Analyzing simulation experiments with common random numbers

Kleijnen, J.P.C.

Published in:
Management Science

Publication date:
1988

MANAGEMENT SCIENCE
Vol. 34, No. 1, January 1988
Printed in U.S.A.

ANALYZING SIMULATION EXPERIMENTS WITH COMMON RANDOM NUMBERS*

JACK P. C. KLEIJNEN
School of Business and Economics, Catholic University Brabant (Katholieke Universiteit Brabant), 5000 LE Tilburg, Netherlands

To analyze simulation runs which use the same random numbers, the blocking concept of experimental design is not needed. Instead, this paper applies a linear regression model with a nondiagonal covariance matrix. This covariance matrix does not need to have a specific pattern such as constant covariances. A simple example yields surprising results. The paper proposes a new framework for the error analysis. This framework consists of three factors (namely, common random numbers, replication, model validity), each with three levels.
(BLOCKING; VARIANCE REDUCTION; ESTIMATED GENERALIZED LEAST SQUARES; GENERAL LINEAR MODEL; ERROR ANALYSIS)

1. Introduction

Using the same random numbers is a popular variance reduction technique in simulation. Not only academic researchers have advocated common random numbers, but practitioners also apply this technique. Actually it is the only variance reduction technique that practitioners find simple and intuitively appealing. The technique is simple indeed, since all it takes—see Table 1—is to reset the random number seed to its old initial value, before executing the next run with different values for the simulation variables $\mathbf{x}$ (boldface denotes matrices including vectors; in a queuing simulation, x_1 may denote the number of servers and x_2 the server speed). We run n combinations of simulation variables or "factors". Common random numbers imply that the n responses within one column of Table 1 use the same seed. As we shall see, good statistical analysis requires that the experiment be *repeated* (with different seeds); that is, the number m_i of replications should satisfy the condition $m_i \geq 2$ $(i = 1, \ldots, n)$. We assume that *all* replications use common seeds; hence m_i is a constant m. We shall discuss the right-hand part of Table 1 in the next section.

Note. In steady-state simulations the m replicates can be interpreted in different ways. For example, we may make m (long) runs, each starting in some fixed state (such as the empty state in queuing simulations) and using m different seeds. Interpretations are more difficult in renewal analysis and other more sophisticated analyses. In terminating simulations (and most practical simulations are terminating) the interpretation is straightforward. Kleijnen (1987) discusses steady-state versus terminating simulations, renewal analysis, and so on.

A problem is that common random numbers complicate the statistical analysis of the simulation data. (Since practitioners tend to neglect that analysis, they may not be aware of any complication.) We assume that the goal of the experiment with the simulation model is to estimate the effects (say) β of Q independent variables $\mathbf{x}$. Hence $n \geq Q$ in Table 1. An efficient and effective solution of this estimation problem, is a factorial design (for example, a 2^{k-p} design); see Kleijnen (1987).

In the literature on factorial designs a classical concept is *blocking*. Originally blocking was introduced to reduce heterogeneous, uncontrollable influences in experiments with real (nonsimulated) systems. This blocking is not simple, neither conceptually nor

* Accepted by George S. Fishman; received August 1985. This paper has been with the author $9\frac{1}{2}$ months for 3 revisions.

0025-1909/88/3401/0065$01.25

TABLE 1
Data of Simulation Experiment

Factor combination (effects) $(\beta_1 \cdots \beta_Q)$	Replicated responses (seed 1)	(seed 2)	$\cdots$	(seed m)	Average response	Estimated (co)variances
$x_{11} \cdots x_{1Q}$	y_{11}	y_{12}	$\cdots$	y_{1m}	$\bar{y}_1$	$\hat{\sigma}_1^2\ \hat{\sigma}_{12} \cdots \hat{\sigma}_{1n}$
$x_{21} \cdots x_{2Q}$	y_{21}	y_{22}	$\cdots$	y_{2m}	$\bar{y}_2$	$\hat{\sigma}_2^2 \cdots \hat{\sigma}_{2n}$
$x_{i1} \cdots x_{iQ}$	y_{i1}	y_{i2}	$\cdots$	y_{im}	$\bar{y}_i$	$\hat{\sigma}_i^2 \cdots \hat{\sigma}_{in}$
$x_{n1} \cdots x_{nQ}$	y_{n1}	y_{n2}	$\cdots$	y_{nm}	$\bar{y}_n$	$\hat{\sigma}_n^2$

technically. Therefore we do not try to describe blocking in a few lines; instead we refer to the statistical literature (see John 1980, Lorenzen 1984, Peres 1981, Shoukri and Ward 1984, Steinberg and Hunter 1984, pp. 85–86). We do discuss the concept and assumptions of blocking in a simulation context; see the end of §2. In this introductory section we only mention that some authors interpreted *common* random numbers in simulation experiments as block effects (see Anderson and Sargent 1974, p. 134, Lin and Rardin 1979, pp. 1261–1262, Naylor et al. 1966, p. 324, Schatzoff 1981, pp. 853–854, Schruben 1979, pp. 239, 247–248). Other authors, however, doubted this interpretation (Kleijnen 1974/1975, p. 355, Nozari et al. 1984, Wilson 1984). This paper shows that we do not need the blocking model; instead we analyze the simulation data through a linear regression model with a nondiagonal covariance matrix. Compared to the blocking model our model is more general, and yet very simple (see §2). We give some examples, so simple that they provide surprising results (§3). These examples illustrate a new analytic framework, consisting of three factors (namely, random number seeds, replication, and model validity), each factor with three levels (§4).

