Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

ESTIMATING SENSITIVITY TO INPUT MODEL VARIANCE

Wendy Xi Jiang
Barry L. Nelson

Dept. of Industrial Engr. & Mgmt. Sci.
Northwestern University
Evanston, IL 60208-3119, USA

L. Jeff Hong

School of Management
School of Data Science
Fudan University
Shanghai 200433, P. R. CHINA

ABSTRACT

Simple question: How sensitive is your simulation output to the *variance* of your simulation input models? Unfortunately, the answer is not simple because the variance of many standard parametric input distributions can achieve the *same* change in multiple ways as a function of the parameters. In this paper we propose a family of output-mean-with-respect-to-input-variance sensitivity measures and identify two particularly useful members of it. A further benefit of this family is that there is a straightforward estimator of any member with no additional simulation effort beyond the nominal experiment. A numerical example is provided to illustrate the method and interpretation of results.

1 INTRODUCTION

A computer model maps its inputs into outputs via a collection of rules and algorithms that mimic the features of the target system. The output of a model can be regarded as a function of the inputs, i.e., $Y = g(\boldsymbol{\theta})$, where each element θ_i of $\boldsymbol{\theta}$ could be a constant, such as the planned release rate of wafers in a semiconductor manufacturing line or the thermal property of a material; or a quantity that is inherently variable such as the daily air temperature or the time that an agent in a call center spends on a call. Computer models are never perfect representations of complex real-life systems or processes, and this includes the inputs that are uncertain due to lack of information, errors of measurement or estimation error due to sampling (Saltelli et al. 2000). Uncertainties in the inputs will imply uncertainty in the outputs. Broadly speaking, sensitivity analysis (SA) studies how the model output responses are affected by the inputs so as to better understand system performance, to quantify risk, or to indicate where input change or management may be desirable. Depending on the type of input and the goals of the analysis, SA methods can be grouped into two categories, global SA and local SA, and further subdivisions within each. This paper deals with a problem in local SA.

Global SA often addresses the case when inputs would be deterministic, if known. To study the effect of uncertainty in the input factors due to lack of information, global SA may impose a distribution, $\Theta_i \sim F_i$, on each input factor based on prior knowledge or data. Then the measured output variability caused by variation in the input factors, $\mathrm{Var}(g(\boldsymbol{\Theta}))$, is apportioned to each input factor as a measure of its contribution to output uncertainty. Measures of global SA provide guidance as to which inputs to control or study to reduce their uncertainty, and which are not significant sources of output uncertainty. The most commonly known global SA measures are variance-based, such as the first-order and total effects in Homma and Saltelli (1996), and the Shapley effects in Song et al. (2016).

Local SA, on the other hand, focuses on the impact of small perturbations of θ_i on the outputs, often in the form of a partial derivative of the output with respect to the input. One justification for this approach is the Taylor Series approximation of g around a nominal value of the input $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ (Saltelli et al. 2000):

$$g(\boldsymbol{\theta}^0 + \Delta\boldsymbol{\theta}) = g(\boldsymbol{\theta}^0) + \sum_j \frac{\partial g}{\partial \theta_j}\Delta\theta_j + \frac{1}{2}\sum_j\sum_l \frac{\partial^2 g}{\partial \theta_j \partial \theta_l}\Delta\theta_j\Delta\theta_l + \cdots.$$

As the partial derivatives implied by a computer model typically have to be estimated via numerical methods, the first-order or gradient terms $\partial g(\boldsymbol{\theta})/\partial \theta_i$ usually have to suffice, meaning the SA is truly local, applying only to an infinitesimal perturbation around the nominal setting. Our SA measure is a partial derivative.

In the context of stochastic simulation models, SA is more complicated because the input factors $\boldsymbol{\theta}$ might be parameters of the input distributions that represent the inherent randomness in the system. In this case there is both sensitivity to the values of these parameters, and also uncertainty about their nominal values $\boldsymbol{\theta}^0$ if they were estimated from data. Much recent work has been done on quantifying or hedging against the uncertainty in the simulation output due to the uncertainties in the values of the input parameters, which is referred as "input uncertainty" in the simulation literature (Lam 2016; Song et al. 2014).

Our goal is to assess the local sensitivity of the simulation output with respect to the *variance* of the input models, assuming the nominal values $\boldsymbol{\theta}^0$ are known. Thus we go beyond sensitivity of the output to the input distribution mean, which is common and even implemented in some commercial software (e.g., Simio®). We consider parameteric input distributions having parameters such as mean, variance, shape, scale, rate, etc. However, sensitivity of the output with respect to the natural input parameters themselves is often difficult to interpret; this can be true even when the variance of the distribution is one of the parameters. Therefore, we propose a new family of local sensitivity measures that enable us to quantify the sensitivity of the output mean to a change in the variance of the input models along a specified *direction* in the input-parameter space.

Like other local SA measures for stochastic simulation, our proposed measure requires the estimation of a stochastic gradient. Existing simulation-based techniques can be categorized into two groups: indirect and direct methods. Indirect methods include finite differences (brute force) and the simultaneous perturbation method, both of which require additional simulation runs that increase with the dimensionality of the problem (Fu 2015). The direct methods includes infinitesimal perturbation analysis, likelihood ratio (score function) method, and measure-valued differentiation (Fu 2015), which all require information about the underlying simulation model. Instead of using these techniques, we incorporate the method of Wieland and Schmeiser (2006) which is particularly well-suited to estimating output gradients with respect to input parameters without additional simulation effort beyond the nominal experiment.

The paper is organized as follows. We define our new family of sensitivity measures in Section 2, and provide an estimator and establish its properties in Sections 3–4. An empirical illustration is found in Section 5, followed by conclusions in Section 6.

2 A NEW FAMILY OF SENSITIVITY MEASURES

Consider a simulation model with K independent, scalar, parametric input distributions denoted $F^{(1)}(\cdot|\boldsymbol{\theta}^{(1)}), F^{(2)}(\cdot|\boldsymbol{\theta}^{(2)}), \ldots, F^{(K)}(\cdot|\boldsymbol{\theta}^{(K)})$, having in total $q \geq K$ input parameters (because for some distributions $\boldsymbol{\theta}$ is a vector). The simulation output of interest can be represented as

$$Y(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(K)}) = \eta(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(K)}) + \varepsilon(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(K)})$$

where $\boldsymbol{\theta}^{(i)} \in \Re^{p_i}$, with $p_i \geq 1$ represents the parameters of input distribution i, $\eta(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(K)})$ is the expected value of the simulation output given the parameters, and $\varepsilon(\cdot)$ is the corresponding stochastic noise with mean 0. In this paper we consider the parameters $\boldsymbol{\theta}^{(i)}$ to be fixed, so where no confusion will arise we will simply write Y. We also let $X^{(i)}$ represent a random variable with distribution $F^{(i)}$.

