*Article*

# Estimation of Expectations and Variance Components in Two-Level Nested Simulation Experiments

**David F. Muñoz** [ID]*

       Departament of Industrial and Operations Engineering, Instituto Tecnológico Autónomo de México, Rio Hondo 1, 01080 Mexico City, Mexico; davidm@itam.mx

\*    Correspondence: davidm@itam.mx; Tel.: +52-55-5628-4118

**Abstract:** When there is uncertainty in the value of parameters of the input random components of a stochastic simulation model, two-level nested simulation algorithms are used to estimate the expectation of performance variables of interest. In the outer level of the algorithm ($n$) observations are generated for the parameters, and in the inner level ($m$) observations of the simulation model are generated with the value of parameters fixed at the value generated in the outer level. In this article, we consider the case in which the observations at both levels of the algorithm are independent, showing how the variance of the observations can be decomposed into the sum of a parametric variance and a stochastic variance. Next, we derive central limit theorems that allow us to compute asymptotic confidence intervals to assess the accuracy of the simulation-based estimators for the point forecast and the variance components. Under this framework, we derive analytical expressions for the point forecast and the variance components of a Bayesian model to forecast sporadic demand; and we use these expressions to illustrate the validity of our theoretical results by performing simulation experiments using this forecast model.

**Keywords:** Bayesian forecasting; stochastic simulation; parameter uncertainty; two-level simulation

## 1. Introduction and Notation

Simulation is widely recognized as an effective technique to produce forecasts, evaluate risk (see, e.g., [1]), animate and illustrate the performance of a system over time (see, e.g., [2]). When there is uncertainty in a component of a simulation model, it is said to be a random component, and it is modeled using a probability distribution and/or a stochastic process that is generated during the simulation run, to produce a stochastic simulation. Random component typically depends on the value of certain parameters, and we will denote by $\theta$ a particular value for the vector of parameters of the random components of a stochastic simulation, and $\Theta$ will denote the random vector that corresponds to the parameter values when there is uncertainty on the value of these parameters.

In general, the output of a stochastic (dynamic) simulation can be regarded as a stochastic process $\{Y(s) : s \geq 0; \Theta\}$, where $Y(s)$ is a random vector (of arbitrary dimension $d$) representing the state of the simulation at time $s \geq 0$. The term transient simulation applies to a dynamic simulation that has a well-defined termination time, so that the output of a transient simulation can be viewed as a stochastic process $\{Y(s) : 0 \leq s \leq T; \Theta\}$, where $T$ is a stopping time (may be deterministic), see, e.g., [3] for a definition of stopping time. Note that this notation includes the case of a discrete-time output $Z_0, Z_1, \ldots$, if we assume that $Y(s) = Z_{\lfloor s \rfloor}$, where $\lfloor s \rfloor$ denotes the integer part of $s$.

A performance variable $W$ in transient simulation is a real-valued random variable (r.v.) that depends on the simulation output up to time $T$, i.e., $W = f(Y(s), 0 \leq s \geq T; \Theta)$, and the expectation of a performance variable $W$ is a performance measure that we usually estimate through experimentation with the simulation model. When there is no uncertainty in the parameters of the random components, the standard methodology that is used to estimate a performance measure in transient simulation is the method of independent

replications, that consists on running the simulation model to produce $n$ replications $W_1, W_2, \ldots, W_n$ that can be regarded as independent and identically distributed (i.i.d.) random variables (see Figure 1) .

For $i = 1$ to $n$

    1. Run a simulation experiment with $\Theta = \theta$ to obtain an independent replication

$$W_i = g(Y_i(s), 0 \le s \le T; \theta)$$

End Loop

Compute a point estimator and an ACI for the expectation $\alpha$: $\hat{\alpha}(n) = \frac{1}{n}\sum_{i=1}^n W_i$.

**Figure 1.** Algorithm for the method of independent replications with parameter fixed at the value $\theta$.

Under the method of independent replications, a point estimator for the expectation $\alpha = E[W_1]$ is the average $\hat{\alpha}(n) = \frac{\sum_{i=1}^n W_i}{n}$. If $E[|W_1|] < \infty$, it follows from the classical Law of Large Numbers (LLN), that $\hat{\alpha}(n)$ is consistent, i.e., it satisfies $\hat{\alpha}(n) \Rightarrow \alpha$ , as $n \to \infty$ (where $\Rightarrow$ denotes weak convergence of random variables), see, e.g., [3] for a proof. Consistency guarantees that the estimator approaches the parameter as the number of replications $n$ increases, and the accuracy of the simulation-based estimator $\hat{\alpha}(n)$ is typically assessed by an asymptotic confidence interval (ACI) for the parameter. The expression for an ACI for a parameter of a stochastic simulation is usually obtained through a Central Limit Theorem (CLT) for the estimator (see, for example, chapter 3 of [4]). For the case of the expectation $\alpha$ in the algorithm of Figure 1, if $E[W_1^2] < \infty$, the classical CLT implies that

$$\frac{\sqrt{n}(\hat{\alpha}(n) - \alpha)}{\sigma} \Rightarrow N(0,1), \qquad (1)$$

as $n \to \infty$, where $\sigma^2 = E[(W_1 - \alpha)^2]$ and $N(0,1)$ denotes a r.v. distributed as normal with mean 0 and variance 1. Then, if $E[W_1^2] < \infty$, it follows from (1) and Slutsky's Theorem (see the Appendix) that

$$\frac{\sqrt{n}(\hat{\alpha}(n) - \alpha)}{\hat{\sigma}(n)} \Rightarrow N(0,1),$$

as $n \to \infty$, where $\hat{\sigma}(n)$ denotes the sample standard deviation, i.e., $\hat{\sigma}^2(n) = \frac{\sum_{i=1}^n (W_i - \hat{\alpha}(n))^2}{n-1}$. This CLT implies that

$$\lim_{n \to \infty} P[|\hat{\alpha}(n) - \alpha| \le z_\beta \hat{\sigma}(n)/\sqrt{n}] = 1 - \beta,$$

for $0 < \beta < 1$, where $z_\beta$ denotes the $(1 - \beta/2)$-quantile of a N(0,1), which is sufficient to establish a $(1 - \beta)100\%$ ACI for $\alpha$ with halfwidth

$$HW_\alpha = z_\beta \hat{\sigma}(n)/\sqrt{n}. \qquad (2)$$

A halfwidth in the form of (2) is the typical measure used in simulation software (e.g., Simio, see [2]) to assess the accuracy of $\hat{\alpha}(n)$ for the estimation of expectation $\alpha$.