We hope to stimulate a new discussion on the fundamental and practical problem of analyzing simulation experiments with common random numbers. Arnold (1981, p. 263) states (be it not in the specific context of simulation): "However, the question of which model to use is a very difficult question, and one for which there is no definitive answer at this time."

2. Common Random Numbers and Least Squares

To analyze simulation experiments with common random numbers, we propose a metamodel based on linear regression (see equation (2.2)) with a statistical submodel for the regression residuals; the submodel reflects the use of common random numbers.

To specify the submodel we observe that simulation runs which use the same seed, yield correlated responses. So the responses within the same column of Table 1 are dependent. Responses in different columns of Table 1 are independent, because they use independent seeds: if $r \neq r'$ then y_{ir} and $y_{i'r'}$ are independent. So there are m independent observations on the n-variate vector $\mathbf{y} = (y_1, \ldots, y_n)'$. If $m \geq 2$, then unbiased estimators of $\sigma_{ii'} = \text{cov}\,(y_{ir}, y_{i'r})$ are

$$\hat{\sigma}_{ii'} = \frac{\sum_{r=1}^{m} (y_{ir} - \bar{y}_i)(y_{i'r} - \bar{y}_{i'})}{m - 1} \qquad (i, i' = 1, \ldots, n). \tag{2.1}$$

The $n \times n$ elements $\hat{\sigma}_{ii'}$ of equation (2.1) define the estimator $\hat{\Omega}_y = (\hat{\sigma}_{ii'})$. Obviously common seeds imply that Ω_y and hence $\hat{\Omega}_y$ are *nondiagonal*. (We hope that all $\sigma_{ii'}$ are positive so that common seeds indeed result in the desired variance reduction.) We emphasize that we do *not* assume a specific pattern for the covariance matrix, i.e., we do

not assume *constant* covariances. Schruben and Margolin (1978, pp. 508–509), however, assume constant variances $\sigma_i^2 = \sigma^2 = 1$ and—in case of common seeds—constant covariances $\sigma_{ii'}$ equal to ρ where $\rho > 0$; also see Schruben (1979), Nozari et al. (1984), Safizadeh (1983). We shall return to Schruben and Margolin (1978) at the end of this section, after we have presented our alternative.

Next we consider the expected value of the simulation output: $E(\mathbf{y}) = E(\bar{\mathbf{y}}) = \boldsymbol{\mu}$ (Table 1 shows m vectors $\mathbf{y}$ and one vector $\bar{\mathbf{y}} = (\bar{y}_1, \ldots, \bar{y}_n)'$, each vector with n elements). We assume that the relationship between $\boldsymbol{\mu}$ and the simulation input $\mathbf{X}$ (an $n \times Q$ matrix) is *linear* in the parameters $\boldsymbol{\beta}$ (a vector with Q effects β_j), but not necessarily linear in the input $\mathbf{X}$ (see Figure 2 in §3). The actual simulation output deviates from the expected output; we assume that these *disturbances* $\mathbf{e} = (e_1, \ldots, e_n)'$ are *additive*:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}. \tag{2.2}$$

We propose two different point estimators for the effects $\boldsymbol{\beta}$, namely the *Ordinary Least Squares* (OLS) estimator

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\bar{\mathbf{y}} \tag{2.3}$$

and the *Estimated Generalized Least Squares* (EGLS) estimator

$$\hat{\hat{\boldsymbol{\beta}}} = (\mathbf{X}'\hat{\boldsymbol{\Omega}}_y^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Omega}}_y^{-1}\bar{\mathbf{y}}. \tag{2.4}$$

Simple (but tedious) linear algebra proves that it is indeed correct to replace the individual responses y_{ir} by the averaged responses $\bar{y}_i$, assuming a constant number of replications ($m_i = m$). It is possible that the random matrix $\hat{\boldsymbol{\Omega}}_y$ has no *inverse*; see the example in §3. OLS does not need this inverse; see (2.3) and (2.5).

Note. If the number of replications were not constant ($m_i \neq m$), then OLS would give more weight to factor combinations replicated more often. This weighting can be achieved by replacing $\bar{\mathbf{y}}$ in equation (2.3) with the vector with the $N = \sum m_i$ individual responses $(y_{11}, y_{12}, \ldots, y_{nm_n})'$ and replacing $\mathbf{X}$ with an $N \times Q$ matrix whose first m_1 rows are identically $(x_{11}, \ldots, x_{1Q}), \ldots$, and whose final m_n rows are $(x_{n1}, \ldots, x_{nQ})$. The EGLS estimator is still given by (2.4) which uses the averages $\bar{\mathbf{y}}$, provided we replace $\hat{\boldsymbol{\Omega}}_y$ by $\hat{\boldsymbol{\Omega}}_{\bar{y}} = (\hat{\sigma}_{ii'}/m_i)$. Also see Arnold (1981), Schmidt (1976).