Consider the M/G/∞ queue as an example. If "G" is the gamma distribution, then there are $K = 2$ input random variables with $q = 3$ parameters: the interarrival time following an exponential distribution with $\boldsymbol{\theta}^{(1)} = \lambda$ ($X^{(1)} \sim \text{exponential}(\lambda)$), and the service time following a gamma distribution with $\boldsymbol{\theta}^{(2)} = (\alpha, \beta)$

($X^{(2)} \sim \text{gamma}(\alpha,\beta)$). The output, Y, could be, for instance, the number of customers in the system, and we are interested in $\mathrm{E}(Y)$. For each input distribution $F^{(k)}$ with parameter $\boldsymbol{\theta}^{(k)}$, denote its mean and variance by $\mu_k(\boldsymbol{\theta}^{(k)})$ and $\sigma_k^2(\boldsymbol{\theta}^{(k)})$, respectively. In the case of the M/G/∞ queue, $\mu_1(\boldsymbol{\theta}^{(1)}) = 1/\lambda$, $\sigma_1^2(\boldsymbol{\theta}^{(1)}) = 1/\lambda^2$, $\mu_2(\boldsymbol{\theta}^{(2)}) = \alpha/\beta$, and $\sigma_2^2(\boldsymbol{\theta}^{(2)}) = \alpha/\beta^2$.

Our local sensitivity analysis is with respect to each input distribution separately, so for ease of exposition we focus first on a single input distribution $X \sim F(\cdot|\boldsymbol{\theta})$, with parameter $\boldsymbol{\theta} \in \Re^p$ implying mean $\mu = \mu(\boldsymbol{\theta})$ and variance $\sigma^2 = \sigma(\boldsymbol{\theta})$.

In this paper we address sensitivity of the output mean, $\mathrm{E}(Y) = \eta(\boldsymbol{\theta})$, with respect to the *variance* of the input F; we call this the *mean sensitivity to variance (MSV)*. What we want, conceptually, is $\partial\mathrm{E}(Y)/\partial\sigma^2$, but this partial derivative is not well defined when there are multiple ways to achieve a change in σ^2. Again consider the M/G/∞ when Y is the number of customers in the system in steady state and σ^2 is the variance of the service-time distribution. We know that $Y \sim \text{Poisson}(\lambda\alpha/\beta)$, $\mathrm{E}(Y) = \lambda\alpha/\beta$, and $\sigma^2 = \alpha/\beta^2$. Therefore, the MSV when changing α with β fixed, or vice versa, are, respectively,

$$\frac{\partial\mathrm{E}(Y)_\alpha}{\partial\sigma^2} = \frac{\partial\mathrm{E}(Y)}{\partial\alpha}\frac{\partial\alpha}{\partial\sigma^2} = \frac{\lambda}{\beta}\beta^2 = \lambda\beta$$
$$\frac{\partial\mathrm{E}(Y)_\beta}{\partial\sigma^2} = \frac{\partial\mathrm{E}(Y)}{\partial\beta}\frac{\partial\beta}{\partial\sigma^2} = \frac{-\lambda\alpha}{\beta^2}\left(\frac{-\beta^3}{2\alpha}\right) = \frac{\lambda\beta}{2}.$$

Clearly *different* changes in $\boldsymbol{\theta}$ may lead to the *same* change in σ^2 but *different* changes in $\mathrm{E}(Y)$. Therefore, the *direction* of the change in (α,β) needs to be specified to obtain a unique MSV. This ambiguity will occur unless θ is scalar, or $X = \mu + \sigma W$ where W has mean 0 and variance 1.

When a function $g(\cdot)$ of a vector argument $\mathbf{x}$ is differentiable, then the *directional derivative* of the function in the direction $\vec{\mathbf{p}}$ is given by

$$D(g(\mathbf{x});\vec{\mathbf{p}})) \stackrel{\text{def}}{=} \lim_{\varepsilon\to 0}\frac{g(\mathbf{x}+\varepsilon\vec{\mathbf{p}}) - g(\mathbf{x})}{\varepsilon\|\vec{\mathbf{p}}\|} = \nabla g(\mathbf{x})^T\frac{\vec{\mathbf{p}}}{\|\vec{\mathbf{p}}\|}$$

where ∇ denotes the gradient operator. The directional derivative can be interpreted as the rate of increase of $g(\cdot)$ per unit of distance moved in the direction given by $\vec{\mathbf{p}}$.

From here on we let $\boldsymbol{\theta}$ denote the vector of parameters, treated as a variable, and $\boldsymbol{\theta}^0$ as the nominal (fixed) setting. Throughout this paper we assume that $\boldsymbol{\theta}^0$ is known, i.e., no input uncertainty. We propose the family of MSV measures obtained from the directional derivative of $\mathrm{E}(Y)$ with respect to σ^2 along the normed direction $\vec{\mathbf{d}}$ from the nominal parameter setting $\boldsymbol{\theta}^0$. Then by using the chain rule for directional derivatives we obtain

$$\mathrm{MSV}_{\vec{\mathbf{d}}} = \frac{\partial\mathrm{E}(Y)}{\partial\sigma^2_{\vec{\mathbf{d}}}} = \frac{\vec{\mathbf{d}}^T\nabla_{\boldsymbol{\theta}^0}\mathrm{E}(Y)}{\vec{\mathbf{d}}^T\nabla_{\boldsymbol{\theta}^0}\sigma^2}. \tag{1}$$

The gradient of the variance of the input X with respect its input parameters $\boldsymbol{\theta}$ at $\boldsymbol{\theta}^0$, $\nabla_{\boldsymbol{\theta}^0}\sigma^2$, will typically be known in closed form or easily computed numerically. Thus, estimating MSV reduces to estimation of $\nabla_{\boldsymbol{\theta}^0}\mathrm{E}(Y)$; we describe a method to do this in Section 3 below.

Although MSV can be computed along any direction, our definition of variance sensitivity will only be valuable if there are practically useful directions. The useful directions we see are the steepest-ascent direction, the minimum-mean-change direction, or a problem-specific "bring your own direction." We describe and illusrate the first two of these in the next two subsections.