In contrast to the estimation of (output) performance measures, parameters of (input) random components of a simulation model are usually estimated from real-data observations ($x$) and, while most applications covered in the relevant literature assume that no uncertainty exists in the value of these parameters, the uncertainty can be significant when little data is available. In these cases, Bayesian statistics can be used to incorporate this uncertainty in the output analysis of simulation experiments via the use of a posterior distribution $p(\theta|x)$. A methodology currently proposed for the analysis of simulation

For $i = 1$ to $n$

   Generate (independently) a value $\theta_i$ from $p(\theta|x)$

   For $j = 1$ to $m$:

        Run a simulation experiment with $\Theta = \theta_i$ to obtain an independent replication

$$W_{ij} = g\left(Y_{ij}(s), 0 \leq s \leq T; \theta_i\right)$$

   End Loop

End Loop

Compute a point estimator and an ACI for the expectation $\alpha$: $\hat{\alpha}(n) = \frac{1}{n}\sum_{i=1}^{n}\hat{\alpha}_i$, $\hat{\alpha}_i = \frac{\sum_{j=1}^{m}W_{ij}}{m}$

**Figure 2.** Two-level algorithm for calculating a point estimator using stochastic simulation under parameter uncertainty.

experiments under parameter uncertainty, is a two-level nested simulation algorithm (see, e.g., [6–8]. In the outer level, we simulate ($n$) observations for the parameters from a posterior distribution $p(\theta|x)$, while in the inner level we simulate ($m$) observations for the performance variable with the parameters fixed at the value ($\theta$) generated in the outer level (see Figure 2). In this paper, we focus on the output analysis of two-level simulation experiments, for the case where the observations at the inner level are independent, showing how the variance of a simulated observation can be decomposed into parametric and stochastic variance components. Afterwards, we derive a CLT for both the estimator of the point forecast and the estimators of the variance components. Our CLTs allow us to compute an ACI for each estimator. Our results are validated through experiments with a forecast model for sporadic demand reported in [10]. This paper is an extended version of results initially reported in [11] and the missing proofs in [11] are provided.

Following this introduction, we present the proposed methodology for the construction of an ACI for the point forecast and the variance components in a two-level simulation experiment. Afterwards, we present an illustrative example that has an analytical solution for the parameters of interest in this paper. This example is used in the next section to illustrate the application and validity of our proposed methodologies for the construction of an ACI. Finally, in the last section, we present conclusions and directions for future research.

## 2. Theoretical Results

To identify the variance components in each observation $W_{ij}$ of the algorithm illustrated in Figure 2, let $\mu(\Theta) = E[W_{11}|\Theta]$, and $\sigma^2(\Theta) = E[W_{11}^2|\Theta] - \mu^2(\Theta)$. Under this notation, the point forecast is $\alpha = E[\mu(\Theta)]$, and the variance of each $W_{ij}$ is:

$$V[W_{ij}] \stackrel{def}{=} E[W_{ij}^2] - E[W_{ij}]^2 = E[E[W_{ij}^2|\Theta] - \mu(\Theta)^2] + E[\mu(\Theta)^2] - E[\mu(\Theta)]^2 = \sigma_S^2 + \sigma_P^2, \quad (3)$$

for $i = 1, ..., n$; $j = 1, ..., m$, where $\sigma_P^2 = V[\mu(\Theta)] \stackrel{def}{=} E[\mu(\Theta)^2] - E[\mu(\Theta)]^2$, and $\sigma_S^2 = E[\sigma^2(\Theta)]$. It is worth mentioning that, in the relevant literature, $\sigma_S^2$ is commonly referred to as stochastic variance and $\sigma_P^2$ is commonly referred to as parametric variance.

### 2.1. Point estimators

In this paper, we are interested in both the estimation of the point forecast $\alpha = E[\mu(\Theta)]$ and the estimators of the variance components of every observations generated in the algorithm of Figure 2 and defined in (3), thus we first consider the natural point estimators

$$\hat{\alpha}(n) = \frac{1}{n}\sum_{i=1}^{n}\hat{\alpha}_i, \hat{\sigma}_T^2(n) = \frac{1}{n-1}\sum_{i=1}^{n}(\hat{\alpha}_i - \hat{\alpha}(n))^2, \hat{\sigma}_S^2(n) = \frac{1}{n}\sum_{i=1}^{n}S_i^2, \quad (4)$$

where $\hat{\alpha}_i = m^{-1} \sum_{j=1}^{m} W_{ij}$, and $S_i^2 = (m-1)^{-1} \sum_{j=1}^{m} (W_{ij} - \hat{\alpha}_i)^2$, $i = 1, ..., m$. Note that the $\hat{\alpha}_i$'s are i.i.d. with expectation $E[\hat{\alpha}_1] = \alpha$ and variance

$$
\begin{aligned}
\sigma_T^2 &\overset{def}{=} E[(\hat{\alpha}_1 - \alpha)^2] = m^{-2}(mE[(W_{11} - \alpha)^2] + m(m-1)E[(W_{11} - \alpha)(W_{12} - \alpha)]) \\
&= m^{-1}(\sigma_S^2 + \sigma_P^2) + m^{-1}(m-1)\sigma_S^2 = \sigma_S^2 + m^{-1}\sigma_P^2.
\end{aligned}
\tag{5}
$$

On the other hand, the $S_i^2$ are i.i.d. with expectation $E[S_1^2] = \sigma_S^2$. Thus, the next proposition follows from the classical LLN .

**Proposition 1.** *Given $m \geq 1$, if $E[W_{11}^2] < \infty$ then $\hat{\alpha}(n)$ and $\hat{\sigma}_T^2(n)$ are unbiased and consistent (as $n \to \infty$) estimators for $\alpha$ and $\sigma_T^2$ (as defined in (5)), respectively. Furthermore, if $m \geq 2$ and $E[W_{11}^2] < \infty$, then $\hat{\sigma}_S^2(n)$ is an unbiased and consistent (as $n \to \infty$) estimator for $\sigma_S^2$ (as defined in (3)).*

*2.2. Accuracy of the point estimators*

As we established in Proposition 1, under mild assumptions the point estimators proposed in (4) are consistent, and thus converge to the corresponding parameter value (as $n \to \infty$). Nonetheless, to establish the level of accuracy of these estimators, we must establish a CLT for each estimator to derive a valid expression for the corresponding ACI. Note that both $\hat{\alpha}(n)$ and $\hat{\sigma}_S^2(n)$ are averages of i.i.d observations, thus the next proposition follows from the classical CLT for i.i.d. observations.