Note. In realistic simulations we experienced nearly singular $\hat{\boldsymbol{\Omega}}_y$ when using common seeds. To solve this problem we may add replications, if we assume that the population covariance matrix $\boldsymbol{\Omega}_y$ is not singular. If the number of replications is given, then we may also estimate $\boldsymbol{\Omega}_y$ after deleting one or more replications (or columns in Table 1). Deleting observations leads to jackknifing, evaluated in Kleijnen et al. (1987). We can manipulate not only the estimator $\hat{\boldsymbol{\Omega}}_y$ but also the population matrix $\boldsymbol{\Omega}_y$, that is, we can delete one or more factor combinations (rows in Table 1) so that $\boldsymbol{\Omega}_y$ changes into a smaller matrix. Later on, we can use the deleted observations to validate the regression model; see Kleijnen (1983, 1987).

Note. OLS and EGLS coincide if the design is saturated ($n = Q$) or if the covariance matrix $\boldsymbol{\Omega}_y$ has a special structure; see Kleijnen (1987, p. 243, note 33).

Besides the point estimators for $\boldsymbol{\beta}$ we need *variance estimators* (standard errors) for $\hat{\boldsymbol{\beta}}$ and $\hat{\hat{\boldsymbol{\beta}}}$. Obviously we have for OLS:

$$\hat{\boldsymbol{\Omega}}_{\hat{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Omega}}_y\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}/m. \tag{2.5a}$$

Following an idea in Schruben and Margolin (1978, pp. 515–516) we can easily prove that (2.5a) is equivalent to

$$\widehat{\text{cov}}\,(\hat{\beta}_j, \hat{\beta}_{j'}) = \frac{\sum_{r=1}^{m} (\hat{\beta}_{jr} - \hat{\beta}_j)(\hat{\beta}_{j'r} - \hat{\beta}_{j'})}{(m-1)m} \qquad (j, j' = 1, \ldots, Q), \tag{2.5b}$$

where $\hat{\beta}_{jr}$ denotes the estimator of effect β_j computed from replication r:

$$\hat{\boldsymbol{\beta}}_r = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_r, \tag{2.6}$$

where $\mathbf{y}_r$ denotes the rth observation on $\mathbf{y}$; $\sum_r \hat{\beta}_{jr}/m$ is identical to $\hat{\beta}_j$ of (2.3). (2.5b) is analogous to (2.1) and will be used in §3. For GLS (with known Ω_y) we have

$$\boldsymbol{\Omega}_{\hat{\beta}} = (\mathbf{X}'\boldsymbol{\Omega}_y^{-1}\mathbf{X})^{-1}/m. \tag{2.7}$$

For EGLS we replace $\boldsymbol{\Omega}_y$ in (2.7) by its estimator $\hat{\boldsymbol{\Omega}}_y$; the resulting $\hat{\boldsymbol{\Omega}}_{\hat{\beta}}$ holds *asymptotically*; see Arnold (1981) and Schmidt (1976). And Kleijnen et al. (1985)'s Monte Carlo experiment suggests that the asymptotic covariance matrix applies if $m \geq 25$; however, their result should be used with care, since they studied EGLS without common seeds ($\sigma_{ii'} = 0$ if $i \neq i'$) albeit with heterogeneous variances ($\sigma_i^2 \neq \sigma^2$).

We emphasize that the analyst should not use the standard formula $\boldsymbol{\Omega}_{\hat{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$ since that formula holds only if the errors are independent with common variance ($\boldsymbol{\Omega}_y = \sigma^2\mathbf{I}$). The estimated covariance matrices for $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ yield *confidence intervals* for β_j based on the Student t statistic:

$$t_\nu = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}}} \qquad (j = 1, \ldots, Q) \tag{2.8}$$

where $\hat{\sigma}_{\hat{\beta}}$ is the square root of element j on the main diagonal of $\hat{\boldsymbol{\Omega}}_{\hat{\beta}}$ and ν denotes the degrees of freedom of the t statistic. Kleijnen et al. (1985) suggest to take $\nu = \min(m_i - 1) = m - 1$. Arnold (1981, p. 343) proves that this procedure is indeed correct (and he further discusses more general tests on $\boldsymbol{\beta}$ using the F statistic). For EGLS we replace $\hat{\beta}_j$ by $\tilde{\beta}_j$ in (2.8). Because its covariance matrix (based on (2.7)) holds only asymptotically, we may replace t_ν by the standard normal variable (say) z, assuming $m \geq 25$. The sensitivity of the t statistic to *nonnormality*, and the extension to alternative regression analyses (such as rank regression and jackknifing) are discussed in Kleijnen (1987) and Kleijnen et al. (1987).

Which estimation procedure should we use, *OLS or EGLS*? If the covariance matrix $\boldsymbol{\Omega}_y$ were known, then the GLS estimator $\tilde{\boldsymbol{\beta}}$ would be the Best Linear Unbiased Estimator (BLUE). Actually we must estimate the covariance matrix $\boldsymbol{\Omega}_y$. We saw that we do not know the exact properties of the resulting OLS and EGLS estimators. A general rule in science is to try different models, when analyzing a problem. Here we recommend applying both OLS and EGLS to the same simulation data, and to see if the two techniques give the same qualitative conclusions (see (2.3) combined with (2.5), and (2.4) with (2.7) where $m \geq 25$). In one case study (a Rotterdam container harbor) the two techniques did give similar conclusions: both OLS and EGLS identified the same factor as being important while all other factors were nonsignificant; see Kleijnen et al. (1979). If OLS and EGLS give different qualitative conclusions, then we may add factor combinations to the n previous combinations.