2.1 Steepest-Ascent Direction

The steepest-ascent direction is the $\boldsymbol{\theta}$ direction along which σ^2 increases the fastest: $\vec{\mathbf{d}} = \nabla_{\boldsymbol{\theta}^0}\sigma^2/\|\nabla_{\boldsymbol{\theta}^0}\sigma^2\|$. Plugging this direction into (1) we have

$$\mathrm{MSV}_{\vec{\mathbf{d}}} = \frac{\partial \mathrm{E}(Y)}{\partial \sigma^2_{\vec{\mathbf{d}}}} = \frac{\vec{\mathbf{d}}^\top \nabla_{\boldsymbol{\theta}^0}\mathrm{E}(Y)}{\vec{\mathbf{d}}^\top \nabla_{\boldsymbol{\theta}^0}\sigma^2} = \frac{(\nabla_{\boldsymbol{\theta}^0}\sigma^2)^\top \nabla_{\boldsymbol{\theta}^0}\mathrm{E}(Y)}{\|\nabla_{\boldsymbol{\theta}^0}\sigma^2\|^2}.$$

Writing each term explicitly, we have $\vec{\mathbf{d}} = \left(\frac{\partial\sigma^2}{\partial\theta_1}\left(\sqrt{\sum_{i=1}^p(\frac{\partial\sigma^2}{\partial\theta_i})^2}\right)^{-1}, \ldots, \frac{\partial\sigma^2}{\partial\theta_p}\left(\sqrt{\sum_{i=1}^p(\frac{\partial\sigma^2}{\partial\theta_i})^2}\right)^{-1}\right)^\top = (d_1, d_2, \ldots, d_p)^\top$ and the resulting $\mathrm{MSV}_{\vec{\mathbf{d}}}$ is

$$\begin{aligned}
\frac{\partial \mathrm{E}(Y)}{\partial \sigma^2_{\vec{\mathbf{d}}}} &= \frac{d_1\frac{\partial \mathrm{E}(Y)}{\partial\theta_1} + \cdots + d_p\frac{\partial \mathrm{E}(Y)}{\partial\theta_p}}{d_1\frac{\partial\sigma^2}{\partial\theta_1} + \cdots + d_p\frac{\partial\sigma^2}{\partial\theta_p}} \\
&= \left(\frac{d_1\frac{\partial\sigma^2}{\partial\theta_1}}{\sum_{i=1}^p d_i\frac{\partial\sigma^2}{\partial\theta_i}}\right)\frac{\frac{\partial \mathrm{E}(Y)}{\partial\theta_1}}{\frac{\partial\sigma^2}{\partial\theta_1}} + \cdots + \left(\frac{d_p\frac{\partial\sigma^2}{\partial\theta_p}}{\sum_{i=1}^p d_i\frac{\partial\sigma^2}{\partial\theta_i}}\right)\frac{\frac{\partial \mathrm{E}(Y)}{\partial\theta_p}}{\frac{\partial\sigma^2}{\partial\theta_p}} \\
&= \left(\frac{(\frac{\partial\sigma^2}{\partial\theta_1})^2}{\sum_{i=1}^p(\frac{\partial\sigma^2}{\partial\theta_i})^2}\right)\frac{\frac{\partial \mathrm{E}(Y)}{\partial\theta_1}}{\frac{\partial\sigma^2}{\partial\theta_1}} + \cdots + \left(\frac{(\frac{\partial\sigma^2}{\partial\theta_p})^2}{\sum_{i=1}^p(\frac{\partial\sigma^2}{\partial\theta_p})^2}\right)\frac{\frac{\partial \mathrm{E}(Y)}{\partial\theta_p}}{\frac{\partial\sigma^2}{\partial\theta_p}} \\
&= w_1\frac{\frac{\partial \mathrm{E}(Y)}{\partial\theta_1}}{\frac{\partial\sigma^2}{\partial\theta_1}} + \cdots + w_p\frac{\frac{\partial \mathrm{E}(Y)}{\partial\theta_p}}{\frac{\partial\sigma^2}{\partial\theta_p}}.
\end{aligned}$$

Notice that $\sum_{i=1}^p w_i = 1$. In fact it is easy to show that the MSV for any direction can be expressed as a convex combination of the terms $\frac{\partial \mathrm{E}(Y)}{\partial\theta_i}/\frac{\partial\sigma^2}{\partial\theta_i}$ for $i = 1, 2, \ldots, p$.

For the $\mathrm{M/G/\infty}$ queue when Y is the number of customers in the system in steady state, the unit-norm, steepest-ascent direction of the service process is

$$\vec{\mathbf{d}} = \left(\frac{\beta}{\sqrt{4\alpha^2+\beta^2}}, -\frac{2\alpha}{\sqrt{4\alpha^2+\beta^2}}\right)$$

which results in the $\mathrm{MSV}_{\vec{\mathbf{d}}}$ of

$$\frac{\partial \mathrm{E}(Y)}{\partial \sigma^2_{\vec{\mathbf{d}}}} = \frac{\vec{\mathbf{d}}^\top \nabla_{\boldsymbol{\theta}}\mathrm{E}(Y)}{\vec{\mathbf{d}}^\top \nabla_{\boldsymbol{\theta}}\sigma^2} = \frac{\left(\frac{\beta}{\sqrt{4\alpha^2+\beta^2}}, -\frac{2\alpha}{\sqrt{4\alpha^2+\beta^2}}\right)^\top \begin{pmatrix} \frac{\lambda}{\beta} \\ -\frac{\lambda\alpha}{\beta^2} \end{pmatrix}}{\left(\frac{\beta}{\sqrt{4\alpha^2+\beta^2}}, -\frac{2\alpha}{\sqrt{4\alpha^2+\beta^2}}\right)^\top \begin{pmatrix} \frac{1}{\beta^2} \\ -\frac{2\alpha}{\beta^3} \end{pmatrix}} = \frac{\lambda\beta^3 + 2\lambda\alpha^2\beta}{\beta^2 + 4\alpha^2} > 0 \qquad (2)$$

evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}^0 = (\lambda^0, \alpha^0, \beta^0)^\top$.

The steepest-ascent direction is a defensive choice: Change $\boldsymbol{\theta}^0$ in the direction that most rapidly increases the variance of the input distribution. Assuming an increase in variance is bad, this is the direction nature would choose to be the most disruptive.

2.2 Minimum-Mean-Change Direction

This direction minimizes the rate of change in the mean of the input while increasing its variance:

$$\begin{aligned} \underset{\vec{\mathbf{d}}\in\Re^p}{\text{Minimize:}}\ & \left|\vec{\mathbf{d}}^\top \nabla_{\boldsymbol{\theta}^0}\mu(\boldsymbol{\theta})\right| \\ \text{subject to: } \vec{\mathbf{d}}^\top \nabla_{\boldsymbol{\theta}^0}\sigma^2(\boldsymbol{\theta}) &> 0 \\ \|\vec{\mathbf{d}}\| &= 1. \end{aligned}$$

For many common input distributions with $p \geq 2$ parameters the optimal objective function value of 0 is achieved at the direction perpendicular to the steep-ascent direction of $\mu(\boldsymbol{\theta})$; this includes the normal, gamma, beta, Weibull, inverse Gaussian, and Pareto distributions. When the input parameter is a scalar, $p = 1$, then the optimal objective function value is typically greater than zero because $\mu(\boldsymbol{\theta})$ is a function of $\sigma^2(\boldsymbol{\theta})$; e.g., the Bernoulli, exponential, geometric and Rayleigh distributions.