**Proposition 2.** *Given $m \geq 1$, if $E[W_{11}^2] < \infty$ then*

$$
\frac{\sqrt{n}(\hat{\alpha}(n) - \alpha)}{\sigma_T} \Rightarrow N(0,1),
$$

*as $n \to \infty$. Furthermore, if $m \geq 2$ and $E[W_{11}^4] < \infty$, then*

$$
\frac{\sqrt{n}(\hat{\sigma}_S^2(n) - \sigma_S^2)}{\sqrt{V_S}} \Rightarrow N(0,1),
$$

*as $n \to \infty$, where $\sigma_S^2$ is defined in (3), $\sigma_T^2$ is defined in (5), $\hat{\alpha}(n)$, $\hat{\sigma}_S^2(n)$, $S_1^2$ are defined in (4), and $V_S = E[(S_1^2 - \sigma_S^2)^2]$.*

Since we have consistent estimators for $\sigma_T^2$ and $V_S$ (under mild assumptions), the next corollary follows from Proposition 1 and Slutsky's Theorem, details of a proof are given in the Appendix.

**Corollary 1.** *Under the same notation and assumptions as in Proposition 2, for $m \geq 1$ we have*

$$
\frac{\sqrt{n}(\hat{\alpha}(n) - \alpha)}{\sqrt{\hat{\sigma}_T^2(n)}} \Rightarrow N(0,1),
$$

*as $n \to \infty$, and for $m \geq 2$ we have*

$$
\frac{\sqrt{n}(\hat{\sigma}_S^2(n) - \sigma_S^2)}{\sqrt{\hat{V}_S(n)}} \Rightarrow N(0,1),
$$

*as $n \to \infty$, where $\hat{\sigma}_S^2(n)$ and $\hat{\sigma}_T^2(n)$ are defined in (4), and*

$$
\hat{V}_s(n) = \frac{1}{n-1} \sum_{i=1}^{n} (S_i^2 - S^2)^2, \quad S^2 = \frac{1}{n} \sum_{i=1}^{n} S_i^2
$$

In order to obtain a CLT for $\hat{\sigma}_T^2(n)$, note that this estimator is the sample variance of a set of i.i.d. observations, thus we can use the following Lemma. A proof using the Delta Method (see, e.g., Proposition 2 of [9] for a proof) is provided in the Appendix.

**Lemma 1.** *If $X_1, X_2, \ldots$ is a sequence of i.i.d. random variables with $E[X_1^4] < \infty$, then*

$$\frac{\sqrt{n}(S^2(n) - \sigma_1^2)}{\sqrt{\sigma_2^2}} \Rightarrow N(0,1),$$

*as $n \to \infty$, where $\sigma_1^2 = \mu_2 - \mu_1^2$, $\sigma_2^2 = \mu_1^2 \mu_2 - 4\mu_1^4 - 4\mu_1\mu_3 + \mu_4 - \mu_2^2$, $\mu_k = E[X_1^k]$, $k = 1,2,3,4$; $S^2(n) = (n-1)^{-1}\sum_{i=1}^n (X_i - \hat{\mu}_1)^2$, $\hat{\mu}_1 = n^{-1}\sum_{i=1}^n X_i$.*

**Corollary 2.** *Under the same assumptions as in Lemma 1 we have*

$$\frac{\sqrt{n}(S^2(n) - \sigma_S^2)}{\sqrt{\hat{\sigma}_2^2(n)}} \Rightarrow N(0,1),$$

*as $n \to \infty$, where $\hat{\sigma}_2^2 = 8\hat{\mu}_1^2\hat{\mu}_2 - 4\hat{\mu}_1^4 - 4\hat{\mu}_1\hat{\mu}_3 + \hat{\mu}_4 - \hat{\mu}_2^2$, $\hat{\mu}_k = n^{-1}\sum_{i=1}^n X_i^k$.*

Corollary 2 follows from the fact that $\hat{\mu}_k$ is an unbiased and consistent estimator of $\mu_k$, and the next corollary follows from the fact that $\hat{\sigma}_T^2(n)$ is the sample variance of the $\hat{\alpha}_i$.

**Corollary 3.** *Given $m \geq 1$, if $E[W_{11}^4] < \infty$ then*

$$\frac{\sqrt{n}(\hat{\sigma}_T^2(n) - \sigma_T^2)}{\sqrt{\hat{V}_T(n)}} \Rightarrow N(0,1),$$

*as $n \to \infty$, where $\hat{V}_T(n) = 8\bar{\alpha}_1^2\bar{\alpha}_2 - 4\bar{\alpha}_1^4 - 4\bar{\alpha}_1\bar{\alpha}_3 + \bar{\alpha}_4 - \bar{\alpha}_2^2$, $\bar{\alpha}_k = n^{-1}\sum_{i=1}^n \hat{\alpha}_i^k$.*

Let $0 < \beta < 1$, and using corollaries 1 and 3 we can establish a $100(1-\beta)\%$ ACI for the point forecast $\alpha$, and variance components $\sigma_S^2$ and $\sigma_T^2 = \sigma_S^2 + m^{-1}\sigma_P^2$; each ACI is centered in the corresponding point estimator ($\hat{\alpha}(n)$, $\hat{\sigma}_S^2(n)$ or $\hat{\sigma}_T^2(n)$) and the corresponding halfwidth is given by:

$$HW_\alpha = z_\beta \frac{\sqrt{\hat{\sigma}_T^2(n)}}{\sqrt{n}}, HW_{\sigma_S^2} = z_\beta \frac{\sqrt{\hat{V}_S(n)}}{\sqrt{n}}, and\ HW_{\sigma_T^2} = z_\beta \frac{\sqrt{\hat{V}_T(n)}}{\sqrt{n}}, \tag{6}$$

for $\alpha$, $\sigma_2^2$ and $\sigma_T^2$, respectively, where $\hat{\sigma}_T^2(n)$ is defined in (4), $\hat{V}_S(n)$ and $\hat{V}_T(n)$ are defined in Corollary 1, and in Corollary 3, respectively.