Note. For example, if the factors are quantitative, we may simulate that combination where the two estimators (OLS, EGLS) give predictors furthest apart, within the area of interest (interpolation, no extrapolation); next we select the estimator with predictor closest to the simulated response $\bar{y}_{n+1}$; see Kleijnen (1987) for a further discussion on "model discrimination".

The General Linear Model of (2.2) is indeed useful in the interpretation of simulation data, which includes validation, sensitivity analysis, and optimization. Kleijnen (1987) gives many references to applications of regression metamodels in simulation.

Schruben and Margolin (1978) were the first to fully formalize an alternative metamodel for simulation with common seeds, namely the *random blocking* model. Their model, however, is conceptually more complicated and requires more assumptions.

They assume a specific covariance pattern, namely constant variances $\sigma_i^2 = \sigma^2 = 1$ and constant covariances $\sigma_{ii'} = \rho$ resulting from common seeds. Moreover, they introduce four assumptions specific for blocking. For example, the error **e** "is decomposed into two random components", namely the block effects **b** and the "remaining unexplained portion" **e***; **b** and **e*** are uncorrelated; **b** does not depend on **x**; see Schruben and Margolin (1978, p. 513). Actually, these authors use the blocking model only to *design* the simulation experiment; to *analyze* the results they use the same formulas as we propose. They use blocking to decide which factor combinations should use the same random numbers **R** and which should use antithetic numbers $\mathbf{1} - \mathbf{R}$. Our equations (2.3) through (2.7) are equivalent to their equations (2.2) and (2.3). We observe that in the classic statistical literature (which does not discuss simulation with common seeds) the Analysis of Variance tests do change when random blocks are introduced; see the references in §1, and Arnold (1981, pp. 209–275).

Note. Actually Arnold's Repeated Measures Model is similar to our Common Seeds Model. First Arnold (1981, pp. 209–241) uses the same restrictive assumptions as Schruben and Margolin's Random Blocks Model; next Arnold (1981, pp. 342–343) uses the same assumptions as we do; also see the Mixed Model (random plus fixed effects) in Arnold (1981, 270–271). The extension to multivariate responses (each run yields several outputs, for example, average waiting time per customer and utilization percentage per server) is discussed in Arnold (1981, pp. 374–378, 388–390), Kleijnen (1974/1975, pp. 735–739 and 1987, p. 149). Tests for a specific covariance matrix (namely, $\sigma_i^2 = 1$ and $\sigma_{ii} = \rho$) are discussed in Arnold (1981, pp. 430–433) and also Kleijnen (1987, p. 245, note 36). Note that Arnold (1981, p. 209) erroneously refers to §19.8 instead of §19.7.

3. An Illustration

To illustrate the preceding section we could have chosen any simulation model. For example, Kleijnen et al. (1979) simulate a container harbor in Rotterdam, and apply the General Linear Model of (2.2); they estimate the factor effects β through OLS and Estimated Weighted Least Squares (EWLS), that is, they do not use common seeds but they do account for variance heterogeneity. The literature often uses queuing networks to illustrate simulation methodology. However, all these examples suffer from specification error, that is the error made when replacing the simulation model by the regression model. If and only if an analytical solution for the simulation model is available, then the specification error is known. This error may be a complicated function of the independent variables. In this section we illustrate the results of the preceding section through the *simplest class of examples* we can imagine in the context of this paper (specification error is absent—see Figure 1—or simple—see the quadratic effect β_2 in Figure 2); this illustration also leads to the framework of the next section. We organize the present section as follows:

(i) First-order model with additive noise and constant variances; see (3.1) and Figure 1.
(ii) Second-order (quadratic) model; see Figure 2.
(iii) EGLS estimation of (i); see (3.2) through (3.5).
(iv) Model (i) with nonconstant variances.
(v) Model (i) with nonadditive noise; see (3.6).
(vi) Linear transformations in general; see (3.7).
(vii) Nonlinear transformations; see (3.10).

As Figure 1 shows we first assume that the true simulation model is

$$y_i = \beta_0 + \beta_1 x_i + e_i \qquad \text{with} \qquad e_i \sim N(0, \sigma_i^2 = \sigma^2) \qquad (i = 1, \ldots, n). \tag{3.1}$$

Notice that (3.1) specifies the conditional distributions y_i, not their joint distribution

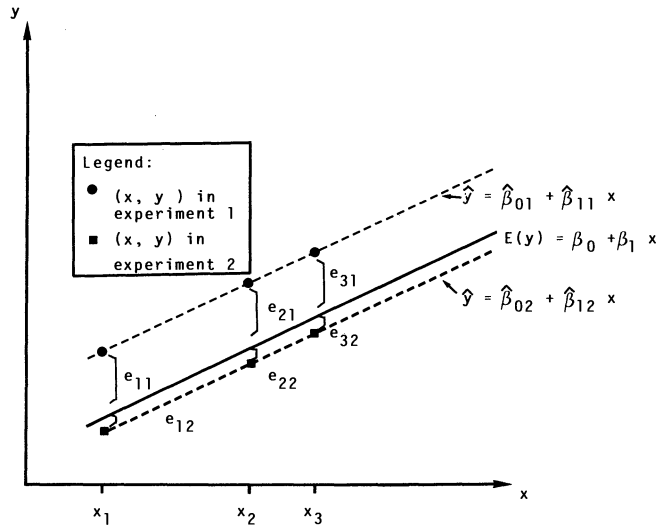

FIGURE 1. Sampling from $y_{ir} = \beta_0 + \beta_1 x_i + e_{ir}$ with Common Random Numbers ($i = 1, 2, 3$) ($r = 1, 2$).