For the $\mathrm{M}/\mathrm{G}/\infty$ queue when Y is the number of customers in the system in steady state, consider the $\mathrm{MSV}_{\vec{\mathbf{d}}}$ with $\vec{\mathbf{d}}$ the minimum-mean-change direction of the service process. The solution to the optimization problem, which has objective function value 0, is

$$\vec{\mathbf{d}} = \left(-\frac{\alpha}{\sqrt{\alpha^2+\beta^2}},\ -\frac{\beta}{\sqrt{\alpha^2+\beta^2}}\right)$$

which results in the $\mathrm{MSV}_{\vec{\mathbf{d}}}$ of

$$\frac{\partial \mathrm{E}(Y)}{\partial \sigma^2_{\vec{\mathbf{d}}}} = \frac{\vec{\mathbf{d}}^\top \nabla_{\boldsymbol{\theta}} \mathrm{E}(Y)}{\vec{\mathbf{d}}^\top \nabla_{\boldsymbol{\theta}} \sigma^2} = \frac{\left(-\frac{\alpha}{\sqrt{\alpha^2+\beta^2}},\ -\frac{\beta}{\sqrt{\alpha^2+\beta^2}}\right)^\top \begin{pmatrix} \frac{\lambda}{\beta} \\ -\frac{\lambda\alpha}{\beta^2} \end{pmatrix}}{\vec{\mathbf{d}}^\top \nabla_{\boldsymbol{\theta}} \sigma^2} = 0. \tag{3}$$

This result makes sense because the expected number of customers in an $\mathrm{M}/\mathrm{G}/\infty$ queue in steady state does not depend on the variance of the service-time distribution, only the mean. Comparing (2) with (3) we see that the MSVs along different directions are dramatically different.

The minimum-mean-change direction is a natural choice for isolating the effect of input-distribution variance with minimal change to its location.

3 ESTIMATING MSV

The key to estimating $\mathrm{MSV}_{\vec{\mathbf{d}}}$ in (1) is estimating $\nabla_{\boldsymbol{\theta}}\mathrm{E}(Y)$. Because $\nabla_{\boldsymbol{\theta}}\mathrm{E}(Y)$ is the gradient with respect to the simulation input-distribution parameters, we are able to use and extend the gradient estimation method of Wieland and Schmeiser (2006) that is easy to implement, computationally cheap, and requires no alteration of the simulation model or supplementary experiments; see also Lin et al. (2015). We describe this method below.

Suppose there is one input distribution with a scalar parameter θ. To execute the simulation we set its value to θ^0. To be concrete, let θ be the rate parameter of the exponential distribution describing the interarrival times in the $\mathrm{M}/\mathrm{G}/\infty$ queue. Among a total of n replications, the jth replication generates m_j independent and identically distributed (i.i.d.) interarrival times, $X_{ij}, i = 1, 2, \ldots, m_j$, where m_j could be random. The input parameter of X under the nominal setting, θ^0, can be estimated (e.g., via maximum likelihood estimation (MLE)) from the random variates generated in each replication. In the $\mathrm{M}/\mathrm{G}/\infty$ queue, $\widehat{\theta}_j = 1/(\sum_{i=1}^{m_j} X_{ij}/m_j)$ is the MLE of θ^0 from replication j.

Replication j also generates output Y_j. For instance, in the $\mathrm{M}/\mathrm{G}/\infty$ queue, if $Z_j(t)$ is the number of customers in the system at time t during a specified period of time $[0,T]$, then $Y_j = T^{-1}\int_0^T Z_j(t)\,dt$ could be the output of interest to estimate the expected number of customers in the system in steady state. Thus, from n replications we observe i.i.d. pairs $(Y_j, \widehat{\theta}_j),\ j = 1, 2, \ldots, n$.

Suppose that the distribution of $(Y_j, \widehat{\theta}_j)$ is bivariate normal with mean $(\eta(\theta^0), \theta^0)$. Then

$$\begin{aligned} \mathrm{E}(Y|\widehat{\theta}_j) &= \beta_0 + \beta\widehat{\theta} \\ Y &\overset{D}{=} \beta_0 + \beta\widehat{\theta} + \varepsilon \end{aligned} \tag{4}$$

where ε is independent of $\widehat{\theta}$, is normally distributed with mean zero, and $\beta = \mathrm{Cov}(Y, \widehat{\theta})/\mathrm{Var}(\widehat{\theta})$. Therefore, the $\partial \mathrm{E}(Y)/\partial\theta$ under the nominal setting θ^0 is β, and the ordinary least squares estimator of β is

$$\widehat{\beta} = \frac{\sum_{j=1}^{n}(Y_j - \bar{Y})(\widehat{\theta}_j - \bar{\theta})}{\sum_{j=1}^{n}(Y_j - \bar{Y})^2}$$

where $\bar{Y} = \sum_{j=1}^{n} Y_j/n$ and $\bar{\theta} = \sum_{j=1}^{n} \widehat{\theta}_j/n$. This is the key idea of Wieland and Schmeiser (2006): fix the value of θ at θ^0 and estimate the sensitivity of the response Y_j to the realized parameter $\widehat{\theta}_j$ as it varies across n replications. Because this relationship is linear when they are bivariate normal, the derivative at $\theta = \theta^0$ can be obtained.

When $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^\top$ is a vector parameter, relationship (4) still applies if the distribution of $(Y_j, \widehat{\boldsymbol{\theta}}_j)$ is multivariate normal: $\boldsymbol{\beta}^\top = \Sigma_{Y,\widehat{\boldsymbol{\theta}}}\left(\Sigma_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}\right)^{-1}$ (Anderson 1984). Therefore, $\boldsymbol{\beta}$ is the gradient of $\mathrm{E}(Y)$ with respect to $\boldsymbol{\theta}$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}^0$, and we can estimate $\boldsymbol{\beta}$ by

$$\widehat{\boldsymbol{\beta}}^\top = \widehat{\Sigma}_{Y,\widehat{\boldsymbol{\theta}}}\left(\widehat{\Sigma}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}\right)^{-1} \tag{5}$$

where $\widehat{\Sigma}_{Y,\widehat{\boldsymbol{\theta}}}$ is the sample covariance matrix between Y and $\widehat{\boldsymbol{\theta}}$ and $\widehat{\Sigma}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}$ is the sample variance matrix of $\widehat{\boldsymbol{\theta}}$. Thus, the estimator of the gradient of $\mathrm{E}(Y)$ with respect to $\boldsymbol{\theta}$ under the nominal setting is

$$\widehat{\nabla}_{\boldsymbol{\theta}^0}\mathrm{E}(Y) \equiv \widehat{\boldsymbol{\beta}} \tag{6}$$

(Lin et al. 2015).