Note that the ACIs proposed in (6) assume that the value of $m$ in the algorithm of Figure 2 is fixed and the accuracy of the estimator improves as $n$ (the number of observations in the outer level) increases (in turn, the halfwidth of the ACI gets smaller). Given that we can build a valid ACI for any value of $m$, a relevant question is how to find an adequate value of $m$ to get an acceptable level of accuracy in a reasonable amount of running time. In order to answer this question for the case of the point estimator of $\alpha$, let us fix the total number of iterations in the algorithm of Figure 2 to $k = nm$, and note from (5) and Proposition 2 that the asymptotic variance of $(\hat{\alpha}(n) - \alpha)$ is

$$n^{-1}\sigma_T^2 = k^{-1}(m\sigma_S^2 + \sigma_P^2), \tag{7}$$

and takes its minimal value for $m = 1$, suggesting that the point estimator $\hat{\alpha}(n)$ defined in (4) is more accurate as $m$ approaches the value of 1. Note that for $m = 1$, a fixed number of iterations $k = nm$ is convenient (from the point of view of running time), when the

computation of $W_{ij}$ requires the same or more computation time as $\Theta_i$, as suggested in the relevant literature (see, for example, [6]). Furthermore, if we allow $m$ to increase with $n$, we can obtain the following proposition (a proof using Lindeberg-Feller Theorem is provided in the Appendix).

**Proposition 3.** *Given $0 < p \leq 1$, if $m = \left\lfloor n^{-1+1/p} \right\rfloor$ and $E[W_{11}^2] < \infty$ then*

$$\frac{\sqrt{n}(\hat{\alpha}(n) - \alpha)}{\sqrt{\sigma_T^2}} \Rightarrow N(0,1),$$

*as $n \to \infty$, where $\sigma_T^2$ is defined in (5).*

Note that the last proposition implies that the ACI defined in equation (6) for the point forecast $\alpha$ is also valid under the assumptions of Proposition 3. If, once again, we set the total number of iterations in the algorithm of Figure 2 to $k = nm$, we let $n \approx k^p$, $m \approx k^{1-p}$, and $nm = k$, it follows from Proposition 3 that the asymptotic variance of $\hat{\alpha}(n)$ is $n^{-1}\sigma_T^2 \approx k^{-p}(k^{1-p}\sigma_S^2 + \sigma_P^2)$ for every $0 \leq p \leq 1$. Note that, for fixed $k$, $n^{-1}\sigma_T^2$ reaches its minimum value when $p = 1$, that is, when $n = k$ and $m = 1$. However, note that we need $m \geq 2$ in order to estimate $\sigma_S^2$. In the following section we report some empirical results that confirm our theoretical results. It is worth mentioning that the case $n = k$ and $m = 1$ has been reported in the literature as the posterior sampling algorithm (see, e.g., [12,13])

### 3. An Example with Analytical Solution

The following model (reported in [10]) has been proposed to forecast sporadic demand by incorporating data on times between arrivals and customer demand; where uncertainty on the model parameters is incorporated using a Bayesian approach. For this model, we will show analytical expressions for the performance measures defined in Section 2. These expressions are used in the following section to illustrate the validity of the ACIs proposed in the previous section.

Customer arrivals for a particular item in a shop follow a Poisson process, yet there is uncertainty in the arrival rate $\Theta_0$, so that given $[\Theta_0 = \theta_0]$, interarrival times between customers are i.i.d. with exponential density:

$$f(y|\theta_0) = \begin{cases} \theta_0 e^{-\theta_0 y}, & y > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{8}$$

where $\theta_0 \in S_{00} = (0, \infty)$. Every client can order $j$ units of this item with probability $\Theta_{1j}$, $j = 1, ..., q$, $q \geq 2$. Let $\Theta_1 = (\Theta_{11}, ... \Theta_{1(q-1)})$ and $\Theta_{1q} = 1 - \sum_{j=1}^{q-1} \Theta_{1j}$, then $\Theta = (\Theta_0, \Theta_1)$ is the parameter vector, and $S_0 = S_{00} \otimes S_{01}$ is the parameter space, where $S_{01} = \{(\theta_{11}, ..., \theta_{1(q-1)}) : \sum_{j=1}^{q-1} \theta_{1j} \leq 1; \theta_{1j} \geq 0, j = 1, ..., q - 1\}$.

Total demand during a period of length $T$ is

$$D = \begin{cases} \sum_{i=1}^{N(T)} U_i, & N(T) > 0 \\ 0, & \text{otherwise,} \end{cases} \tag{9}$$

where $N(s)$ is the number of customer arrivals during the interval $[0, s]$, $s \geq 0$, and $U_1, U_2, ...$ are the individual demands (conditionally independent relative to $\Theta$). The information about $\Theta$ consists of i.i.d. observations $v = (v_1, ..., v_r)$, $u = (u_1, ..., u_r)$ of past customers, where $v_i$ is the interarrival time between customer $i$ and customer $(i - 1)$, and $u_i$ is the number of units ordered by client $i$. By taking Jeffrey's non-informative prior as the prior

density for $\Theta$, we obtain the posterior density (see [10] for details) $p(\theta|x) = p(\theta_0|v)p(\theta_1|u)$, where $x_i = (v_i, u_i), i = 1, ..., r$ , $x = (x_1, ..., x_r)$, $\theta = (\theta_0, \theta_1)$ as

$$p(\theta_0|v) = \frac{\theta_0^{r-1}(\sum_{i=1}^{r} v_i)^r e^{-\theta_0 \sum_{i=1}^{r} v_i}}{(r-1)!}, p(\theta_1|u) = \frac{(1 - \sum_{j=1}^{q-1} \theta_{1j})^{c_q - 1/2} \Pi_{j=1}^{q-1} \theta_{1j}^{c_j - 1/2}}{B(c_1 + 1/2, ..., c_q + 1/2)}, \quad (10)$$

where $c_j = \sum_{i=1}^{r} I[u_i = j]$, and $B(a_1, ...a_q) = \Pi_{j=1}^{q} \Gamma(a_j)/\Gamma(\sum_{j=1}^{q} a_j)$, for $a_1, ..., a_q > 0$. Using this notation, we can show that (see [1] for details)

$$\alpha = E[T\Theta_0] \sum_{j=1}^{q} j p_j,$$

$$\sigma_P^2 = \frac{E[T^2\Theta_0^2]}{(q_0 + 1)} \sum_{j=1}^{q} j^2 p_j + \frac{E[T\Theta_0]^2[(q_0/n) - 1]}{(q_0 + 1)} (\sum_{j=1}^{q} j p_j)^2,$$

$$\sigma_S^2 = E[T\Theta_0] \sum_{j=1}^{q} j^2 p_j,$$

where $E[T\Theta_0] = Tr(\sum_{i=1}^{r} v_i)^{-1}$, $E[T^2\Theta_0^2] = T^2 r(1 + r)(\sum_{i=1}^{r} v_i)^{-2}$, $p_j = q_j/q_0$, $q_j = c_j + 1/2, j = 1, ..., q$, $q_0 = \sum_{j=1}^{q} q_j$, and $c_j$ are defined in (10).