(simulation models typically express conditional distributions; the regression metamodel should specify the statistical submodel for the regression errors). If we specify the metamodel correctly, then the simulation and the regression model have the same structure: *no specification error.* We examine three input "combinations": $n = 3$. To estimate factor effects we might use the same seeds and a single replicate ($m_i = 1$). Then we obtain identical errors: $e_{11} = e_{21} = e_{31}$ in Figure 1. First we use the OLS point estimator $\hat{\beta}$; we shall discuss EGLS in (3.2) through (3.5). Obviously, we obtain a *perfect slope estimate* and an imperfect intercept estimate: $\hat{\beta}_{11} = \beta_1$ and $\hat{\beta}_{01} = \beta_0 + e_{.1}$ where $e_{.1} = \sum_i e_{i1}/n$.

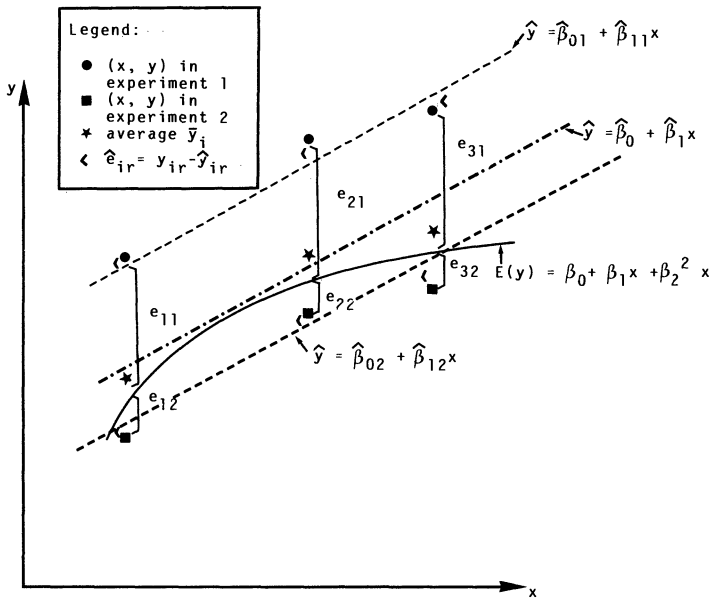

FIGURE 2. Sampling from $y_{ir} = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_{ir}$ with Common Random Numbers, While Estimating $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$.

Now we consider *replications* r ($r = 1, \cdots m \geq 2$) and we still use common seeds: $e_{12} = e_{22} = e_{32} = e_{.2}$ in Figure 1. Then we again get perfect OLS point estimates of the slope in each replication ($\hat{\beta}_{1r} = \beta_1$) and imperfect estimates of the intercept ($\hat{\beta}_{0r} = \beta_0 + e_{.r}$). Repetition also yields a perfect estimate of the *variability* of the slope estimator [$\widehat{\text{var}}(\hat{\beta}_1) = \text{var}(\hat{\beta}_1) = 0$] and a valid estimator $\widehat{\text{var}}(\hat{\beta}_0)$ of the variability of the intercept estimator; also see (2.5b).

If we have no repetitions ($m = 1$), then we cannot estimate the *variability* of the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$, unless we assume constant variances $\sigma_i^2 = \sigma^2$, so that we can estimate σ^2 through the estimated residuals $\hat{\mathbf{e}}$ where $\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$. We specified the regression model correctly, and we obtain "perfect fit", that is, all estimated residuals are equal to zero ($\hat{e}_{i1} = 0$). Hence we correctly conclude that var ($\hat{\beta}_1$) = 0 and we *incorrectly* conclude that var ($\hat{\beta}_0$) = 0.

Figure 2 illustrates a *misspecified* regression model, i.e., the regression model is still specified by (3.1) but the true simulation model now equals (3.1) augmented with the second-order term $\beta_2 x_i^2$. How can we *detect* this specification error, and what are the consequences if we do not detect the lack of fit? If we have *replications* ($m \geq 2$) then the following two tests are possible:

(i) We compare the estimated residuals $\bar{y}_i - \hat{y}_i$ to the "pure error" $\hat{\sigma}_i^2$ defined in (2.1). This comparison leads to the "F test for lack of fit" popular in experimental design, which assumes constant variances $\sigma_i^2 = \sigma^2$ and normality; see Kleijnen (1974/75, 1987).