Notice that a sufficient condition to apply the method of Wieland and Schmeiser (2006) is the normality of $(Y_j, \widehat{\boldsymbol{\theta}}_j)$. When both Y_j and $\widehat{\boldsymbol{\theta}}_j$ are the average of a large number of outputs within replication j, or MLEs of their respective parameters, then it is plausible to approximate the distribution as normal. When this is not the case batching the replications can be used to induce normality, as suggested in Wieland and Schmeiser (2006).

4 VARIANCE OF THE MSV ESTIMATOR

In this section we derive the variance of the MSV estimator when $(Y_j, \widehat{\boldsymbol{\theta}}_j)$ are multivariate normal. From here on $\boldsymbol{\theta}$ and $\widehat{\boldsymbol{\theta}}_j$ are $q \times 1$, containing the parameters across all K input distributions.

Plugging (6) into (1), the estimator of MSV under the nominal setting is

$$\widehat{\mathrm{MSV}}_{\vec{\mathbf{d}}} = \vec{\mathbf{d}}^\top\widehat{\boldsymbol{\beta}}\left(\vec{\mathbf{d}}^\top\nabla_{\boldsymbol{\theta}^0}\sigma^2\right)^{-1}. \tag{7}$$

The estimator is clearly unbiased as it is a linear combination of unbiased estimators. The only random quantity in (7) is $\widehat{\boldsymbol{\beta}}$. Therefore,

$$\mathrm{Var}\left(\widehat{\mathrm{MSV}}_{\vec{\mathbf{d}}}\right) = \vec{\mathbf{d}}^\top\mathrm{Var}(\widehat{\boldsymbol{\beta}})\vec{\mathbf{d}}\left(\vec{\mathbf{d}}^\top\nabla_{\boldsymbol{\theta}^0}\sigma^2\right)^{-2}. \tag{8}$$

We know that $\widehat{\boldsymbol{\beta}}$ in (5) is equivalent to the ordinary least-squares (OLS) estimator of the slope coefficients:

$$\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_{1,\text{OLS}} \text{ where } \widehat{\boldsymbol{\beta}}_{\text{OLS}} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y} = \begin{bmatrix} \widehat{\beta}_{0,\text{OLS}} \\ \widehat{\boldsymbol{\beta}}_{1,\text{OLS}} \end{bmatrix},$$

$\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^\top$ and

$$\mathbf{X} = \begin{bmatrix} 1 & \widehat{\theta}_{11} & \cdots & \widehat{\theta}_{1q} \\ 1 & \widehat{\theta}_{21} & \cdots & \widehat{\theta}_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \widehat{\theta}_{n1} & \cdots & \widehat{\theta}_{nq} \end{bmatrix}.$$

Here $\widehat{\theta}_{ji}$ is the ith element of $\widehat{\boldsymbol{\theta}}_j$ obtained from the jth replication. Therefore, by the law of total variance we have

$$\begin{aligned} \text{Var}\left[\widehat{\boldsymbol{\beta}}_{\text{OLS}}\right] &= \text{E}\left[\text{Var}\left(\widehat{\boldsymbol{\beta}}_{\text{OLS}}|\mathbf{X}\right)\right] + \text{Var}\left[\text{E}\left(\widehat{\boldsymbol{\beta}}_{\text{OLS}}|\mathbf{X}\right)\right] \\ &= \text{E}\left[\sigma_\varepsilon^2\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right] + \text{Var}[\boldsymbol{\beta}_{\text{OLS}}] \\ &= \sigma_\varepsilon^2 \text{E}\left[\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right] + 0 \end{aligned}$$

where σ_ε^2 is the conditional variance of Y given $\mathbf{X}$ and $\boldsymbol{\beta}_{\text{OLS}}$ is the true regression coefficient. Here the first term in the second equality is derived via standard multiple linear regression analysis and the third equality is because the OLS estimators are unbiased under our assumptions.

To obtain the variance of $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_{1,\text{OLS}}$, we need to focus on the lower right sub-matrix of $\text{E}\left[\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right]$. In the Appendix we show that it is equivalent to the inverse of $(n-1)\widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}$, which under the multivariate normality assumption has a $\mathbf{W}_q\left(\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}, n-1\right)$ distribution, i.e., a Wishart distribution of dimension q with $n-1$ degrees of freedom and variance matrix $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}$. Thus,

$$\begin{aligned} \widehat{\Sigma}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}} &\sim \frac{1}{n-1}\mathbf{W}_q\left(\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}, n-1\right) \\ \Rightarrow \left(\widehat{\Sigma}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}\right)^{-1} &\sim (n-1)\mathbf{W}_q^{-1}\left(\boldsymbol{\Sigma}^{-1}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}, n-1\right) \\ \Rightarrow \text{E}\left[\left(\widehat{\Sigma}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}\right)^{-1}\right] &= \frac{n-1}{n-q-2}\boldsymbol{\Sigma}^{-1}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}. \end{aligned}$$

Therefore,

$$\text{Var}(\widehat{\boldsymbol{\beta}}) = \sigma_\varepsilon^2 \text{E}\left[\frac{1}{n-1}\left(\widehat{\Sigma}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}\right)^{-1}\right] = \frac{\sigma_\varepsilon^2}{n-q-2}\boldsymbol{\Sigma}^{-1}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}.$$

To estimate $\text{Var}(\widehat{\boldsymbol{\beta}})$, we estimate σ_ε^2 using $s_\varepsilon^2 = \text{SSE}/(n-q-1)$, where SSE is the sum of squared errors of the multiple linear regression of Y on $\widehat{\boldsymbol{\theta}}$, and estimate $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}$ by the sample variance-covariance matrix $\widehat{\Sigma}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}$. Therefore, our estimator of the variance of $\widehat{\boldsymbol{\beta}}$ is

$$\widehat{\text{Var}}(\widehat{\boldsymbol{\beta}}) = \frac{s_\varepsilon^2}{n-q-2}\left(\widehat{\Sigma}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}\right)^{-1}$$

which can be plugged into (8) to obtain an estimator of the variance of the MSV estimator:

$$\widehat{\mathrm{Var}}\left[\widehat{\mathrm{MSV}}_{\vec{\mathbf{d}}}\right] = \vec{\mathbf{d}}^\top \left(\frac{s_\varepsilon^2}{n-q-2}\left(\widehat{\Sigma}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}\right)^{-1}\right)\vec{\mathbf{d}}\left(\vec{\mathbf{d}}^\top \nabla_{\boldsymbol{\theta}^0}\sigma^2\right)^{-2}.$$

