UNIVERSITY
CATHOLIEKE
UNIVERSITEIT
BRABANT

POSTBOX 90153
5000 LE TILBURG
THE NETHERLANDS

[redacted]

R41
Simulation Models
Experimental Design
Regression Analysis

DEPARTMENT OF ECONOMICS

RESEARCH MEMORANDUM

TWO-STAGE VERSUS SEQUENTIAL SAMPLE-SIZE DETERMINATION IN REGRESSION ANALYSIS OF SIMULATION EXPERIMENTS

Jack. P.C. Kleijnen
Willem van Groenendaal

FEW 572

Communicated by Prof.Dr. F.A. van der Duyn Schouten

TWO-STAGE VERSUS SEQUENTIAL SAMPLE-SIZE DETERMINATION IN REGRESSION ANALYSIS OF SIMULATION EXPERIMENTS

JACK P.C. KLEIJNEN and WILLEM VAN GROENENDAAL

Katholieke Universiteit Brabant (Tilburg University)
School of Management and Economics
Department of Information Systems and Auditing
5000 LE Tilburg, Netherlands

In practice, simulation models usually have responses with variances that vary with the inputs. Then the number of observations (also called simulation runs or replications) per combination of simulation inputs can be selected such that the variances of the average simulation responses become (approximately) equal; in other words, combinations with high variability are replicated more often. These average responses can be analyzed through a regression metamodel. To estimate the regression parameters (or simulation input effects), Weighted Least Squares (WLS) can be applied. WLS becomes identical to Ordinary Least Squares (OLS) applied to the average simulation responses. Because the response variances are unknown, they are estimated by repeating the simulation runs with different random numbers. The estimated response variances yield the number of runs required to obtain approximately equal variances per average simulation response. Two rules for selecting the required number of simulation runs are presented, namely a two-stage and a sequential rule. The rules are first formalized and analyzed, and then evaluated through a Monte Carlo experiment. Both procedures turn out to yield confidence intervals for the estimated regression parameters that have acceptable coverage probabilities. The sequential rule demands more complicated computations to select the number of simulation runs, but this rule saves runs.

(Key Words: EXPERIMENTAL DESIGN; METAMODEL; PILOT SAMPLE; STEIN ESTIMATOR; ITERATIVE PROCEDURE; STOPPING RULE)

1. Introduction

Regression analysis of simulation data gives metamodels, which are gaining increasing popularity. Nevertheless a number of research issues rémain to be investigated, as Sargent (1991) witnesses. One of those issues is that it is not known how to satisfy the assumption of constant response variance for the various combinations of simulation inputs (or factors) that are specified by the experimental design (for example, a 2^{k-p} design). In practice the response variances may indeed differ substantially; for example, Kleijnen, Van den Burg and Van der Ham (1979, p. 60) report a simulation case-study in which 16 input combinations give estimated variances that range between the values 64 and 93,228. To obtain appropriate run lengths Sargent uses the guidelines in Whitt (1989). These guidelines, however, are heuristic

formulas to be used to estimate required simulation run lengths in the early planning stages before any data have been collected. Moreover these formulas are entirely for queuing models. In this paper, however, we consider sequentialized (iterative or dynamic) sample size determination for any type of random simulation model. Examples of such models are provided by queuing networks but also by Monte Carlo studies on the robustness of certain statistics (Kleijnen, 1974/1975, pp. 677-745, presents a Monte Carlo experiment with a multiple ranking procedure and applies a 2^{7-3} design analyzed through OLS). A popular idea in the simulation field is to obtain more observations for those input combinations (or system variants) that have higher variability. Note that "observations" are also called "replications" or "runs" in terminating simulations, and "subruns" in steady-state simulations (more research is needed for sample-size selection rules applied to steady-state simulations analyzed by renewal analysis and other time-series analyses). Unfortunately, until now the idea of taking more observations for systems with more noise was never formalized and analyzed, although Welch (1990) sketches two approaches that resemble our two procedures. He also concludes that there is a deep need to investigate sample size selection for regression metamodels.

We study two sample-size rules (also called stopping rules) that are well-known in the statistics literature: two-stage and sequential rules. Based on that literature we conjecture that two-stage sampling will be simpler but also less efficient; in other words, sequentialization leads to more complicated sample-size selection rules for the users but saves them simulation runs. The statistical analysis of such sample rules is notoriously difficult. We shall show that an exact formalization of sample-size selection in regression metamodeling is indeed cumbersome. The statistics literature virtually ignores dynamic sample-size selection in experimental designs, because in real-life experiments it is impossible or difficult to implement sequentialization. In computer simulation, however, observations become available in a sequential way, so these selection rules certainly deserve to be investigated. We hope that our paper will stimulate further analytical research on the statistical performance of sample-size selection rules in simulation, and on their implementation in simulation software. In the mean time, our Monte Carlo experimental results give simulation users some advise on how to apply the popular idea of taking more observations for noisy input combinations.

There are several other approaches to the problem of variance heterogeneity in experimental design, which we do not investigate in detail.

(a) Variance heterogeneity may be accepted and the analysis may be adjusted, that is, prefixed numbers of observations per input combination are taken and WLS is used. Our paper, however, concentrates on an alternative design of the experiment. Of course, each design requires a concomitant analysis. We shall see that in our design the WLS estimator can be computed simply by applying the OLS algorithm to the average response of each input combination.

(b) Another alternative is to transform the simulation responses, for example, to take the logarithm of the response. Such transformations, however, do not always give the desired constant variances. Moreover, they may fail to give the desired insight into the underlying simulation model. Variance stabilizing transformations are discussed in many places in Kleijnen (1987) (see its subject index on page 427).

(c) The statistics literature also discusses a prefixed sample-size selection rule that has little appeal to simulation practitioners. This rule assumes that the responses are Normally and Independently Distributed (NID) with constant variances. The user is supposed to test the null-hypothesis that the inputs have no effects, with prespecified type I and II errors. Then a fixed sample size can be determined, provided the user also specifies a value for the parameter of the non-central F statistic; see Neter, Wasserman and Kutner (1985, p. 547) and also Kleijnen (1987, pp. 207-208).

How much help does the literature provide for the issue under investigation? The statistical literature investigates stopping rules primarily in situations with only one or two populations (system variants); see the textbooks by Govindarajulu (1981) and Wetherill (1986) and the handbook edited by Ghosh and Sen (1991). The simulation literature concentrates on sample-size selection for a single population; for a survey see Kleijnen (1987).

In the remainder of this paper we use the following notation. We assume that $n \geq 2$ input combinations are simulated, and that $\underline{m}_i$, the number of replications of combination i with $i = 1,...,n$, is determined sequentially. Sequentialization means that during the simulation experiment the responses $\underline{y}_{ij}$ $(j = 1,..., \underline{m}_i)$ are obtained successively, and are used to estimate the n variances; these estimators are employed in a stopping rule. Once the simulation experiment has stopped, its data are analyzed through a Least Squares metamodel. We underline random variables, because the random character of the number of replications is essential and becomes explicit in this notation. The simulation data are analyzed through the classic linear regression model

$$\underline{y} = \mathbf{X}\beta + \underline{e} \tag{1.1}$$

with the vector of normally distributed simulation responses $\underline{y} = (\underline{y}_1,...,\underline{y}_n)'$, the matrix of independent variables $\mathbf{X} = (x_{iq})$ with $i = 1,...,n$ and $q = 1,...,Q$, the vector of input effects or regression parameters $\beta = (\beta_1,...,\beta_Q)'$, and the vector of fitting error $\underline{e} = (\underline{e}_1,...,\underline{e}_n)'$; we use bold upper case letters to denote matrices and bold lower case letters for vectors.