(ii) We use the estimated effects $\hat{\boldsymbol{\beta}}$ to derive the predictor $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ for a new factor combination $\mathbf{x}_{n+1} \neq \mathbf{x}_i$ and we compare this $\hat{y}$ to the observed simulation response y_{n+1}. This validation test leads to a t test. This test is simplest if we make $\hat{y}_{n+1}$ (which depends on y_i) and y_{n+1} independent, i.e., if we use a new seed for y_{n+1}. Also see Kleijnen (1983, 1987). We may not detect the specification error, especially if the second-order effect β_2 is small compared to the first-order effect β_1 and the noise σ_i^2 (we tend to reject H_0: $\beta_1 = 0$ and accept H_0: $\beta_2 = 0$). If we have replications, then we estimate σ_i^2 using $\hat{\sigma}_i^2$ of (2.1); the estimators $\hat{\sigma}_i^2$ remain unbiased, even if we use the wrong regression model ($\hat{\sigma}_i^2$ in (2.1) uses actual simulation responses y_{ir}, not regression predictors $\hat{y}_{ir} = \hat{y}_i$). If we have *single replicates* ($m = 1$), then we saw above that we must assume constant variances $\sigma_i^2 = \sigma^2$. Unfortunately, in case of specification error the estimated residuals $\hat{e}_{i1} = \hat{y}_{i1} - y_{i1}$ do not provide an unbiased estimator of σ^2. We note that in Figure 2 not all estimated residuals $\hat{e}_{i1}$ are zero. Hence if we use $\hat{\mathbf{e}}$ to estimate var ($\hat{\beta}_1$) then $\widehat{\text{var}}(\hat{\beta}_1) \neq 0$; actually it is simple to prove that all $\hat{\beta}_{1r}$ are equal ($\hat{\beta}_{1r} = \hat{\gamma}_1 \neq \beta_1$) so that var ($\hat{\beta}_1$) = 0 (in Figure 1 we had $\hat{e}_{i1} = 0$ so that we correctly concluded that var ($\hat{\beta}_1$) = 0 and incorrectly concluded that var ($\hat{\beta}_0$) = 0).

We now return to the simplest case, where both the simulation and the regression model are specified by (3.1), and we prove that *EGLS* does not exist. As we saw, common seeds imply that all n input combinations have the same error per replication:

$$e_{ir} = e_{.r} = \frac{\sum_{i'=1}^{n} e_{i'r}}{n} \qquad (r = 1, \ldots, m). \tag{3.2}$$

Equation (3.1) yields

$$\bar{y}_i = \frac{\sum_{r=1}^{m} y_{ir}}{m} = \beta_0 + \beta_1 x_i + e_{i.} \qquad (i = 1, \ldots, n), \tag{3.3}$$

where $e_{i.} = \sum_{r=1}^{m} e_{ir}/m$. In this example, however, common seeds imply that all input combinations have the same error (say) $e_{..}$, that is, (3.2) yields

$$e_{i.} = \frac{\sum_{r=1}^{m} e_{ir}}{m} = \frac{\sum_{r=1}^{m} e_{.r}}{m} = e_{..} \qquad (i = 1, \ldots, n). \tag{3.4}$$

Substitution of (3.1) through (3.4) into (2.1) yields

$$\hat{\sigma}_{ii'} = \frac{\sum_{r=1}^{m} (e_{ir} - e_{i.})(e_{i'r} - e_{i'.})}{m-1} = \frac{\sum_{r=1}^{m} (e_{.r} - e_{..})^2}{m-1} \tag{3.5}$$

so that $\hat{\sigma}_{ii'}$ reduces to a constant, say $\hat{\sigma}^2$. Consequently the estimator $\hat{\Omega}_y$ has all $n \times n$ elements equal to a common constant, and the estimated correlation coefficients $\hat{\rho}_{ii'} = \hat{\sigma}_{ii'}/\hat{\sigma}_i\hat{\sigma}_{i'}$ are all equal to one: *maximum linear correlation*. Hence $\hat{\Omega}_y$ is *singular* and EGLS does not exist. Fortunately the estimator $\hat{\Omega}_y$ clearly warns the researcher not to apply EGLS in the example of (3.1) where all true variances and covariances are constant.

Next we introduce *variance heterogeneity* into the simple model of (3.1): $\sigma_i^2 \neq \sigma^2$. In other words, $e_i = \sigma_i z$ with $z \sim N(0, 1)$ or $e_i = e_1\sigma_i/\sigma_1$. Then (3.2) and (3.4) hold no longer, and the $\sigma_{ii'}$ in (3.5) do not reduce to a constant. Nevertheless the correlations $\rho_{ii'}$ still equal plus one, so that $\hat{\Omega}_y$ remains singular. So EGLS does not apply; OLS does (the assumptions of Schruben and Margolin's random block model, namely $\sigma_i^2 = 1$ and $\sigma_{ii'} = \rho$, do not hold).

Let us consider a variation on the simple model where the errors are *multiplicative*, not additive:

$$y_i = (\beta_0 + \beta_1 x_1)e_i. \tag{3.6}$$

It is simple to prove that we obtain perfect fit within each replication. Hence var $(\hat{\beta}_1)$ estimated from the residuals of a single replicate, yields $\widehat{\text{var}}(\hat{\beta}_1) = 0$; actually var $(\hat{\beta}_1) = \beta_1^2\sigma^2$. The perfect fit for each replication obtained by OLS cannot be improved by EGLS, that is, (3.5) still holds $(\hat{\rho}_{ii'} = \hat{\rho} = 1)$ so that $\hat{\Omega}_y$ is singular.