## 4. Empirical Results

To validate the ACIs proposed in (4), we conducted some experiments with the Bayesian model of the previous section to illustrate the estimation of $\alpha$, $\sigma_S^2$ and $\sigma_T^2$. We considered the values $T = 15$, $r = 20$, $\sum_{i=1}^{r} x_i = 10$, $q = 5$, $c_1 = 5$, $c_2 = 3$, $c_3 = 2$, $c_4 = 3$, $c_5 = 7$. With this data, the point forecast is $\alpha \approx 95.333$, and the variance components are $\sigma_S^2 \approx 380.667$, $\sigma_P^2 \approx 568.598$. The empirical results that we report below illustrate a typical behavior that we should experiment for any other feasible data set.
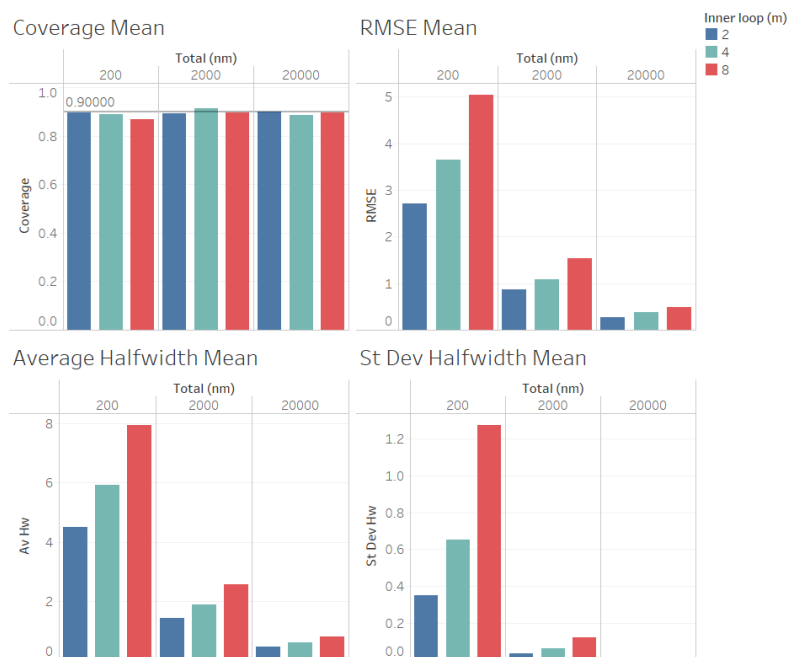


**Figure 3.** Performance of the estimation of point forecast $\alpha$ for *nm* fixed comparing different values of *m*.

In all the experiments reported in this Section we considered 1000 independent replications of the algorithm of Figure 2 for different number of observations in the outer level ($n$) and in the inner level ($m$); in each replication we computed the point estimators for $\alpha$, $\sigma_S^2$, and $\sigma_T^2$, and the corresponding halfwidths of 90% ACI's according to equations (6). Since we know the value of the parameters we are estimating, we were able to report (for $n$ and $m$ given), the empirical coverage (i.e., the fraction of independent replications that the corresponding ACI covered the true parameter value), the average and standard deviation of halfwidths, and the squared root of the empirical mean squared error defined by

$$RMSE = \sqrt{\frac{1}{n_0}\sum_{i=1}^{n_0}(\hat{\theta}_i - \theta)^2},$$

where $\hat{\theta}_i$ denotes the value obtained in the $i$-th replication for the estimation of a parameter $\theta$, $i = 1, 2, \ldots, n_0$ ($n_0 = 1000$ in our experiments).

In a first set of experiments we considered $nm = 200, 2000, 20000$, and $m = 2, 4, 8$ for each value of $nm$, to compare the effect of increasing the number of observations in the inner level for a given value value of $nm$. The results of this set of experiments are summarized in figures 3, 4 and 5. Note that we are not considering $m = 1$ in this set of experiments to be able to construct an ACI for the stochastic variance $\sigma_S^2$.
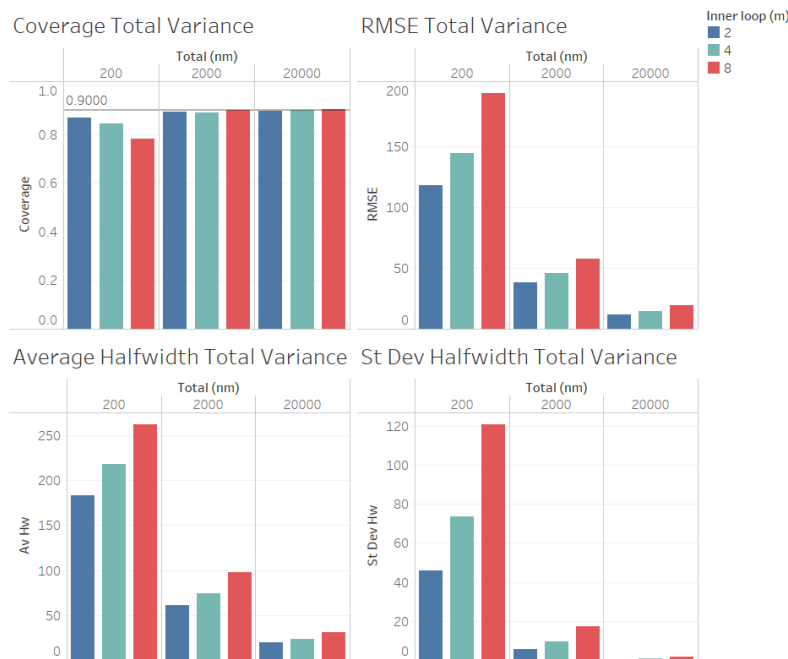
In Figure 3 we illustrate the performance measures for the quality of the estimation procedure that we obtained for the estimation of the point forecast $\alpha$. As we observe from Figure 3, the coverages are acceptable (very close to the nominal value of 0.9, even for $n = 100$). These results validate the ACI defined in (6) for the point forecast $\alpha$. We also observe from Figure 3 that the RMSE, average halfwidth and standard deviation of halwidths improve (decrease) as the number of observations in the outer level ($n$) increases, as suggested by Corollary 1. Note also from Figure 3 that a smaller value of $m$ provides smaller RMSE, average halfwidths and standard deviations of halwidths, validating our theoretical results.
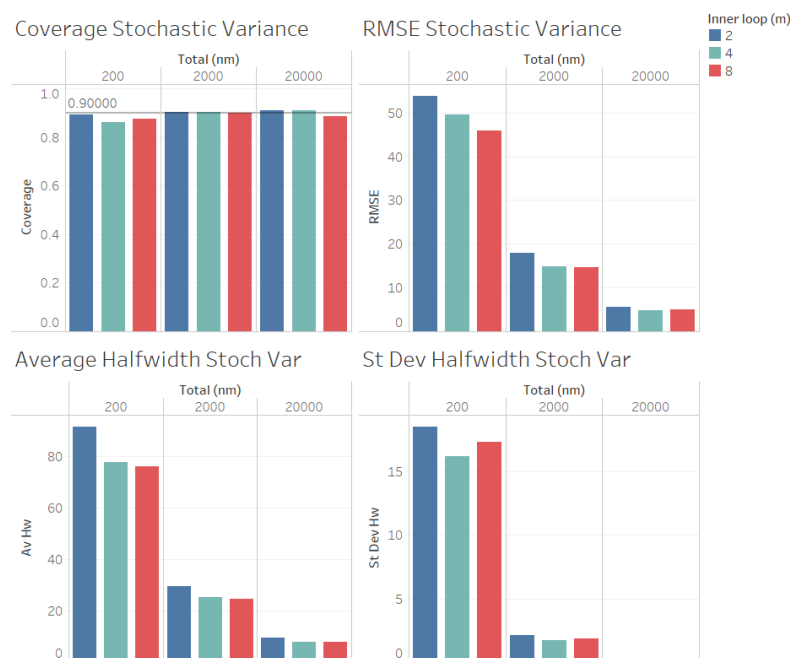