The remainder of this paper is organized as follows. In §2 we analyze a sample-size selection rule in case the heterogenous variances were known. This rule gives average responses with constant variances. The WLS estimator can then be computed through the familiar OLS algorithm applied to these averages. This estimator is the Best Linear Unbiased Estimator (BLUE). In §3 we examine a two-stage sampling plan for regression metamodels. In stage #1 we take a pilot sample to estimate the unknown variances. Based on these estimates we decide how many more responses to simulate for each system variant. We also present an example (in §3.2). In §4 we present a sequential procedure, that is, after the pilot sample we take only one observation from each combination that has a relatively large estimated variance for its sample average. We also introduce a novelty: to estimate variances we use all simulation responses available; we drop certain responses when estimating means. To illustrate this procedure we apply it to the same example as we used in two-stage sampling (see §4.2). In §5 we first specify a Monte Carlo experiment, that is, we select 27 combinations of three factors, namely (a) the matrix of regression variables **X**, (b) the variances of the individual simulation responses, and (c) the size of the pilot sample (used to obtain an initial estimate of these variances). Next we analyze the results of this Monte Carlo experiment. In §6 we summarize our results and indicate future research.

2.2.Known Response Variances

In this section (and only in this section) we assume that the response variances σ_i^2 are known. Intuitively it seems wise to obtain more observations for those input combinations that have high variability. More specifically, suppose we take

$$m_i = c\,\sigma_i^2 \text{ with } i = 1,...,n, \tag{2.1}$$

where c is a (common) positive constant such that the m_i become integers. This implies that we assume that the variances σ_i^2 are rational numbers; we shall return to the selection of c. Then the average response of combination i is $\bar{y}_i = \sum_{j=1}^{m_i} y_{ij}/m_i$ and has variance

$$\text{var}(\bar{y}_i) = 1/c. \tag{2.2}$$

The sample-size rule of eq. (2.1) is not necessarily 'optimal' (we shall return to this issue in the final section, which discusses future research), but it simplifies the regression analysis of the simulation data, as follows.

It is well-known that in case of known unequal variances the unbiased linear estimator with minimum variance (BLUE) is provided by WLS applied to the individual responses y_{ij}. It is simple to prove that this WLS estimator (say) $\tilde{\beta}$ follows from

$$\min \sum_{i=1}^{n} \{m_i(\bar{y}_i - \hat{y}_i)^2/\sigma_i^2\} \tag{2.3}$$

with the predicted responses $\hat{y}_i = \mathbf{x}_i' \tilde{\beta}$ where $\mathbf{x}_i'$ denotes the i^{th} row of $\mathbf{X}$ so $\mathbf{x}_i' = (x_{i1},\ldots,x_{iQ})$. Eq. (2.3) is easily interpreted: combination i gets more weight if it is replicated more often (see m_i) or if it has smaller variance (see σ_i^2). The OLS estimator $\hat{\beta}$ is still unbiased (but not BLUE) and follows from

$$\min \sum_{i=1}^{n} m_i(\bar{y}_i - \hat{y}_i)^2. \tag{2.4}$$

Under condition (2.1) we can substitute $m_i/\sigma_i^2 = c$ into (2.3), so the WLS estimate follows from

$$\min \sum_{i=1}^{n} (\bar{y}_i - \hat{y}_i)^2. \tag{2.5}$$

In other words, sample-size rule (2.1) implies that $\tilde{\beta}$, the WLS estimator computed from the individual responses (see eq. 2.3), is identical to the OLS estimator computed from the average responses ignoring the number of replications (see eqs. 2.4 and 2.5). The latter algorithm is simpler, so we shall use (2.5).

In summary, the sample size rule (2.1) gives average responses with constant variances. The WLS estimator can be computed through the familiar OLS algorithm applied to these

averages. This estimator is BLUE.

3. Two-stage Sampling

In practice the response variances are unknown, so we must estimate σ_i^2 and hence m_i in eq. (2.1). Many sampling schemes are conceivable: two-stage (leading to so-called Stein estimators), multistage, and sequential schemes. In this section we concentrate on a two-stage procedure. In § 3.1 we give some elementary theory; in § 3.2 we give a simple example.

3.1 Theory

In stage #1 we take a pilot sample of $m_0 \geq 2$ responses from each input combination. Then we compute the n classic variance estimators

$$\underline{s}_i^2(m_0) = \sum_{j=1}^{m_0} (\underline{Y}_{ij} - \bar{\underline{Y}}_i(m_0))^2/(m_0-1) \tag{3.1}$$

with the sample average of combination i (computed from m_0 simulation runs)

$$\bar{\underline{Y}}_i(m_0) = \sum_{j=1}^{m_0} \underline{Y}_{ij}/m_0 . \tag{3.2}$$

Because sample size is a crucial variable in our study, we explicitly show the dependence of the various estimators on the sample size. Different values for the size of the pilot sample will be investigated in our Monte Carlo experiment (see §5).

Now we use (2.1), the sample-size formula for known variances. In that equation we replace the variances by the estimators defined in (3.1). Moreover, if we would require (2.1) to hold exactly, then the m_i might be very large integers. 'Nearly' constant variances require smaller integers. So we decide to take no additional observations from the combination with the smallest estimated variance; (Welch, 1990, p. 393 proposes a similar approach). For the other combinations we estimate the required number of observations through

$$\hat{\underline{m}}_i = [m_0 \, \underline{s}_i^2(m_0)/ \min_{1 \leq i' \leq n} \underline{s}_{i'}^2(m_0)], \tag{3.3}$$

where [x] denotes the integer closest to x. In stage #2 we take $\hat{\underline{m}}_i - m_0$ additional observations from combination i (so at the end of stage #2 we have $\hat{\underline{m}}_i$ responses from system variant i; at least one system variant has exactly m_0 responses). We use all observations from a particular system variant to estimate the mean and the variance of the response of that system:

$$\bar{\underline{y}}_i(\hat{\underline{m}}_i) = \sum_{j=1}^{\hat{\underline{m}}_i} \underline{y}_{ij} / \hat{\underline{m}}_i \tag{3.4}$$

and

$$\underline{s}_i^2(\hat{\underline{m}}_i) = \sum_{j=1}^{\hat{\underline{m}}_i} (\underline{y}_{ij} - \bar{\underline{y}}_i(\hat{\underline{m}}_i))^2 / (\hat{\underline{m}}_i - 1). \tag{3.5}$$

We shall need the following properties of these estimators for the n means and n variances respectively. The estimator of the expected simulation response is the average of a random number of responses, that is, it is a ratio estimator. In general, ratio estimators are biased (remember renewal analysis of simulation output). However, our estimator (3.4) remains unbiased, under our assumption of normally distributed simulation responses, as we prove as follows. For a fixed sample size the estimators of the mean and the variance are independent if and only if the responses are normally distributed; see Mathai and Pederzoli (1977). The denominator $\hat{\underline{m}}_i$ of (3.4) depends on the estimated variances $\underline{s}_i^2(m_0)$ through (3.3), but these $\underline{s}_i^2(m_0)$ are independent of $\bar{\underline{y}}_i(\hat{\underline{m}}_i)$. This yields:

$$E(\bar{\underline{y}}_i(\hat{\underline{m}}_i)) = \underset{\hat{\underline{m}}_i}{E} \{ E(\bar{\underline{y}}_i(\hat{\underline{m}}_i) \mid \hat{\underline{m}}_i = m_i) \} = \underset{\hat{\underline{m}}_i}{E} (\mathbf{x}_i' \boldsymbol{\beta}) = \mathbf{x}_i' \boldsymbol{\beta} . \tag{3.6}$$

However, the estimator of the variance of the sample *average*

$$\hat{\text{var}}(\bar{\underline{y}}_i(\hat{\underline{m}}_i)) = \sum_{j=1}^{\hat{\underline{m}}_i} (\underline{y}_{ij} - \bar{\underline{y}}_i(\hat{\underline{m}}_i))^2 / \{(\hat{\underline{m}}_i - 1)\hat{\underline{m}}_i\} \tag{3.7}$$

is biased, as we prove as follows. If we condition on all variance estimates $s_i^2(m_0)$ -and hence on $\hat{\underline{m}}_i$- then we can prove (see appendix 1) that

$$E\{\underline{s}_i^2(\hat{m}_i) \mid \underline{s}_i^2(m_0) = s_i^2(m_0)\} = s_i^2(m_0) + \{\sigma_i^2 - s_i^2(m_0)\}(\hat{m}_i - m_0)/\hat{m}_i . \tag{3.8}$$