When there is concern that $(Y_j, \widehat{\boldsymbol{\theta}}_j)$ may not be approximately normally distributed, then batching can be used to improve the approximation. Let $1 \le b \le \lfloor n/2 \rfloor$ be the batch size and $k = \lfloor n/b \rfloor$ be the number of batches; for convenience of exposition we assume $n = kb$ from here on. The batch means are

$$\bar{Y}_h(b) = \frac{1}{b}\sum_{j=(h-1)b+1}^{hb} Y_j$$

$$\bar{\boldsymbol{\theta}}_h(b) = \frac{1}{b}\sum_{j=(h-1)b+1}^{hb} \widehat{\boldsymbol{\theta}}_j$$

for $h = 1, 2, \ldots, k$. Both $\bar{Y}_j(b)$ and $\bar{\boldsymbol{\theta}}_j(b)$ are the average of b replications. As b increases it is more and more plausible that approximate multivariate normality of $(\bar{Y}_j(b), \bar{\boldsymbol{\theta}}_j(b))$ holds, and we may estimate the gradient (and therefore MSV) from the batch means rather than the raw replication results. However, there is a loss of degrees of freedom: If the batch means are multivariate normal then

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}(b)) = \frac{\sigma_\varepsilon^2}{k-q-2}\boldsymbol{\Sigma}^{-1}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}.$$

One way to quantify the loss is to note that if multivariate normality holds without batching, then the variance inflation due to batching is

$$\frac{\mathrm{Var}\left[\widehat{\mathrm{MSV}}_{\vec{\mathbf{d}}}(b)\right]}{\mathrm{Var}\left[\widehat{\mathrm{MSV}}_{\vec{\mathbf{d}}}\right]} = \frac{n-q-2}{k-q-2} \ge 1.$$

5 ILLUSTRATION

In this section we illustrate the estimation and interpretation of MSV for a simulation model of a simplified semiconductor wafer fab. The process consists of two basic steps, diffusion and lithography, each of which contains sub-steps as indicated in Figure 1.

In this fab cassettes are released exactly 1 every hour, 24 hours a day, 7 days a week. The raw cassette begins at the Clean station of diffusion, and passes through diffusion and lithography three times before leaving the process. The times spent at Oxidize, Coat, and Stepper are deterministic, specifically 2 hours, 1.5 hours, and 1.5 hours, respectively. The times spent at Clean, Load, Unload and Develop are variable with the $K = 4$ distributions having the $q = 7$ parameters specified in Table 1. The goal is to estimate the sensitivity of the mean cycle time of the manufactured cassettes to the variance of each of the four input distributions when their variances change along the steepest-ascent and the minimum-mean-change directions. The simulation was run for 900 replications, each with a run-length of 7 days (168 hours).

Because the distributions of the MLEs of these parameters are asymptotically normal and the mean cycle time is the average of a large number of cycle times collected within each replication, it is plausible to approximate the distribution of $(Y, \widehat{\boldsymbol{\theta}})$ as multivariate normal. If so, then the relationship between the conditional expectation of the mean cycle time and the MLE estimators is linear and we can obtain the gradient estimator, $\widehat{\boldsymbol{\beta}}$, with its variance via multiple linear regression. The summary of the fitted model is shown in Table 2.

In Table 2 we see that all predictors are significant and the Clean time is the largest one, which is not surprising because the Clean time has the largest mean among the four input models and the main

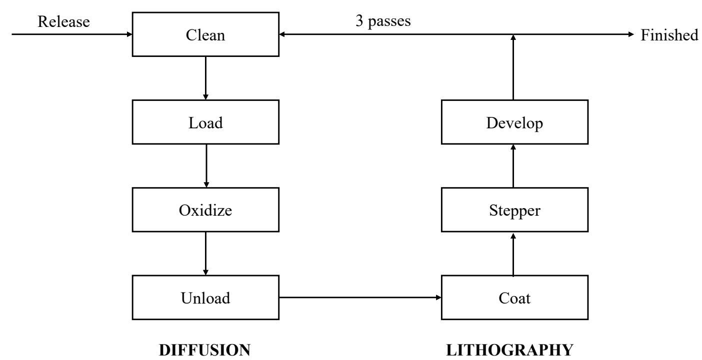

Figure 1: Diagram of a simplified wafer fab.

Table 1: Experiment setup.

Input	Distribution	Parameter	Nominal Value
Clean Time	exponential	rate	$\lambda^0 = 0.5490$
Develop Time	gamma	(shape, scale)	$(\alpha^0, \beta^0) = (5.0, 10.1297)$
Load & Unload Time	lognormal	(meanlog, sdlog)	$(\mu^0, \sigma^0) = (-1.4283, 0.5762)$

driver of the congestion in the queue is the mean. Also, it makes sense that the coefficient associated with CL_MLE is negative because the increase of the rate of the exponential distribution will decrease its mean which in turn helps mitigate the congestion of the station. The adjusted R^2 of 0.812 suggests that this linear model fits well to the set of 900 observations. We also verified the normality, homoscedasticity, and linearity assumptions of the model via diagnostic plots including Q-Q plot, Residual vs. Fitted Values plot, and Added Variable plot. Multicollinearity and outliers were checked through calculating variance inflation factors of the coefficients and Cook's distance of each observation. In summary, we conclude that the linear model fits well to the data.

The MSV estimators and their corresponding standard errors were obtained through plugging the values of the coefficients into Equations (7) and (8). The gradients of the variances of the input distributions, $\nabla_{\boldsymbol{\theta}}\sigma^2$, for Clean (exponential), Develop (gamma), and load/unload (lognormal) are

$$
\begin{aligned}
\nabla_{\boldsymbol{\theta}}\sigma^2_{\mathrm{CL}} &= -\frac{2}{\lambda^3} \\
\nabla_{\boldsymbol{\theta}}\sigma^2_{\mathrm{DE}} &= \left(\frac{1}{\beta^2},\ -\frac{2\alpha}{\beta^3}\right) \\
\nabla_{\boldsymbol{\theta}}\sigma^2_{\mathrm{L/UL}} &= \left(2(e^{\sigma^2}-1)e^{2\mu+\sigma^2},\ 2\sigma(2e^{\sigma^2}-1)e^{2\mu+\sigma^2}\right).
\end{aligned}
$$

The MSV estimates are shown in Table 3. Consider $\mathrm{MSV}_{\mathrm{DE}}$, the sensitivity of mean cycle time to the variance of the Develop station, as an example. The sensitivity in the steepest-ascent and minimum-mean-change directions are signficantly different; this is because the mean processing time is the primary driver of queueing congestion, and the steepest-ascent direction changes the mean as well as the variance, while the minimum-mean-change direction does not. Throughout we see that the sensitivity to variance when the mean is held constant is less, and in some cases may not be statistically significant. To illustrate why, the

Table 2: Regression results (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 2e-16$).