**Figure 4.** Performance of the estimation of stochastic variance $\sigma_S^2$ for $nm$ fixed comparing different values of $m$.

In Figure 4 we illustrate the performance measures for the quality of the estimation procedure that we obtained for the estimation of the stochastic variance $\sigma_S^2$. As we observe

from Figure 4, the coverages are acceptable (very close to the nominal value of 0.9, even for $n = 100$). These results validate the ACI defined in (6) for the stochastic variance $\sigma_S^2$. We also observe from Figure 3 that the RMSE, average halfwidth and standard deviation of halwidths improve (decrease) as the number of observations in the outer level ($n$) increases, as suggested by Corollary 2. However, contrary to what we observe for the estimation of $\alpha$, a larger value of $m$ provides smaller RMSE, average halfwidths and standard deviations of halwidths, suggesting that, for a fixed value of $nm$, the quality of the estimation for the stochastic variance $\sigma_S^2$ improves as the number of the observations in the inner loop ($m$) increases.
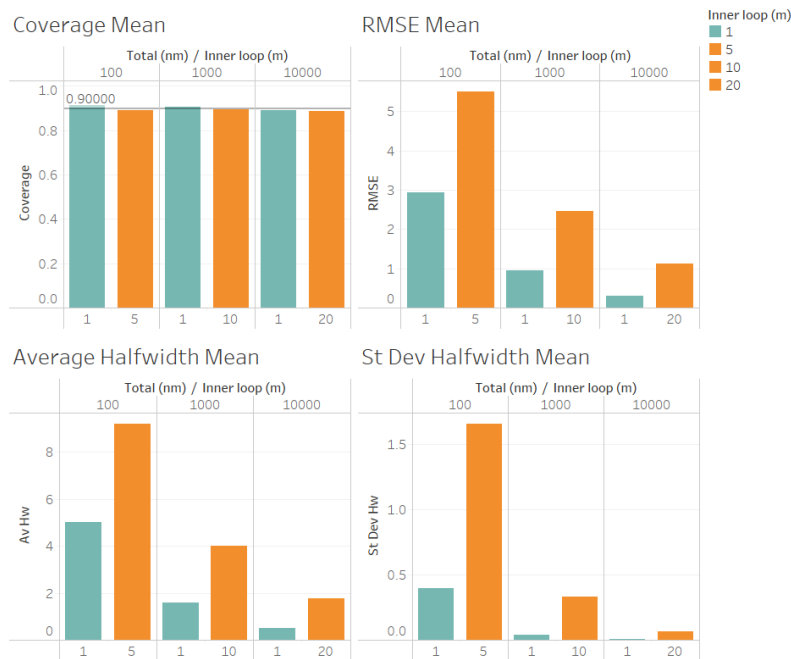


**Figure 5.** Performance of the estimation of total variance $\sigma_T^2$ for $nm$ fixed comparing different values of $m$.

For the estimation of the total variance $\sigma_T^2$ (illustrated n Figure 5) we obtained similar results for the quality of the estimation as for the estimation of the point forecast $\alpha$, except that a larger values of $n$ is required to obtain reliable coverages. As we observe from Figure 3, the coverages are acceptable (very close to the nominal value of 0.9, for $n = 1000$ and 10000). These results validate the ACI defined in (6) for the total variance $\sigma_T^2$. We also observe from Figure 3 that the RMSE, average halfwidth and standard deviation of halfwidths improve (decrease) as the number of observations in the outer level ($n$) increases, as suggested by Corollary 3. Note also from Figure 5 that a smaller value of $m$ provides smaller RMSE, average halfwidths and standard deviations of halwidths, validating our theoretical results.

In a second set of experiments we considered $nm = 100, 1000, 10000$, with $m = 1$ and $m \approx (nm)^{1/3}$ for each value of $nm$, to compare the quality of the estimation procedures using the value of $m$ that we suggest as optimal for the estimation of point forecast $\alpha$ with the value of $m$ suggested in [6] as an adequate choice for $m$ in the case of biased estimators in the inner level of the algorithm of Figure 2. The results of this set of experiments are summarized in figures 6 and 7. Note that we are not considering the estimation of the stochastic variance $\sigma_S^2$ in this set of experiments because $m \geq 2$ is required to construct an ACI for the stochastic variance $\sigma_S^2$. Note also that we considered $100^{1/3} \approx 5$, $1000^{1/3} \approx 10$, and $10000^{1/3} \approx 20$, and we are using the same color for $m = 5, 10, 20$ in figures 6 and 7.