This equation means that $\underline{s}_i^2(\underline{\hat{m}}_i)$ underestimates the true variance σ_i^2: if $s_i^2(m_0)$ happens to be small, then no more observations are taken in the second stage ($\hat{m}_i = m_0$) so this underestimate is not recomputed. If $s_i^2(m_0)$ happens to be large, then more observations are taken, so the overestimate is recomputed. Averaging over both possibilities shows that the variance estimator $\underline{s}_i^2(\underline{\hat{m}}_i)$ underestimates σ_i^2. Note that the combination with the smallest (true) variance is probably the combination with the smallest number of observations; it is also the combination that can be estimated best (because $\text{var}(\underline{s}^2) = 2\sigma^4/(m-1)$). Finally, $\hat{\text{var}}(\bar{\underline{y}}_i(\underline{\hat{m}}_i))$ is biased because it equals $\underline{s}_i^2(\underline{\hat{m}}_i) / \underline{\hat{m}}_i$, which has a numerator that underestimates, and a denominator that is random (see eq. 3.7). Whether this bias is important, will be examined in our Monte Carlo experiment. As these derivations illustrate, many statistical properties that are known for fixed sample sizes, must be re-evaluated for random sample sizes.

Obviously the estimated variances computed at the end of stage #2 may differ substantially. The two-stage procedure, however, accepts this variance heterogeneity (whereas the sequential procedure realizes a common estimated variance of the sample averages, at the time the simulation stops, as we shall see in §4).

Finally we obtain the WLS estimators for the regression parameters. Equation (2.5) showed how to compute the WLS estimator $\tilde{\beta}$ for fixed sample sizes and known response variances. This equation inspires us to estimate the regression parameters in two-stage sampling through

$$\tilde{\underline{\beta}}(\underline{\hat{m}}) = (X'X)^{-1}X'\,\bar{\underline{y}}(\underline{\hat{m}}) \tag{3.9}$$

with $\underline{\hat{m}} = (m_0, \underline{\hat{m}}_1, \ldots, \underline{\hat{m}}_n)'$ the vector of n+1 sample sizes and $\bar{\underline{y}}(\underline{\hat{m}})$ the vector of n sample means defined in (3.4). This regression estimator is unbiased because (as we saw in eq. 3.6) the response averages $\bar{\underline{y}}(\underline{\hat{m}})$ are unbiased. Its covariance matrix is

$$\text{cov}(\tilde{\underline{\beta}}(\underline{\hat{m}})) = (X'X)^{-1}X'\,\text{cov}(\bar{\underline{y}}(\underline{\hat{m}}))X(X'X)^{-1}\,, \tag{3.10}$$

where $\text{cov}(\bar{\underline{y}}(\underline{\hat{m}}))$ denotes the covariance matrix of the average responses at the end of stage #2. To estimate $\text{cov}(\tilde{\underline{\beta}}(\underline{\hat{m}}))$ we need to estimate $\text{cov}(\bar{\underline{y}}(\underline{\hat{m}}))$. The simulation user can not estimate the covariances among the sample averages, from a single simulation experiment. Therefore we assume these covariances to be zero. We will investigate this assumption in the

Monte Carlo experiment. The user can estimate the variances of the sample means through (3.7). So we assume $\hat{cov}(\bar{\underline{y}}(\hat{\underline{m}}))$ to be a diagonal matrix with elements defined by (3.7). Since users know **X**, they can compute $\hat{cov}(\hat{\underline{\beta}}(\hat{\underline{m}}))$ through the sample analogue of (3.10).

The square roots of the diagonal elements of $\hat{cov}(\hat{\underline{\beta}}(\hat{\underline{m}}))$ give the 'standard errors' (estimated standard deviations) of $\hat{\beta}_q$ (with q = 1,..., Q), the individual estimated regression parameters. To obtain confidence intervals for β_q we propose the classic Student statistic t_υ. *What is the correct number of degrees of freedom* υ? Scheffé (1970, p. 1502) proposes $\upsilon = \min(m_i - 1)$ in his study of a case with only two populations (n = 2: Behrens-Fisher problem); see Dudewicz and Mishra (1988, pp. 503-514) and Kleijnen (1974/1975, p. 472). Further, Kleijnen, Cremers and Van Belle (1985) study regression metamodels with fixed and equal sample sizes $m_i = m$, and investigate several degrees of freedom, namely $\min(m_i - 1)$, $\sum_{i=1}^{n}(m_i - 1)$ and infinity (t_∞ is a standard normal variate, say, z). In two-stage (and sequential) experiments $\sum_{i=1}^{n}(m_i - 1)$ is high. Therefore we restrict ourselves to $\min(\hat{m}_i - 1) = m_0 - 1$ and infinity. Note that MacNair and Welch (1991, p.824) also reduce the degrees of freedom of the t statistic in order to realize good coverage probabilities of their sequential procedure, albeit for a situation with a single population (n = 1). Fortunately, our choice implies non-stochastic υ. This results in the $1 - \alpha$ confidence intervals

$$\hat{\beta}_q(\hat{\underline{m}}) \pm t_\upsilon^{\alpha/2} \sqrt{\hat{var}(\hat{\beta}_q(\hat{\underline{m}}))} \quad \text{with } q = 1,...,Q. \tag{3.11}$$

In the Monte Carlo experiment we shall investigate whether these intervals cover the true values with prescribed probability of $1 - \alpha$. Note that joint confidence intervals can be derived from the individual confidence intervals by applying Bonferroni's inequality.

3.2 Example

To illustrate two-stage sampling we could use the M/M/1 queue as Sargent (1991) did. However, we would then have to rely on the asymptotic normality of the simulation responses. Moreover, the metamodel would then have specification error; in other words, the fitting errors would not have zero expectation so the estimator of the regression effects β would be biased. Therefore we conduct a Monte Carlo experiment with perfectly normal 'simulation' responses and with fitting errors that have zero expectation (in future research we can explore the robustness of two-stage regression analysis with respect to nonnormality

and the validation of such metamodels). We consider an example with two simulation inputs. Suppose that all inputs have effects of +1; the overall mean (or intercept, often denoted by β_0, but in this paper denoted by β_1) is also +1 (so $\beta_1 = \beta_2 = \beta_3 = 1$). We use a 2^2 design to specify the four input combinations of the 'simulation'. So **X** in (1.1) has column 1 with all elements equal to 1; column 2 with elements -1, 1, -1 , 1; column 3 with -1, -1, 1, and 1. (Consequently the expected value of the first combination is $E(\underline{y}_1) = \mathbf{x}'_1\boldsymbol{\beta} = \beta_1 - \beta_2 - \beta_3 = -1$.) For the response variances we select the values 1, 4, 8, 11.38 respectively ($\sigma_1^2 = 1, \ldots, \sigma_4^2 = 11.38$). The pilot sample-size is 25 ($m_0 = 25$). To save space we do not display the m_0 individual responses for the four combinations, but we mention only that the responses in combination 1 range between -2.25 and +1.32. Their average is -0.546 and their estimated sample variance is 0.803 (so this part of the Monte Carlo simulation program seems correct). For the other three combinations we get the following sample estimators: $s_2^2(m_0) = 4.915$, $s_3^2(m_0) = 8.816$, and $s_4^2(m_0) = 18.615$ (comparing these estimates with the true variances, we conclude that this part of the computer program seems indeed without bugs). These variance estimates imply that the number of additional responses to be taken in stage #2, are 0, 128, 249, and 554 respectively (for example, $\hat{m}_4 - m_0 = [25(18.615/0.803)] - 25 = 554$). After taking these extra observations we get $\bar{y}_1(\hat{m}_1=25) = -0.546$, $\bar{y}_2(\hat{m}_2=153) = 0.900$, $\bar{y}_3(\hat{m}_3=274) = 0.997$, and $\bar{y}_4(\hat{m}_4=579) = 2.873$. The estimated variances of these averages are 0.03213, 0.02492, 0.02880, and 0.02283. The estimates of the regression parameters are 1.056, 0.831, and 0.879 (remember that their true values are 1). Their estimated variances are 0.007 for all three parameters. Hence their standard errors are 0.08. How about the confidence intervals? Let us concentrate on the tightest interval, namely the interval based on the normal approximation ($\upsilon = \infty$) and a high value for the type I error, say $\alpha = 0.10$. Then the $1 - \alpha/2$ quantile of $\underline{z}$ is $z_{\alpha/2} = 1.645$. The upper bound of the 90% confidence interval for β_2 (which is the parameter with the largest deviation from its expected value) is then 0.831 + (1.645)(0.08) = 0.96, which does not cover the true value. Is this bad luck? Does this also happen when users select a lower type I error rate and other degrees of freedom for the t statistic? We also wonder whether the two-stage procedure achieves its goal of (approximately) constant variances for the average responses. These and other questions will be answered in the section on the Monte Carlo experiment.

