. Dean and S. M. Lewis (Eds.), *Screening*. New York: Springer-Verlag.

Montgomery, D. C. 2009. *Design and Analysis of Experiments.* 7th ed. New York: John Wiley.

Persson, F. and J. Olhager. 2002. Performance simulation of supply chain designs. *International Journal of Production Economics* **77**: 231–245.

Sun, Y., D. W. Apley and J. Staum. 2011. Efficient Nested Simulation for Estimating the Variance of a Conditional Expectation. *Operations Research* **59**: 998–1007.

Wan, H., B. E. Ankenman and B. L. Nelson. 2006. Controlled sequential bifurcation: A new factor-screening method for discrete-event simulation. *Operations Research* **54**: 743–755.

Zouaoui, F. and J. R. Wilson. 2003. Accounting for parameter uncertainty in simulation input modeling. *IIE Transactions* **35**: 781-172.

Zouaoui, F. and J. R. Wilson. 2004. Accounting for input-model and input-parameter uncertainties in simulation. *IIE Transactions* **36**: 1135-1751.

ACKNOWLEDGMENTS

This research was partially supported by National Science Foundation Grant CMMI-1068473.

AUTHOR BIOGRAPHIES

BRUCE E. ANKENMAN is a Charles Deering McCormick Professor of Teaching Excellence and an Associate Professor in the Department of Industrial Engineering and Management Sciences at North-

western University's McCormick School of Engineering and Applied Sciences. He received a BS in Electrical Engineering from Case Western Reserve University and an MS and PhD in Industrial Engineering from the University of Wisconsin-Madison. His research interests primarily deal with the design and analysis of experiments that are used to build models for physical systems or metamodels for simulated systems. Professor Ankenman is the co-director of the Segal Design Institute. For more info see www.iems.northwestern.edu/~bea.

BARRY L. NELSON is the Walter P. Murphy Professor and Chair of the Department of Industrial Engineering and Management Sciences at Northwestern University. He is a Fellow of INFORMS and IIE. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems. His e-mail and web addresses are nelsonb@northwestern.edu and www.iems.northwestern.edu/~nelsonb.