So GLS cannot be applied if Ω_y is *singular*, and Ω_y will be singular indeed whenever the responses y_i and $y_{i'}$ are *perfectly correlated*: if $\rho_{ii'} = 1$ then cov $(y_i, y_{i'})$ reduces to $\sigma_i\sigma_{i'}$ so that the determinant of Ω_y becomes zero. It is simple to prove that $\rho_{ii'}^2 = 1$ if $y_i = h(\mathbf{x}_i, e_{ir})$ can be written as

$$y_i = h_0(\mathbf{x}_i) + h_1(\mathbf{x}_i)h_2(e_{.r}), \tag{3.7}$$

that is, a deterministic component $h_0(\mathbf{x}_i)$ plus the product of two components, one component being a function of the deterministic input $\mathbf{x}_i$, say $h_1(\mathbf{x}_i)$, and one component being dependent on the errors e_{ir} which for common seeds reduce to $e_{.r}$, say $h_2(e_{.r})$. Then $\sigma_{ii'} = h_1(\mathbf{x}_i)h_1(\mathbf{x}_{i'})E[h_2(e_{.r}) - E\{h_2(e_{.r})\}]$ so that $\rho_{ii'} = \sigma_{ii'}/\sigma_i\sigma_{i'} = 1$. Examples of (3.7) are (3.1) and (3.6); other examples are

$$y_i = (\beta_0 + \beta_1 x_i + \beta_2 x_i^2)e_i^2 \qquad \text{and} \tag{3.8}$$

$$y_i = \exp(\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i^2)$$

$$= \exp(\beta_0 + \beta_1 x_i + \beta_2 x_i^2)\exp(e_i^2). \tag{3.9}$$

Next we consider the model

$$y_i = \exp\{(\beta_0 + \beta_1 x_i)e_i\} \qquad \text{or} \tag{3.10a}$$

$$\ln(y_i) = (\beta_0 + \beta_1 x_i)e_i, \tag{3.10b}$$

which resembles (3.6). Obviously the responses y_i have nonconstant variances and, when using common seeds such that $e_{ir} = e_{.r}$, no perfect linear correlations result: $\rho_{ii'} \neq 1$. So in general $\hat{\Omega}_y$ is not singular and EGLS yields results different from OLS and different from the Blocking Model; also see Arnold (1981) who discusses the use of alternative models to analyze experimental data, each model resulting in different statistical tests (see Chapters 14, 15 and §18.7 in Arnold 1981). We proposed the general linear model with additive noise of (2.2) as an approximation to actual non-linear

simulation models such as (3.10); both OLS and EGLS should be applied if $\hat{\Omega}_y$ is nonsingular.

4. A New Framework for Error Analysis

Following the examples of the preceding section, we propose a novel framework for the "error" analysis of simulation data, i.e., the analysis of the additive error component **e** in the general linear metamodel of (2.2). We utilize the concepts of experimental design, i.e., we distinguish three factors, each with three levels. The three factors are:

1. Random number seed.
2. Replication.
3. Validity of the regression metamodel.

For factor 1, *seed*, we distinguish the levels a, b and c:

(a) We use the same random number seeds (per replication r of the n combinations of the k factors of the simulation model; these k factors correspond to Q independent regression variables, where $Q \geq k + 1$; $r = 1, \ldots, m_i = m$ and $i = 1, \ldots, n$; see §1).

(b) We sample all seeds independently ($\sum m_i$ independent responses).

(c) We synthesize (a) and (b), i.e., assuming that the simulation model has multiple inputs ($k > 1$), we sample some seeds (namely $k_1 \geq 1$) independently and we keep some seeds ($k_2 \geq 1$) constant ($k = k_1 + k_2$) per replication; see Mihram (1972, 1983), Chang et al. (1985), Wilson (1984).

For factor 2, *replication*, we also distinguish three levels:

(a) There are no repetitions ($m_i = 1$ with $i = 1, \ldots, n$).

(b) There are repetitions ($m_i \geq 2$ for all i).

(c) There are pseudoreplications, i.e., we assume that the simulation model has a steady state (and satisfies additional technical assumptions) so that we can estimate the variances σ_i^2 from single runs (we can use subruns, renewal analysis, spectral analysis, standardized time series, etc; see Kleijnen 1987).

For factor 3, *metamodel validity*, we again distinguish three levels:

(a) The regression metamodel, used to analyze the simulation data, is correctly specified. Then we estimate the variance from the estimated residuals ($\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}}$), provided the variances are constant ($\sigma_i^2 = \sigma^2$) and we have degrees of freedom available ($N = \sum m_i > Q$).

(b) The regression model is not valid. For example, the metamodel ignores higher-order effects (such as quadratic effects); see the next level.

(c) The regression model is approximately valid. For example, the regression model is a first-order polynomial whereas the simulation model should be approximated by a second-order polynomial (curvature and interactions); however, within the area of interest (local approximation!) the neglected second-order effects may be small compared to the first-order effects and the noise σ_i^2; again see Figure 2.

We have not yet worked out this concise framework (of three factors with three levels) in full detail, but we did illustrate its use in the preceding section. We hope that our new framework will be used by other researchers too.

5. Conclusion

In the statistical analysis of a simulation experiment we must specify a statistical model. Our metamodel provides an alternative to Schruben and Margolin (1978)'s Blocking Model. Arnold (1981, p. 263) states that in general the question of which model to use has no definitive answer. We argue that the ideal model should be both realistic and simple. Our general linear regression model is indeed more realistic, since we do not assume a specific covariance pattern. Our model is also quite simple, as we can use the well-known OLS point estimator combined with the corrected standard

errors (see (2.3) and (2.5)); we can also use EGLS (see (2.4) and (2.7) assuming we have "many" replications, say $m \geq 25$, and a nonsingular covariance matrix). We further proposed a new framework that may be useful in future research on simulation output analysis.[1]

[1] I greatly appreciate the comments made by two referees which lead to the expansion of the original manuscript, the correction of some technical errors, and a much better presentation.