4. Sequential Sampling

The first stage of sequential sampling is identical to stage #1 of two-stage sampling. So we take a pilot sample of size m_0 from each of the n combinations and estimate the n response variances through (3.1). We use these variance estimators $\underline{s}_i^2(m_0)$ to obtain our *first* estimators of the required number of replications $\underline{\hat{m}}_i$; see (3.3). Whereas under two-stage sampling we jump ahead and obtain $\underline{\hat{m}}_i - m_0$ observations, we now proceed more cautiously: for each population with $\hat{m}_i > m_0$ we obtain a single new response.

After we have acquired one new simulation output for each system variant -except for the system with the smallest estimated response variance after the pilot stage- we *re*compute all estimated response variances. To estimate the variance σ_i^2 at a particular stage we use all responses from combination i that are available at that stage. Let $\underline{m}_{it}$ denote the number of responses from combination i available after stage t. The total number of stages at the end of the sequential procedure is not known beforehand; it is a random variable, which we denote by $\underline{T}$ (so t runs from 1 through $\underline{T}$; for t = 1 we have $\underline{m}_{i1} = m_0$). Hence we replace (3.5) by

$$\underline{s}_i^2(\underline{m}_{it}) = \sum_{r=1}^{\underline{m}_{it}} (\underline{y}_{ir} - \underline{\bar{y}}_i(\underline{m}_{it}))^2 / (\underline{m}_{it} - 1) , \qquad (4.1)$$

where (analogous to eq. 3.4) $\underline{\bar{y}}_i(\underline{m}_{it})$ denotes the average computed from the available $\underline{m}_{it}$ responses. Note that Kleijnen and Van Groenendaal (1992, pp. 12-13) discuss the numerical accuracy of different algorithms for the computation and updating of estimated means and variances.

In each stage we re-estimate the required number of replications, using the most recent update of the variance estimate, which was defined in (4.1). We continue until (for the first time) the desired number of replications does not exceed the available number, for all n system variants simulated:

$$\forall i : \underline{\hat{m}}_{it} \leq \underline{m}_{it} . \qquad (4.2)$$

So $\underline{T}$ denotes the minimum number of stages for which condition (4.2) holds. Once we have stopped generating simulation responses, we use a regression metamodel to analyze the simulation data.

A novelty of our sequential procedure is the distinction we introduce between $\underline{m}_{it}$, the number of available responses for combination i in stage t, and $\underline{\hat{m}}_{it}$, the estimator of the number of responses for combination i in stage t that is necessary to realize nearly constant response variances. To use the simple formula (2.5) (which leads to eqs. 3.9 and 4.3) for our estimators of β we assumed that the sample averages have constant variances; in other words, we estimate the expected simulation responses $E(\underline{y}_i)$ from $\underline{\hat{m}}_{it}$ responses. However, to estimate the nuisance parameters σ_i^2 we use as much information as possible, that is, we use $\underline{m}_{it}$ observations (also see eq. 4.6). In practice, the difference between $\underline{\hat{m}}_{it}$ and $\underline{m}_{it}$ is negligible at the *end* of the sequential procedure (t = T), because $s_i^2(m_{it})$ changes only slowly (also see the experimental data in Table 1).

Note that our 'novelty' is reminiscent of a technique devised by Scheffé in 1943 for the derivation of an exact confidence interval for the difference between the means of two normal populations with unequal variances (Behrens-Fisher problem). He used all observations when estimating the mean of the population with the largest sample. His variance estimator, however, used fewer observations (such that both populations have equal sample sizes). We estimate the mean, not the variance, from fewer observations; moreover we extend his idea to regression analysis. For a further discussion of Scheffé's procedure we refer to Kleijnen (1974/1975, p. 473).

Our WLS estimator based on a sequentially determined number of simulation responses is

$$\underline{\tilde{\beta}}(\underline{\hat{m}}_T) = (X'X)^{-1}X'\,\underline{\bar{y}}(\underline{\hat{m}}_T) \tag{4.3}$$

with $\underline{\hat{m}}_T = (m_0, \underline{\hat{m}}_{1T}, \ldots, \underline{\hat{m}}_{nT})'$ the vector of (n+1) sample sizes and $\underline{\bar{y}}(\underline{\hat{m}}_T)$ the vector of n sample means

$$\underline{\bar{y}}_i(\underline{\hat{m}}_{iT}) = \sum_{i=1}^{\underline{\hat{m}}_{iT}} \underline{y}_{ii} \;/\; \underline{\hat{m}}_{iT} \;. \tag{4.4}$$

This regression estimator is unbiased under normality; see (3.6). Its covariance matrix is

$$\mathrm{cov}(\underline{\tilde{\beta}}(\underline{\hat{m}}_T)) = (X'X)^{-1}X'\,\mathrm{cov}(\underline{\bar{y}}(\underline{\hat{m}}_T))X(X'X)^{-1} \;, \tag{4.5}$$

where $\mathrm{cov}(\underline{\bar{y}}(\underline{\hat{m}}_T))$ denotes the covariance matrix of the average responses at the end of stage

#$\underline{T}$. As in two-stage sampling we assume zero correlations among the average simulation responses at the end of the sequential procedure. We estimate the variances of the sample means through

$$\hat{var}(\bar{\underline{y}}_i(\underline{\hat{m}}_{iT})) = \underline{s}_i^2(\underline{m}_{iT}) / \underline{\hat{m}}_{iT} \tag{4.6}$$

where the variance of an individual simulation response is estimated from all $\underline{m}_{iT}$ available observations (see eq. 4.1), whereas the estimated variance of the average response accounts for the fact that this average uses only $\underline{\hat{m}}_{iT}$ responses (see the denominator of eq. 4.6). This estimator is biased (see the discussion of eq. 3.7), but this bias may be negligible in large samples (see the Monte Carlo experiment). We further assume that our sequential procedure indeed yields an (approximately) common variance for the averages $\bar{\underline{y}}_i(\underline{\hat{m}}_{iT})$. Our sequential procedure implies that the n variance estimates of (4.6) are indeed virtually the same at t = T (we continue until these estimates are essentially equal). Therefore we estimate this common variance by *pooling* these n variance estimators:

$$\bar{\underline{s}}^2 = \sum_{i=1}^{n} \hat{var}(\bar{\underline{y}}_i(\underline{\hat{m}}_{iT}))/n \; . \tag{4.7}$$

Hence (4.5) reduces to a simple and familiar formula:

$$\hat{cov}(\underline{\tilde{\beta}}(\underline{\hat{m}}_{iT})) = \bar{\underline{s}}^2(\mathbf{X}'\mathbf{X})^{-1} \; . \tag{4.8}$$

Note that **X** is often orthogonal in well-designed simulation experiments, as the example in § 3.2 demonstrated. Then the individual regression estimators become uncorrelated with constant estimated variances $\bar{\underline{s}}^2/n$. These properties, however, are not essential for our approach.