Parameter	Coefficient	Significance	StdErr
CL_Rate	-10.714	$****$	(0.177)
DE_Alpha	0.346	$***$	(0.047)
DE_Beta	-0.173	$***$	(0.022)
L_LMean	0.916	$***$	(0.181)
L_LStDev	0.736	$**$	(0.245)
UL_LMean	0.821	$***$	(0.177)
UL_LStDev	0.851	$***$	(0.248)
Intercept	31.544	$****$	(0.431)
Observations	900		
R^2	0.813		
Adjusted R^2	0.812		
Residual Std. Error	0.140 (df = 892)		
F Statistic	554.042^{***} (df = 7; 892)		

Table 3: MSV values with standard errors.

$MSV_{Input,Direction}$	Direction	MSV (hr/hr^2)	Std. Error (hr/hr^2)	$\Delta\mu$ (hr)
MSV_{CL}	Steepest Ascent	0.8870	0.0146	0.0249
$MSV_{DE,A}$	Steepest Ascent	26.848	3.4814	7.6299
$MSV_{DE,M}$	Min Mean Change	0.3343	1.6009	0
$MSV_{L,A}$	Steepest Ascent	5.7417	1.3413	1.4344
$MSV_{L,M}$	Min Mean Change	1.6170	2.0725	0
$MSV_{UL,A}$	Steepest Ascent	6.1576	1.3665	1.4344
$MSV_{UL,M}$	Min Mean Change	2.9342	2.0745	0

$\Delta\mu$ column shows $\Delta\mu(\boldsymbol{\theta}^0) = \vec{\mathbf{d}}^T \nabla_{\boldsymbol{\theta}^0}\mu / \vec{\mathbf{d}}^T \nabla_{\boldsymbol{\theta}^0}\sigma^2$ which is approximately how much the mean of each input distribution, $\mu(\boldsymbol{\theta})$, would change if the variance of the distribution $\sigma^2(\boldsymbol{\theta})$ changed one unit. Of course this is 0 in the minimum-mean-change direction. Notice in particular the substantial change in the mean develop time when the variance of the develop time changes in the steepest-ascent direction. Of course, for the exponential distribution of cleaning time there is only a steepest-ascent direction since the mean and variance are linked.

6 CONCLUSIONS

In this paper we defined a new family of sensitivity measures for the expected value of a simulation output with respect to the *variances* of its parametric input distributions, which we call mean sensitivity to variance (MSV). Since probabilities may be represented as expected values, MSV addresses probability sensitivities as well. As variance is often the "corrupting influence" in system performance (e.g., Hopp and Spearman 2011), identifying the inputs whose variance reduction or inflation has the greatest impact is relevant for system design and control. Two members of our new family seem particularly relevant for applications, and any member of the family is easy to estimate requiring no more than OLS regression.

The *definition* of MSV does not depend on the gradient estimator in use, but the properties of our *estimator* of MSV do. Further theoretical and empirical evaluation of estimator performance is clearly in order, including consideration of the infinitesimal-perturbation-analysis, likelihood-ratio and measure-valued-differentiation gradient-estimation methods (Fu 2015) for situations when the method of Wieland

and Schmeiser (2006) is not effective. Empirical evaluation is challengining in that MSV is only derivable mathematically for very simple examples such as the M/G/∞ queue. Therefore, substantial off-line simulation using, say, finite differences, is required to create challenging examples with known MSV.

Sensitivity measures for the *variance* of a simulation output with respect to the variance of its input distributions, which we call variance sensitivity to variance (VSV), may be defined analogously to MSV. However, estimation of VSV is more challenging, so we will present these results in a later paper.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation Grant Number CMMI-1634982.

APPENDIX

Claim: The lower right sub-matrix of the matrix $(\mathbf{X}^\top\mathbf{X})^{-1}$ is equivalent to $\left[(n-1)\widehat{\Sigma}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}\right]^{-1}$, where $\mathbf{X} = [\mathbf{1}, \widehat{\boldsymbol{\theta}}^{(1)}, \ldots, \widehat{\boldsymbol{\theta}}^{(q)}]$, $\mathbf{1} = [1,1,\ldots,1]^\top$, $\widehat{\boldsymbol{\theta}}^{(i)} = [\widehat{\theta}_{i1}, \widehat{\theta}_{i2}, \ldots, \widehat{\theta}_{in}]^\top$ for $i = 1,2,\ldots,q$, and $\widehat{\Sigma}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}}$ is the sample variance-covariance matrix of $\widehat{\boldsymbol{\theta}}$. Here $\widehat{\theta}_{ij}$ is the ith element of $\widehat{\boldsymbol{\theta}}_j$ obtained from the jth replication.

Proof. We first write matrix $\mathbf{X}^\top\mathbf{X}$ as a block matrix and find its inverse

$$\mathbf{X}^\top\mathbf{X} = \left[\begin{array}{c|ccc} \mathbf{1}^\top\mathbf{1} & \mathbf{1}^\top\widehat{\boldsymbol{\theta}}^{(1)} & \ldots & \mathbf{1}^\top\widehat{\boldsymbol{\theta}}^{(q)} \\ \hline (\widehat{\boldsymbol{\theta}}^{(1)})^\top\mathbf{1} & (\widehat{\boldsymbol{\theta}}^{(1)})^\top\widehat{\boldsymbol{\theta}}^{(1)} & \ldots & (\widehat{\boldsymbol{\theta}}^{(1)})^\top\widehat{\boldsymbol{\theta}}^{(q)} \\ \vdots & \vdots & \ddots & \vdots \\ (\widehat{\boldsymbol{\theta}}^{(q)})^\top\mathbf{1} & (\widehat{\boldsymbol{\theta}}^{(q)})^\top\widehat{\boldsymbol{\theta}}^{(1)} & \ldots & (\widehat{\boldsymbol{\theta}}^{(q)})^\top\widehat{\boldsymbol{\theta}}^{(q)} \end{array}\right] = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right]$$

$$\Rightarrow (\mathbf{X}^\top\mathbf{X})^{-1} = \left[\begin{array}{c|c} C_1^{-1} & -A_{11}^{-1}A_{12}C_2^{-1} \\ \hline -C_2^{-1}A_{21}A_{11}^{-1} & C_2^{-1} \end{array}\right]$$

where $C_1 = A_{11} - A_{12}A_{22}^{-1}A_{21}$ and $C_2 = A_{22} - A_{21}A_{11}^{-1}A_{12}$. Writing out C_2 explicitly we have