In Figure 6 we illustrate the performance measures for the quality of the estimation procedure that we obtained for the estimation of the point forecast $\alpha$ in our second set of

**Figure 6.** Performance of the estimation of point forecast $\alpha$ for $nm$ fixed comparing $m = 1$ and $m \approx (nm)^{1/3}$.

experiments . As we observe from Figure 6, the coverages are acceptable (very close to the nominal value of 0.9, even for $n = 100$). These results validate the ACI defined in (6) for the point forecast $\alpha$, and the ACI suggested by Proposition 3. We also observe from Figure 6 that the RMSE, average halfwidth and standard deviation of halwidths are worse for $m \approx (nm)^{1/3}$, confirming our finding that, for the same number of replications $nm$, $m = 1$ produces better point estimators for $\alpha$ than $m \approx (nm)^{1/3}$ confirming the result of Proposition 3.
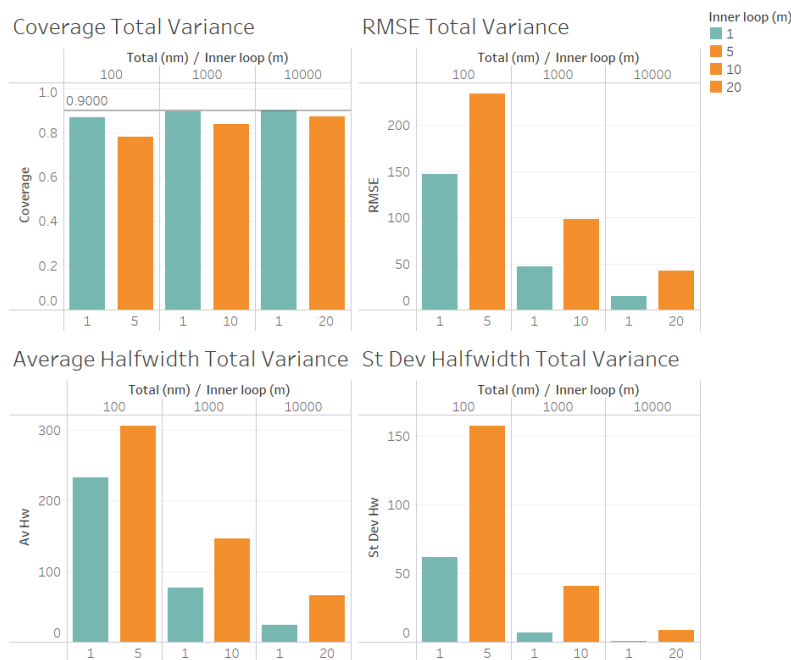


**Figure 7.** Performance of the estimation of total variance $\sigma_T^2$ for $nm$ fixed comparing $m = 1$ and $m \approx (nm)^{1/3}$.

Finally, in Figure 7 we show the results of our second set of experiments for the estimation of the total variance $\sigma_T^2$. We found similar resulta as for the case of the estimation of the point forecast $\alpha$, coverages are very good (even for $n = 100$), and all performance measure for the ACI (RMSE, average and standard deviation of halfwidths) are worse for $m \approx (nm)^{1/3}$, suggesting that, for the same number of replications $nm$, $m = 1$ produces better point estimators for $\sigma_T^2$ than $m \approx (nm)^{1/3}$.

## 5. Conclusions

In this paper, we propose methodologies to calculate point estimators (and their corresponding halfwidths), for both the point forecast and the variance components in two-level nested stochastic simulation experiments, for the case where the observations at both levels of the algorithm are independent. These methods can be applied to the construction of Bayesian forecasts based on experiments using a simulation model under parameter uncertainty.

Both our theoretical and our experimental results confirm that the proposed point estimators and their corresponding halfwidths are asymptotically valid, i.e., the point estimators converge to the corresponding parameter values and the halfwidths converge to the nominal coverage as the number of replications ($n$) of the outer level increases.

Furthermore, given a fixed number of total observations ($nm$), we show that the choice of only one replication in the inner level ($m = 1$) provides more accurate estimators for both the point forecast ($\alpha$), and the variance of the point forecast ($\sigma_T^2$). However, $m \geq 2$ is required for the estimation of $\sigma_S^2$.

Directions for future research on this topic includes experimentation with other point estimators, such as, quasi Monte Carlo or Simpson integration, with the objective of finding more accurate point estimators for the parameters considered in this paper.

**Data Availability Statement:** The raw data corresponding to our experiments are available in the repository of *AppliedMath*.

**Conflicts of Interest:** The Author declare no conflict of interest.

## Appendix A

For completeness, we first write three well known theorems. Proofs of Theorem A1 and Theorem A2 can be found, e.g., in [5], and a proof of Theorem A3 can be found, e.g., in [9]. In what follows, we write $\Rightarrow$ for weak converge (as $n \to \infty$ without explicit mention).

**Theorem A1.** *(Slutsky). Let $X, Y, X_1, X_2, ..., Y_1, Y_2, ...$ be random variables and c be a real constant. If $X_n \Rightarrow X$, and $Y_n \Rightarrow c$, then:*

  *(i)*    $X_n + Y_n \Rightarrow X + c$
  *(ii)*   $X_n Y_n \Rightarrow cX$
  *(iii)*  $X_n / Y_n \Rightarrow X/c$, *if $c \neq 0$*

**Theorem A2.** *(Continuous mapping). Let $X, X_1, X_2, ...$ be $\Re^k$-valued random vectors, and let $g : \Re^k \to \Re$ be a function such that $P[X \in D(g)] = 0$, where $D(g) = \{x : g(x)$ is not continuous at $x\}$, then $g(X_n) \Rightarrow g(X)$.*

**Theorem A3.** *(Delta method). Let $Y_1, Y_2, ...$ be $\Re^k$-valued random vectors, and let $g:\mathbb{R}^k \to \Re$ be a function that is differentiable in a vecinity of $\mu \in \Re^k$. If there exists a $k \times k$ matrix G such that the TLC*

$$\sqrt{n}[\bar{Y}(n) - \mu] \Rightarrow GN_k(0,1)$$

is satisfied, where $\bar{Y}(n) = n^{-1}\sum_{i=1}^{m} Y_i$, and $N_k(0, I)$ denotes a (k -variate) normal distribution with mean 0 and variance I (the identity), then

$$\sqrt{n}[g(\bar{Y}(n)) - g(\mu)] \Rightarrow \sigma N(0, 1),$$

where $\sigma = \sqrt{\nabla g(\mu)^T GG^T \nabla g(\mu)}$.