To obtain confidence intervals for β_g we use the Student statistic $t_{n'}$ as in two-stage sampling. We again restrict ourselves to degrees of freedom equal to $\min(\hat{m}_i - 1) = m_0 - 1$ and infinity respectively. This results in the $1 - \alpha$ confidence intervals of (3.11) where we replace $\underline{\hat{m}}$ by $\underline{\hat{m}}_T$. In the Monte Carlo experiment we shall investigate whether these intervals cover the true values with prescribed probability of $1 - \alpha$, but first we illustrate sequential sampling through an example.

4.2 An Example

We consider the same example as in §3.2 on two-stage sampling. The pilot sample-size is again 25 ($m_0 = 25$); that is, we obtain the same number of responses as in stage #1 of the two-stage procedure. However, we use different pseudorandom number streams for the two sampling procedures (the pseudorandom numbers would get out of step: synchronization problem; see Kleijnen, 1974/1975, p. 202). We get the following variance estimates: $s_1^2(m_0) = 1.165$, $s_2^2(m_0) = 2.881$, $s_3^2(m_0) = 7.495$, and $s_4^2(m_0) = 5.849$ (remember that $\sigma_1^2 = 1$, $\sigma_2^2 = 4$, $\sigma_3^2 = 8$, $\sigma_4^2 = 11.38$ so the pilot sample wrongly suggests that the variance of system 3 is largest). These variance estimates imply that we take one new observation from combinations 2, 3, and 4 respectively. After taking these extra observations we get $\bar{y}(\hat{m}_1 = 25) = -0.782$, $\bar{y}(\hat{m}_2 = 26) = 0.945$, $\bar{y}(\hat{m}_3 = 26) = 1.661$, $\bar{y}(\hat{m}_4 = 26) = 3.147$. The re-estimated variances of the individual responses are 1.165, 2.967, 7.228, and 6.201 (these updates lag behind). These variance estimates imply estimated required sample sizes of 25, 64, 155, and 133. So we take one more observation from combinations 2, 3, and 4. And so on. Table 1 gives some more data.

INSERT TABLE 1

As this table shows, after 221 stages the available numbers of observations are at least as big as the required numbers. For example, for combination 4 there are 25 responses from the pilot sample (stage #1) plus 220 responses from stage #2 through stage #221, which adds up to 245 responses (at the end of stage #220 we estimate that [(11.4461/1.165)25] = 246 responses are required for combination 4, and this number 246 is entered under stage 221; at the end of stage #221 we compute $\hat{m}_4$ = [(11.405/1.165)25] = 245 and stop). So condition (4.2) is met and we stop sampling. For combination 2 the mean simulation response is estimated from 89 (not 90) responses. In stage #65 this number of responses was reached; then the estimated mean was 1.014 (not 1.012, which is the estimated mean in stage #66). The variance is estimated from all 90 observations, and equals 4.156. At the end of the sequential procedure the estimated variances of the four *average* responses are 0.04662, 0.04669, 0.04664, and 0.04655 (for example 11.405/245 = 0.04655), so these values are indeed 'approximately' equal. Their average is 0.04663. The estimates of the regression parameters are 0.964, 0.883, and 0.848 (their true values are 1). Their estimated variances

are 0.012 for all three parameters. Hence their standard errors are 0.11. The tightest confidence interval is based on the normal approximation and $\alpha = 0.10$ (so $z_{\alpha/2} = 1.645$). The upper bound of the 90% confidence interval for β_3 (which is the parameter with the largest deviation from its expected value) is then $0.848 + (1.645)(0.11) = 1.03$, so the interval covers the true value ($\beta_3 = 1$).

Let us compare these results with those of the two-stage procedure (which are independent, as we saw above). The total number of simulation responses is 1031 in two-stage sampling and 503 in sequential sampling. Consequently the estimated variances of the estimated regression parameters is larger in sequential sampling: 0.012 versus 0.007. A criterion for comparing the two procedures may be $\sqrt{\hat{var}(b_q)} \times \sum_{r=1}^{n} m_r$, which equals 55 ($= 0.11 \times 503$) for sequential and 82 ($= 0.08 \times 1031$) for two-stage sampling. So sequential sampling seems better. Are the results of this example accidental? This question will be answered in the next section.

5. Monte Carlo Experiment

In §5.1 we specify the design of our Monte Carlo experiment; in §5.2 we present the resulting output.

5.1 Monte Carlo inputs

In our Monte Carlo experiment we use nearly the same inputs as in Kleijnen, Cremers and Van Belle (1985), who studied fixed sample sizes. So we have the following four **X** matrices.

(a) **X** is a 2^{6-2} design including all six main effects, only six two-factor interactions, and the intercept (16 combinations to estimate 13 effects).

(b) **X** is an 8 x 4 matrix following from a 2^3 design (with three main effects and β_0). Note that **X** is a submatrix of **X** in (a).

(c) **X** follows from a 2^2 design, and is used in the examples of the preceding two sections.

We combine these three matrices with different degrees of variance heterogeneity quantified by

$$H = \{\max(\sigma_i^2) - \min(\sigma_i^2)\}/\min(\sigma_i^2) \quad (5.1)$$

We fix H at the values 0, 4, and (roughly) 11. The individual response variances σ_i^2 are shown in appendix 2 and are also taken from Kleijnen et al. (1985, p. 95). We assume that the simulation practitioner does not use common random numbers; hence $\mathbf{cov}(\underline{y})$, the covariance matrix of the individual simulation responses, is diagonal.

We further use the β values of Kleijnen et al. (1985, p. 95). The precise values of β do not affect the interpretation of the Monte Carlo experiment; so we move β to appendix 2.

We study different sizes of m_0, the pilot sample, namely 4, 9 and 25.

We repeat each of the 27 ($= 3^3$) Monte Carlo cases (specified by $\mathbf{X}$, $\mathbf{cov}(\underline{y})$, and m_0) 150 times, since in Kleijnen et al. (1985) 150 'macroreplications' reduce noise so much that clear patterns emerge from the Monte Carlo experiment.

We use the NAG (Numerical Algorithms Group) multiplicative pseudorandom number generator, which has multiplier 13^{13} and modulo 2^{59}. All 27 cases use different pseudorandom number streams (we continue a next case where we stopped the preceding case). The initial seed is selected randomly (using the computer's internal clock).

5.2 Monte Carlo Output

In the discussion of the examples in §3.2 and §4.2 we have already mentioned some outputs that seem to confirm the correctness of parts of our computer program.

Before we present the estimated coverage probabilities of the two sampling procedures, we mention some *intermediate* results that we obtained by repeating the examples in §3.2 and §4.2 150 times using independent pseudorandom numbers. (Those examples refer to only one of the 27 cases, whereas the coverage probabilities are investigated for all 27 cases.)

First we test whether $\bar{y}_i(\hat{\underline{m}}_i)$, the average response of combination i at the end of the two-stage procedure, is *normally* distributed. We use the 150 observations on $\bar{y}_i(\hat{\underline{m}}_i)$ to form a histogram with 14 classes. We know the mean, but we must estimate the variance of these observations. So we use a chi-square statistic for goodness of fit with 12 degrees of freedom. We test, at a type I error rate of 0.05 and 0.025, whether the histogram fits a normal distribution. For none of the four combinations we reject the null-hypothesis. The same conclusion (normal averages) holds for sequential sampling.