$$\begin{aligned} C_2 &= A_{22} - A_{21}A_{11}^{-1}A_{12} \\ &= \begin{bmatrix} (\widehat{\boldsymbol{\theta}}^{(1)})^\top\widehat{\boldsymbol{\theta}}^{(1)} & \ldots & (\widehat{\boldsymbol{\theta}}^{(1)})^\top\widehat{\boldsymbol{\theta}}^{(q)} \\ \vdots & \ddots & \vdots \\ (\widehat{\boldsymbol{\theta}}^{(q)})^\top\widehat{\boldsymbol{\theta}}^{(1)} & \ldots & (\widehat{\boldsymbol{\theta}}^{(q)})^\top\widehat{\boldsymbol{\theta}}^{(q)} \end{bmatrix} - \frac{1}{n}\begin{bmatrix} (\widehat{\boldsymbol{\theta}}^{(1)})^\top\mathbf{1} \\ \vdots \\ (\widehat{\boldsymbol{\theta}}^{(q)})^\top\mathbf{1} \end{bmatrix}\begin{bmatrix} \mathbf{1}^\top\widehat{\boldsymbol{\theta}}^{(1)} & \ldots & \mathbf{1}^\top\widehat{\boldsymbol{\theta}}^{(q)} \end{bmatrix} \\ &= \begin{bmatrix} (\widehat{\boldsymbol{\theta}}^{(1)})^\top\widehat{\boldsymbol{\theta}}^{(1)} & \ldots & (\widehat{\boldsymbol{\theta}}^{(1)})^\top\widehat{\boldsymbol{\theta}}^{(q)} \\ \vdots & \ddots & \vdots \\ (\widehat{\boldsymbol{\theta}}^{(q)})^\top\widehat{\boldsymbol{\theta}}^{(1)} & \ldots & (\widehat{\boldsymbol{\theta}}^{(q)})^\top\widehat{\boldsymbol{\theta}}^{(q)} \end{bmatrix} - n\begin{bmatrix} \bar{\theta}_1 \\ \vdots \\ \bar{\theta}_p \end{bmatrix}\begin{bmatrix} \bar{\theta}_1 & \ldots & \bar{\theta}_p \end{bmatrix} \\ &= \begin{bmatrix} (\widehat{\boldsymbol{\theta}}^{(1)})^\top\widehat{\boldsymbol{\theta}}^{(1)} - n(\bar{\theta}_1)^2 & (\widehat{\boldsymbol{\theta}}^{(1)})^\top\widehat{\boldsymbol{\theta}}^{(2)} - n\bar{\theta}_1\bar{\theta}_2 & \ldots & (\widehat{\boldsymbol{\theta}}^{(1)})^\top\widehat{\boldsymbol{\theta}}^{(q)} - n\bar{\theta}_1\bar{\theta}_p \\ (\widehat{\boldsymbol{\theta}}^{(2)})^\top\widehat{\boldsymbol{\theta}}^{(1)} - n\bar{\theta}_2\bar{\theta}_1 & (\widehat{\boldsymbol{\theta}}^{(2)})^\top\widehat{\boldsymbol{\theta}}^{(2)} - n(\bar{\theta}_2)^2 & \ldots & (\widehat{\boldsymbol{\theta}}^{(2)})^\top\widehat{\boldsymbol{\theta}}^{(q)} - n\bar{\theta}_2\bar{\theta}_p \\ \vdots & \vdots & \ddots & \vdots \\ (\widehat{\boldsymbol{\theta}}^{(q)})^\top\widehat{\boldsymbol{\theta}}^{(1)} - n\bar{\theta}_p\bar{\theta}_1 & (\widehat{\boldsymbol{\theta}}^{(q)})^\top\widehat{\boldsymbol{\theta}}^{(2)} - n\bar{\theta}_p\bar{\theta}_2 & \ldots & (\widehat{\boldsymbol{\theta}}^{(q)})^\top\widehat{\boldsymbol{\theta}}^{(q)} - n(\bar{\theta}_p)^2 \end{bmatrix} \\ &= (n-1)\widehat{\Sigma}_{\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}} \quad \text{where } \bar{\theta}_i = \textstyle\sum_{k=1}^n \widehat{\theta}_{ik}/n. \end{aligned}$$

□

REFERENCES

Anderson, T. W. 1984. *An Introduction to Multivariate Statistical Analysis*. 2nd ed. New York: John Wiley & Sons.

Fu, M. C. 2015. "Stochastic Gradient Estimation". In *Handbook of Simulation Optimization*, edited by M. C. Fu, Volume 216 of *International Series in Operations Research & Management Science*, 105–147. New York: Springer.

Homma, T., and A. Saltelli. 1996. "Importance Measures in Global Sensitivity Analysis of Nonlinear Models". *Reliability Engineering & System Safety* 52(1):1–17.

Hopp, W. J., and M. L. Spearman. 2011. *Factory Physics*. Long Grove, Illinois: Waveland Press.

Lam, H. 2016. "Advanced Tutorial: Input Uncertainty and Robust Analysis in Stochastic Simulation". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder et al., 178–192. Piscataway, New Jersey: IEEE.

Lin, Y., E. Song, and B. L. Nelson. 2015. "Single-experiment Input Uncertainty". *Journal of Simulation* 9(3):249–259.

Saltelli, A., K. Chan, and E. M. Scott. 2000. *Sensitivity Analysis*. New York: John Wiley & Sons.

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

Song, E., B. L. Nelson, and J. Staum. 2016. "Shapley Effects for Global Sensitivity Analysis: Theory and Computation". *SIAM/ASA Journal on Uncertainty Quantification* 4(1):1060–1083.

Wieland, J. R., and B. W. Schmeiser. 2006. "Stochastic Gradient Estimation using a Single Design Point". In *Proceedings of the 2006 Winter Simulation Conference*, edited by F. Perrone et al., 390–397. Piscataway, New Jersey: IEEE.

AUTHOR BIOGRAPHIES

WENDY XI JIANG is a Ph.D. student in the Department of Industrial Engineering and Management Sciences at Northwestern University. She is majoring in applied statistics and statistical learning, with minors in analytics and optimization. Her research focus is on stochastic simulation methodology. Her e-mail address is xijiang2020@u.northwestern.edu.

BARRY L. NELSON is the Walter P. Murphy Professor in the Department of Industrial Engineering and Management Sciences at Northwestern University. He is a Fellow of INFORMS and IISE. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems, and he is the author of *Foundations and Methods of Stochastic Simulation: A First Course*, from Springer. His e-mail and web addresses are nelsonb@northwestern.edu and www.iems.northwestern.edu/~nelsonb/

L. JEFF HONG is a Chair Professor in the School of Management and the School of Data Science at Fudan University in China. His research interests include simulation modeling and optimization, stochastic programming, and business analytics. He is currently the simulation area editor of *Operations Research* and an associate editor of *Management Science* and *ACM Transactions on Modeling and Computer Simulation*. His email address is hong_liu@fudan.edu.cn.