**Proof of Corollary 1.** Since $\hat{\alpha}_1, \hat{\alpha}_2, ...$ are i.i.d. with $E[\hat{\alpha}_1^2] < \infty$, it follows from the Law of Large Numbers that $n^{-1}(\sum_{i=1}^{n} \hat{\alpha}_i^2, \sum_{i=1}^{n} \hat{\alpha}_i) \Rightarrow (E[\hat{\alpha}_1^2], E[\hat{\alpha}_1])$. Therefore, by taking $g(x_1, x_2) = \sqrt{x_1 - x_2^2}$, for $x_1 - x_2^2 \geq 0$ in Theorem A2, we have $\sqrt{(n-1)\hat{\sigma}_T^2(n)/n} \Rightarrow \sqrt{\sigma_T^2}$, so that $Y_n = \sqrt{\hat{\sigma}_T^2(n)}/\sqrt{\sigma_T^2} \Rightarrow 1$. Finally, by taking $X_n = \sqrt{n}(\hat{\alpha}(n) - \alpha)/\sqrt{\sigma_T^2}$ in Theorem A1, it follows from Proposition 2 that

$$\frac{\sqrt{n}(\hat{\alpha}(n) - \alpha)}{\sqrt{\hat{\sigma}_T^2(n)}} \Rightarrow N(0, 1)$$

Similarly, since $S_1^2, S_2^2, ...$ are i.i.d. with $E[S_1^4] < \infty$, it follows from Theorem A1 and Proposition 2 that

$$\frac{\sqrt{n}(\hat{\sigma}_S^2(n) - \sigma_S^2)}{\sqrt{\hat{V}_S(n)}} \Rightarrow N(0, 1).$$

$\square$

**Proof of Lemma 1.** Let $k = 2$, $Y_i = (X_i, X_i^2)$, $\mu = (\mu_1, \mu_2)$, then the TLC of Theorem A3 is satisfied for

$$GG^T = \begin{bmatrix} \mu_2 - \mu_1^2 & \mu_3 - \mu_1\mu_2 \\ \mu_3 - \mu_1\mu_2 & \mu_4 - \mu_2^2 \end{bmatrix}$$

By taking $g(\mu) = \mu_2 - \mu_1^2 = \sigma_1^2$, we have $g(\bar{Y}(n)) = (n-1)S^2(n)/n$, $\nabla g(\mu)^T = (-2\mu_1, 1)$, and

$$\nabla f(\mu)^T GG^T \nabla f(\mu) = 8\mu_1^2\mu_2 - 4\mu_1^4 - 4\mu_1\mu_3 + \mu_4 - \mu_2^2 = \sigma_2^2.$$

Then, it follows from Theorem A3 that

$$\sqrt{n}\left[(n-1)S^2(n)/n - \sigma_1^2\right] \Rightarrow \sigma_2 N(0, 1),$$

and the final conclusion follows from Theorem A1. $\square$

**Proof of Proposition 3.** In this proof we follow the notation of Lindeberg-Feller Theorem as in Theorem 7.2.1 of [3].

For $n = 1, 2, ...$, let $m_n = \left\lfloor n^{-1+1/p} \right\rfloor$ and $\alpha_j(n) = (\sum_{i=1}^{m_n} W_{ij})/m_n, j = 1, ..., n$. Then $\alpha_1(n), \alpha_2(n), ..., \alpha_n(n)$ are independent, and for $X_{nj} = (\alpha_j(n) - \alpha)/\sqrt{n\sigma_T^2}$ we also have that $X_{n1}, X_{n2}, ...X_{nn}$ are independent.

Then if $Y_{nj} = (\alpha_j(n) - \alpha)/\sigma_T$, we have $E[Y_{nj}] = 0$ and $E[Y_{nj}^2] = 1$, so that given $\epsilon > 0$ there exists $\eta_0 > 0$ such that $\int_{|y|<\eta_0} y^2 dF y_{nj}(y) < \epsilon$.

Therefore, given $\eta > 0$, for $n \geq max\{1, (\eta_0/\eta)^2\}$ we have

$$\sum_{j=1}^{n} \int_{|x|<\eta} x^2 dF x_{nj}(x) \leq \sum_{j=1}^{n} \frac{1}{n} \int_{|y|<\eta_0} y^2 dF y_{nj}(y) < \epsilon,$$

so that (1) of Theorem 7.2.1 of [? ] is satisfied, and it follows from this Theorem that $S_n \Rightarrow N(0, 1)$, where

$$S_n = \sum_{j=1}^{n} X_{nj} = \frac{\sqrt{n}(\hat{\alpha}(n) - \alpha)}{\sqrt{\sigma_T^2}}.$$

□

## References

1.  Muñoz, D.F. Simulation output analysis for risk assessment and mitigation. In *Multi-Criteria Decision Analysis for Risk Assessment and Management*; Ren J., Ed.; Springer: Heidelberg, Germany, 2021.
2.  Smith J.S.; Sturrock, D.T. *Simio and Simulation: Modeling, Analysis, Aplications* , 6th ed.; Simio LLC: Sewickley, Pennsylvania, 2022.
3.  Chung K.L. *A Course in Probability Theory*; Academic Press: Cambridge, Massachusetts, 2001.
4.  Asmussen S.; Glynn P.W. *Stochastic Simulation Algorithms and Analysis*; Springer: Heidelberg, Germany, 2007.
5.  Serfling R.J. *Approximation Theorems of Mathematical Statistics*; John Wiley & Sons: Hoboken, New Jersey, 2009.
6.  Andradóttir, S.; Glynn, P.W. Computing bayesian means using simulation. *ACM TOMACS* **2016**, *26*, paper 10.
7.  L'Ecuyer P. Quasi-Monte Carlo methods with applications in finance. *Finance and Stochastics* **2009**, *13*, 307-349.
8.  Zouaoui F., Wilson J.R. Accounting for parameter uncertainty in simulation input modeling. *IIE Transactions* **2003**, *35*, 781-792.
9.  Muñoz, D.F.; Glynn, P.W. A batch means methodology for estimation of a nonlinear function of a steady-state mean. *Manag Sci* **1997**, *43*, 1121-1135.
10. Muñoz, D.F.; Muñoz, D.F. Bayesian forecasting of spare parts using simulation. In *Service Parts Management: Demand Forecasting and Inventory Control*; Altay N., Litteral L.A., Eds.; Springer: Heidelberg, Germany, 2011.
11. Muñoz, D.F. Estimation of expectations in two-level nested simulation experiments. In Proceedings of the Name of the 29th European Modeling and Simulation Symposium, Barcelona, Spain, 18–20 Sep 2017; 233-238.
12. Russo, D.; Van Roy, B. Learning to optimize via posterior sampling. *Mathematics of Operations Research* **2014**, *39*, 1221-1243.
13. Muñoz, D.F.; Muñoz, D.F.; Ramírez-López A. On the incorporation of parameter uncertainty for inventory management using simulation. *International Transactions in Operational Research* **2013**, *20*, 493-513.