Our Monte Carlo experiment gives unbiased estimators of $\text{var}(\bar{\underline{y}}_i(\hat{\underline{m}}_i))$, the *variance of the sample average* of combination i at the end of the two-stage procedure, and of their covariances:

$$\hat{\text{cov}}(\bar{\underline{y}}_i(\hat{\underline{m}}_i), \bar{\underline{y}}_j(\hat{\underline{m}}_j)) = \sum_{r=1}^{150} \{\bar{\underline{y}}_{ir}(\hat{\underline{m}}_i) - \bar{\bar{\underline{y}}}_i(\hat{\underline{m}}_i)\}\{\bar{\underline{y}}_{jr}(\hat{\underline{m}}_j) - \bar{\bar{\underline{y}}}_j(\hat{\underline{m}}_j)\}/149 \tag{5.2}$$

with

$$\bar{\bar{\underline{y}}}_i(\hat{\underline{m}}_i) = \sum_{r=1}^{150} \bar{\underline{y}}_{ir}(\hat{\underline{m}}_i)/150. \tag{5.3}$$

We emphasize that simulation practitioners can *not* compute these estimators, since they execute their simulation experiments only once (not 150 times as in our Monte Carlo experiment). They do compute $\hat{\text{var}}(\bar{\underline{y}}_i(\hat{\underline{m}}_i))$, the estimated variance of $\bar{\underline{y}}_i(\hat{\underline{m}}_i)$, which is biased (see the discussion below eq. 3.7). The example in §3.2 gave one observation on the latter estimator, for each of the four combinations of simulation inputs: 0.032, 0.025, 0.029, and 0.023. The unbiased estimator (5.2) yields 0.038, 0.052, 0.051, and 0.044 for these four combinations. We wish to test whether these differences are significant. Therefore we compare $\overline{\hat{\text{var}}(\bar{\underline{y}}_i(\hat{\underline{m}}_i))}$, the average value of the 150 realizations of the simulationist's estimator, with $\hat{\text{cov}}(\bar{\underline{y}}_i(\hat{\underline{m}}_i), \bar{\underline{y}}_i(\hat{\underline{m}}_i))$, the one value obtained after our 150 Monte Carlo experiments, which follows from (5.2) upon replacing j by i. Note that we use the symbol $\hat{\text{cov}}(\bar{\underline{y}}_i(\hat{\underline{m}}_i), \bar{\underline{y}}_i(\hat{\underline{m}}_i))$ to avoid confusion with $\hat{\text{var}}(\bar{\underline{y}}(\hat{\underline{m}}_i))$ defined in (3.7). It is well known that the variance of the variance estimated from n observations on a normally distributed variable with variance (say) σ^2 is $2\sigma^4/(n-1)$. Therefore we compute the statistic

$$\underline{z} = \frac{\{\overline{\hat{\text{var}}(\bar{\underline{y}}_i(\hat{\underline{m}}_i))} - \hat{\text{cov}}(\bar{\underline{y}}_i(\hat{\underline{m}}_i), \bar{\underline{y}}_i(\hat{\underline{m}}_i))\}}{\sqrt{\sum_{r=1}^{150} \{\hat{\text{var}}(\bar{\underline{y}}_i(\hat{\underline{m}}_i))_r - \overline{\hat{\text{var}}(\bar{\underline{y}}_i(\hat{\underline{m}}_i))}\}^2/(149\times150) + 2\,\hat{\text{cov}}(\bar{\underline{y}}_i(\hat{\underline{m}}_i), \bar{\underline{y}}(\hat{\underline{m}}_i))^2/150}} \tag{5.4}$$

with

$$\overline{\hat{\text{var}}(\bar{\underline{y}}_i(\hat{\underline{m}}_i))} = \sum_{r=1}^{150} \hat{\text{var}}(\bar{\underline{y}}_i(\hat{\underline{m}}_i))_r/150. \tag{5.5}$$

Note that the last factor of (5.4) is 150: although $E(\underline{s}^2) = \sigma^2$, we know that $E(\underline{s}^4) \neq \sigma^4$ so the exact value of this factor is rather arbitrary. For the four combinations we obtain the following values for this z: 0.502, -1.631, -1.267, and -0.188. We assume that (5.4) indeed gives

a standard normal variable. Then none of these values is significant at type I error rates of 0.10, 0.05, or 0.01. So these results suggest that there is no significant bias. For the sequential procedure we get -0.675, -2.313, -1.067, and -0.545, so only one z value is significant at type I error rates of 0.10 and 0.05. This one case does not suggest that the variance estimator for the average simulation response at the end of the procedure (two-stage or sequential) is biased, even though it ignores the random character of the sample sizes.

The simulationist can not estimate the correlations among the sample averages, from a single simulation experiment. Therefore we assumed them to be zero. This assumption can be investigated in the Monte Carlo experiment. The estimated correlations range between -0.125 and +0.099 in two-stage sampling, and between -0.133 and +0.107 in sequential sampling, so they may indeed be assumed zero.

Our estimator of β (the regression parameters) is a linear transformation of the vector of average simulation responses at the end of the procedure. So we expect this estimator to be *normally* distributed. The chi-square goodness of fit test is indeed not significant, neither in two-stage nor in sequential sampling.

Next we ask whether the Studentized $\hat{\beta}_q$ has indeed a *t distribution*; see (3.11). In the discussion of this distribution we proposed degrees of freedom equal to $m_0 - 1 = 24$. So we determine the histogram of t_{24}, again with 14 classes, which leads to a chi-square statistic with 13 degrees of freedom. For none of the three regression parameters χ^2_{13} is significant. The same conclusion ($\hat{\beta}_q$ is distributed like t_{m_0-1}) holds for the sequential procedure.

Simulation practitioners, however, use only the *tails* of the t distribution. So the distribution of the Studentized β estimator may resemble a t distribution, but their critical values (such as the .90 quantile) may differ significantly. This takes us to the final output of our Monte Carlo experiment, the estimated *coverage probability*. So we consider two-sided $1 - \alpha$ confidence intervals for β_q based on the tabulated $1 - \alpha/2$ quantile (say) $t_{\upsilon;\alpha/2}$ of the Student statistic t_υ with $\upsilon = m_0 - 1$ and $\upsilon = \infty$ respectively. We suppose that simulationists use one of the following classic values: $\alpha = 0.01$, 0.05 or 0.10. Each of the 150 Monte Carlo repetitions yields one confidence interval per regression parameter β_q, given a value for α and υ. So we can check if that interval covers the true value β_q. If the interval does not cover the true parameter value, then we score (say) a one; otherwise a zero. So we obtain a binomial variable $\hat{p}$ based on 150 observations. In this way we estimate the expected coverage of the confidence interval: is this value $1 - \alpha$?

We formulate a one-sided null-hypothesis:

$$H_0': E(\hat{p}) \leq \alpha \text{ versus } H_1': E(\hat{p}) > \alpha. \tag{5.6}$$

The alternative hypothesis means that the coverage probability is smaller than the nominal one, that is, the confidence intervals are too tight.

Note that we also test the related two-sided hypothesis

$$H_0: E(\hat{p}) = \alpha \text{ versus } H_1: E(\hat{p}) \neq \alpha. \tag{5.7}$$

This test gives similar conclusions. So in this paper we concentrate on the one-sided null-hypothesis.

It is convenient to approximate the binomial distribution of $\hat{p}$ through the normal distribution $N(\alpha, \alpha(1-\alpha)/150)$ with α defined by (5.6).

A complication is that there are Q parameters β_q: the more parameters, the more likely it is that we find some estimated coverage probability deviating from its expected value by pure chance. Therefore we apply Bonferroni's inequality, that is, we test H_0 with a type-I error rate of 0.05/Q so that the experimentwise error rate is 0.05 at most (see Kleijnen, 1987, p. 42).

In the Monte Carlo experiment we consider confidence intervals for β_q with α is 0.01, 0.05 and 0.10, while the degrees of freedom are $m_0 - 1$ and infinity respectively. For the two-stage procedure with $m_0 - 1$ degrees of freedom we reject H_0 in none of the 27 cases, with the exception of one case. In the latter case we have $m_0 = 25$ and $\alpha = 0.10$ (in this case the one-sided hypothesis is rejected, whereas the two-sided hypothesis is not). So we conclude that, in general, coverage intervals based on the t distribution with $m_0 - 1$ degrees of freedom give estimated coverage probabilities that are not significantly lower than their nominal value $1 - \alpha$.