References

ANDERSON, H. A. AND R. G. SARGENT, "Investigation into Scheduling for an Interactive Computing System," *IBM J. Res. Develop.*, (March 1974), 125–137.

ARNOLD, S. F., *The Theory of Linear Models and Multivariate Analysis*, John Wiley & Sons, New York, 1981.

CHANG, Y., R. S. SULLIVAN, J. R. WILSON AND U. BAGCHI, "Experimental Investigation of Real-Time Scheduling in Flexible Manufacturing Systems," *Ann. Oper. Res.*, (1985).

JOHN, J. A., "New Developments in Classical Design," *Operations Research and Mathematical Statistics, Statistics (Math. Operationsforsch. Statist., Ser. Statist.)*, 11, 3 (1980), 389–402.

KLEIJNEN, J. P. C., *Statistical Techniques in Simulation. Volumes I, II*, Marcel Dekker, Inc., New York. 1974, 1975. (Russian translation: Publishing House "Statistics", Moscow, 1978.)

———, "Cross-Validation using the *t* Statistic," *European J. Oper. Res.*, 13, 2 (1983), 133–141.

———, *Statistical Tools for Simulation Practitioners*, Marcel Dekker, Inc., New York, 1987.

———, P. CREMERS AND F. VAN BELLE, "The Power of Weighted and Ordinary Least Squares with Estimated Unequal Variances in Experimental Design," *Comm. Statist., Simulation and Computation*, B14, 1 (1985), 85–102.

———, P. C. A. KARREMANS, W. K. OORTWIJN AND W. J. H. VAN GROENENDAAL, "Jackknifing Estimated Weighted Least Squares," *Communications in Statist.—Theory Methods*, 16, 3 (1987), 747–764.

———, A. J. VAN DEN BURG AND R. T. VAN DER HAM, "Generalization of Simulation Results: Practicality of Statistical Methods," *European J. Oper. Res.*, 3 (1979), 50–64.

——— AND W. VAN GROENENDAAL, *Regression Analysis of Factorial Designs with Sequential Replication*, Catholic University Brabant, 1986.

LIN, B. W. AND R. L. RARDIN, "Controlled Experimental Design for Statistical Comparison of Integer Programming Algorithms," *Management Sci.*, 25, 12 (1979), 1258–1271.

LORENZEN, T. J., "Randomization and Blocking in the Design of Experiments," *Commun. Statist.—Theory Methods*, 13, 21 (1984), 2601–2623.

MIHRAM, G. A., *Simulation: Statistical Foundations and Methodology*, Academic Press, New York, 1972.

———, "Simulation Methodology: Statistical Aspects," *Proc. Winter Simulation Conf.*, S. Roberts, J. Banks and B. Schmeiser (Eds.), IEEE, 1983.

NAYLOR, T. H., J. L. BALINTFY, D. S. BURDICK AND K. CHU, *Computer Simulation Techniques*, John Wiley & Sons, Inc., New York, 1966.

NOZARI, A., S. F. ARNOLD AND C. D. PEGDEN, "Statistical Analysis under Schruben and Margolin Correlation Induction Strategy," School of Industrial Engineering, University of Oklahoma, 1984. (Accepted for publication in *Oper. Res.*)

PERES, C. A., "Testing the Effect of Blocking in a Randomized Complete Block Design (RCBD)," *Commun. Statist.—Theory Methods*, 10, 23 (1981), 2447–2459.

SAFIZADEH, M. H., *Pseudorandom Number Assignment to Response Surface Designs*, College of Business Administration, Wichita State University, Wichita, 1983.

SCHATZOFF, M., "Design of Experiments in Computer Performance Evaluation," *IBM J. Res. Develop.*, 25, 6 (1981), 848–859.

SCHMIDT, P., *Econometrics*, Marcel Dekker, Inc., New York, 1976.

SCHRUBEN, L. W., "Designing Correlation Induction Strategies for Simulation Experiments," *Current Issues in Computer Simulation*, N. R. Adam and A. Dogramaci (Eds.), Academic Press, Inc., New York, 1979.

——— AND B. H. MARGOLIN, "Pseudorandom Number Assignment in Statistically Designed Simulation and Distribution Sampling Experiments," *J. Amer. Statist. Assoc.*, 73, 363 (1978), 504–525.

SHOUKRI, M. M. AND R. H. WARD, "On the Estimation of the Intraclass Correlation," *Commun. Statist.—Theory Methods*, 13, 10 (1984), 1239–1255.

STEINBERG, D. M. AND W. G. HUNTER, "Experimental Design: Review and Comment (and Discussion)," *Technometrics*, 26, 2 (1984), 71–130.

WILSON, J. R., "Variance Reduction Techniques for Digital Simulation," *Amer. J. Math. Management Sciences*, 4, 3, and 4 (1984), 277–312.

References

Experimental Design: Review and Comment
David M. Steinberg; William G. Hunter
Technometrics, Vol. 26, No. 2. (May, 1984), pp. 71-97.
Stable URL:
http://links.jstor.org/sici?sici=0040-1706%28198405%2926%3A2%3C71%3AEDRAC%3E2.0.CO%3B2-Y