When the confidence interval for β_q uses the normal distribution (which gives tighter intervals), then the estimated coverage probabilities are significantly smaller than their nominal values, whatever α value the simulationist uses (1, 5, or 10 %); that is, in most cases we reject H_0. For a 'large' pilot sample ($m_0 = 25$) the t statistic with $m_0 - 1$ degrees of freedom approaches the standard normal distribution. Indeed for those cases the estimated coverage probabilities are often not significantly smaller than the nominal values.

For the sequential procedure we get similar results. In other words, confidence intervals based on the standard normal distribution give too tight confidence intervals, in

general. Exceptions are cases with large pilot-samples (then there is not much difference between the standard normal and the t distribution with m_0 - 1 degrees of freedom). The t distribution with m_0 - 1 degrees of freedom gives confidence intervals with acceptable estimated coverage probabilities, except in two of the 91 situations (27 Monte Carlo cases, each evaluated for three α values). Actually these two situations count as one situation, since they concern the same case with H_0 tested for α is 0.05 and 0.10.

Because both the two-stage and the sequential procedure give acceptable coverage probabilities (for the confidence intervals for the regression parameters β_q), we next compare their statistical efficiency measured through the expected number of simulation responses $\underline{y}_i$. So we observe $\hat{m}_1$ for the two-stage procedure and $\hat{m}_{\pi}$ for the sequential procedure. We do this for all 27 cases, and repeat this 300 times. So for each case we obtain the average number of simulation responses. In 20 of the 27 cases, two-stage sampling requires more simulation responses, which according to the sign test is statistically significant at a type I error rate of 0.05 or higher.

6. Conclusions and Further Research

We examined two procedures for determining the number of simulation responses (replications) per combination of simulation inputs, such that the variances of the average responses become 'approximately' constant. The first procedure is a two-stage rule; the second one is sequential. We use Weighted Least Squares estimators to estimate the Q regression parameters (input effects) β_q (with q = 1,...,Q) from the average simulation responses at the end of the sample size selection rule. These regression estimators are unbiased. Confidence intervals for β_q are too tight, when we use the standard normal variable. These intervals have acceptable coverage probabilities when we use the Student statistic with degrees of freedom equal to m_0-1 where m_0 denotes the pilot sample size. These conclusions hold for both the two-stage and the sequential procedure. The two-stage procedure is easier to understand and easier to program and implement. However, the sequential rule requires fewer simulation responses.

For the sequential approach we introduced a novelty: we use all available simulation responses to estimate the response variances, but (to realize sample-size ratios required for constant response variances) we use fewer responses to estimate the mean simulation

response per input combination.

A refinement of this idea, which needs more research, is as follows. If the estimated response variances differ greatly, then the simulated system with the maximum estimated variance may require too many replications, in practice. But suppose that the maximum divided by the minimum estimated variance is $\underline{r}$. Then for the system with the smallest variance estimate we compute the response average from a single response ($\hat{m} = 1$); to estimate its variance, however, we use all responses for that system. The system with the largest estimated variance is simulated $\underline{r}$ times.

The sample-size rule of (2.1) is not necessarily 'optimal'. Several optimality criteria are popular in the theory of optimal design; see Kleijnen (1987, p. 335). These criteria are usually not applied to sequential designs. For example, a closely related rule replaces the variances σ_i^2 in (2.1) by the standard errors σ_i. If the sample size were fixed and there would be only two populations (n=2), then this alternative rule minimizes the variance of $\bar{Y}_1 - \bar{Y}_2$; see Kleijnen (1987,p. 51). However, the advantage of the rule in (2.1) is that it simplifies the regression analysis of the simulation data, as we saw. Nevertheless we might use the standard deviations (not the variances) of the simulation responses to select the sample sizes. Then the variances will not be constant. There are two alternatives: either apply WLS (which does not reduce to OLS applied to the averages) or rely on the robustness of OLS since the variance heterogeneity will be reduced through the sample sizes selected.

Welch (1990, p. 394) proposes to select the sample sizes such that the confidence intervals for β_u will be of fixed widths.

In future research we can further explore the robustness of the two rules with respect to nonnormality. Further, MacNair and Welch (1991,p. 827) investigate sequential procedures that include testing the fit of the metamodel.

We hope that our paper will stimulate others to further explore the various heuristic procedures to select the number of simulation responses per combination of simulation inputs.

Appendix 1: Proof of (3.8)

For the proof of (3.8) we rearrange the sum of squares and introduce the symbols

$\bar{Y}_i(*) = (\hat{m}_i - m_0)^{-1} \sum_{i=m_0+1}^{\hat{m}_i} Y_{iu}$ and $S_i^2 = S_i^2(m_0)$:

$$\sum_{j=1}^{\underline{\hat{m}}_i} (\underline{y}_{ij} - \bar{\underline{y}}_i(\underline{\hat{m}}_i))^2 = \sum_{j=1}^{m_0} (\underline{y}_{ij} - \bar{\underline{y}}_i(m_0) + \bar{\underline{y}}_i(m_0) - \bar{\underline{y}}_i(\underline{\hat{m}}_i))^2 + \sum_{j=m_0+1}^{\underline{\hat{m}}_i} (\underline{y}_{ij} - \bar{\underline{y}}_i(*) + \bar{\underline{y}}_i(*) - \bar{\underline{y}}_i(\underline{\hat{m}}_i))^2$$

$$= \sum_{j=1}^{m_0} (\underline{y}_{ij} - \bar{\underline{y}}_i(m_0))^2 + m_0(\bar{\underline{y}}_i(m_0) - \frac{m_0}{\underline{\hat{m}}_i} \bar{\underline{y}}_i(\underline{m}_0) - \frac{\underline{\hat{m}}_i - m_0}{\underline{\hat{m}}_i} \bar{\underline{y}}_i(*))^2$$

$$+ \sum_{j=m_0+1}^{\underline{\hat{m}}_i} (\underline{y}_{ij} - \bar{\underline{y}}_i(*))^2 + (\underline{\hat{m}}_i - m_0)(\bar{\underline{y}}_i(*) - \frac{m_0}{\underline{\hat{m}}_i} \bar{\underline{y}}_i(\underline{\hat{m}}_i) - \frac{\underline{\hat{m}}_i - m_0}{\underline{\hat{m}}_i} \bar{\underline{y}}_i(*))^2$$

$$= (m_0 - 1)\, \underline{s}_i^2 + \sum_{j=m_0+1}^{\underline{\hat{m}}_i} (\underline{y}_{ij} - \bar{\underline{y}}_i(*))^2 + \frac{m_0(\underline{\hat{m}}_i - m_0)}{\underline{\hat{m}}_i} (\bar{\underline{y}}_i(m_0) - \bar{\underline{y}}_i(*))^2 .$$

Assume all $\underline{s}_{i'}^2$, $i' = 1,\ldots, n$ are fixed, so the $\underline{\hat{m}}_{i'}$ are fixed. The expectation of $\sum_{j=1}^{\underline{\hat{m}}_i} (\underline{y}_{ij} - \bar{\underline{y}}_i(\underline{\hat{m}}_i))^2$ conditioned on $\underline{s}_{i'}^2(m_0) = s_{i'}^2(m_0)$ is

$$E\{\sum_{j=1}^{\underline{\hat{m}}_i} (\underline{y}_{ij} - \bar{\underline{y}}_i(\underline{\hat{m}}_i))^2 \mid \underline{s}_{i'}^2(m_0) = s_{i'}^2(m_0),\ i' = 1,\ldots,n\} = (m_0 - 1)\, s_i^2 + (\hat{m}_i - m_0 - 1)\, \sigma_i^2$$

$$+ \frac{m_0(\hat{m}_i - m_0)}{\hat{m}_i} (\frac{1}{m_0} s_i^2 + \frac{1}{\hat{m}_i - m_0} \sigma_i^2)$$

$$= \frac{m_0(\hat{m}_i - 1)}{\hat{m}_i} s_i^2 + \frac{(\hat{m}_i - m_0)(\hat{m}_i - 1)}{\hat{m}_i} \sigma_i^2 .$$

So we get

$$E(\underline{s}_i^2(\underline{\hat{m}}_i)) = \frac{m_0}{\hat{m}_i} s_i^2 + \frac{(\hat{m}_i - m_0)}{\hat{m}_i} \sigma_i^2$$

$$= s_i^2 + \frac{(\hat{m}_i - m_0)}{\hat{m}_i} (\sigma_i^2 - s_i^2) \quad \text{QED.}$$

Appendix 2: Monte Carlo Input

In this appendix we add some details that have not already been mentioned in the main text. Most values are taken from Kleijnen et al. (1985) who took the vector β in Case 1 from the case study in Kleijnen, Van den Burg and Van der Ham (1979).

Case 1: $\beta' = (-1.42, -0.769, 13.4, -11.508, 3.5, -1.375, 140.918,$
$15.391, 0.046, 281.098, 21.25, 11.875, -49.483).$

H = 0: σ_i^2 = 1 for all i.

H = 4: σ_i^2 = (4, 4.5, 5, 6, 8, 9, 10, 11, 12, 13.5, 14, 14.5, 16, 18, 19.5, 20).

H = 11.84: σ_i^2 = (1, 2, 3, 4, 4.5, 5, 6, 7, 7.5, 8, 9, 9.5, 10, 11, 12, 12.84).

Case 2: β' = (-1.42, -0.769, 13.44, -11.508).

H = 0: σ_i^2 = 1 for all i.

H = 4: σ_i^2 = (4, 6, 8, 10, 12, 14, 16, 20).

H = 10.83: σ_i^2 = (1, 2, 4, 5, 6, 7, 9, 11.83).

Case 3: β' = (1, 1, 1).

H = 0: σ_i^2 = 1 for all i.

H = 4: σ_i^2 = (4, 10, 16, 20).

H = 10.38: σ_i^2 = (1, 4, 8, 11.38).

Acknowledgement

The Monte Carlo experiment was programmed and executed by Richard Veltman, student in the School of Management and Economics of Tilburg University. Valuable comments on an earlier version were made by our colleague Bert Bettonvil.

References

Dudewicz, E.J. and S.N. Mishra, *Modern mathematical statistics*. John Wiley & Sons, New York, 1988.

Ghosh, B.K. and P.K. Sen, editors, *Handbook of sequential analysis*. Marcel Dekker, Inc. New York, 1991.

Govindarajulu, Z., *The sequential statistical analysis*. American Sciences Press, Columbus (Ohio), 1981.

Kleijnen, J.P.C., *Statistical techniques in simulation*, Volumes I and II. Marcel Dekker, Inc., New York, 1974/1975. (Russian translation, Publishing House "Statistics", Moscow, 1978.)

Kleijnen, J.P.C., *Statistical tools for simulation practitioners*. Marcel Dekker, Inc., New York, 1987.

Kleijnen, J.P.C., P. Cremers and F. Van Belle, The power of weighted and ordinary least squares with estimated unequal variances in experimental design. *Communications in Statistics, Simulation and Computation, B* 14, no. 1, 1985, pp. 85-102.

Kleijnen, J.P.C., A.J. Van Den Burg and R.T. Van Der Ham, Generalization of simulation

results: practicality of statistical methods. *European Journal of Operational Research*, 3, 1979, pp. 50-64.

Kleijnen, J.P.C. and W. van Groenendaal, *Simulation: a statistical perspective*. John Wiley , Chichester (England), 1992.

MacNair, E.A. and P.D. Welch, *Toward a higher level output analysis interface*. Proceedings of the 1991 Winter Simulation Conference, W.D. Kelton and G.M. Clark, pp. 822-831.

Mathai, A. and G. Pederzoli, *Characterizations of the normal probability Law*. Wily Eastern Limited, New Delhi (India), 1977.

Neter, J., W. Wasserman and M.H. Kutner, *Applied linear statistical models: regression, analysis of variance, and experimental design*. Homewood, IL: R.D. Irwin, 1985.

Sargent Robert G., Research Issues in Metamodeling. *Proceedings of 1991 Winter Simulation Conference*. Edited by B.L. Nelson, W.D. Kelton, and G.M. Clark Phoenix, AZ, pp. 888-893

Scheffé, H., Practical solutions of the Behrens-Fisher problem. *Journal American Statistical Association*, 65, no. 332, Dec. 1970, p. 1501-1508.

Welch, P.D., *Simulation and regression: dealing with the assumption of a common error variance*. Proceedings of the 1990 Winter Simulation Conference, edited by O. Balci R.P. Asdowski, and R.E. Nance, pp. 392-394.

Wetherill, G.B., *Sequential methods in statistics*. Methuen & Co., Ltd., London, and John Wiley & Sons, Inc., New York 1966. Revised third edition, Chapman and Hall Ltd, London, 1986.

Whitt, W., *Planning queueing simulations*. Management Science, 35, no. 11, Nov. 1989, pp. 1341-1366.

TABLE 1

Example Data for Sequential Sampling

Stage t	Comb. i	Available m	Required $\hat{m}$	Mean $\bar{y}$	Variance s^2
2	1	25	25	-0.782	1.165
	2	26	64	0.945	2.967
	3	26	155	1.661	7.228
	4	26	133	3.147	6.201
3	1	25	25	-0.782	1.165
	2	27	66	0.857	3.066
	3	27	153	1.579	7.132
	4	27	144	2.980	6.710
...	...	...	...	...	...
65	1	25	25	-0.782	1.165
	2	89	90	1.014	4.203
	3	89	126	1.005	5.892
	4	89	180	2.927	8.380
66	1	25	25	-0.782	1.165
	2	90	89	1.012	4.156
	3	90	125	1.007	5.826
	4	90	178	2.912	8.304
...	...	...	...	...	...
220	1	25	25	-0.782	1.165
	2	90	89	1.012	4.156
	3	143	142	0.945	6.623
	4	244	246	2.684	11.446
221	1	25	25	-0.782	1.165
	2	90	89	1.012	4.156
	3	143	142	0.945	6.623
	4	245	245	2.679	11.405

IN 1991 REEDS VERSCHENEN

466 Prof.Dr. Th.C.M.J. van de Klundert - Prof.Dr. A.B.T.M. van Schaik
Economische groei in Nederland in een internationaal perspectief

467 Dr. Sylvester C.W. Eijffinger
The convergence of monetary policy - Germany and France as an example

468 E. Nijssen
Strategisch gedrag, planning en prestatie. Een inductieve studie binnen de computerbranche

469 Anne van den Nouweland, Peter Borm, Guillermo Owen and Stef Tijs
Cost allocation and communication

470 Drs. J. Grazell en Drs. C.H. Veld
Motieven voor de uitgifte van converteerbare obligatieleningen en warrant-obligatieleningen: een agency-theoretische benadering

471 P.C. van Batenburg, J. Kriens, W.M. Lammerts van Bueren and R.H. Veenstra
Audit Assurance Model and Bayesian Discovery Sampling

472 Marcel Kerkhofs
Identification and Estimation of Household Production Models

473 Robert P. Gilles, Guillermo Owen, René van den Brink
Games with Permission Structures: The Conjunctive Approach

474 Jack P.C. Kleijnen
Sensitivity Analysis of Simulation Experiments: Tutorial on Regression Analysis and Statistical Design

475 C.P.M. van Hoesel
An $O(nlogn)$ algorithm for the two-machine flow shop problem with controllable machine speeds

476 Stephan G. Vanneste
A Markov Model for Opportunity Maintenance

477 F.A. van der Duyn Schouten, M.J.G. van Eijs, R.M.J. Heuts
Coordinated replenishment systems with discount opportunities

478 A. van den Nouweland, J. Potters, S. Tijs and J. Zarzuelo
Cores and related solution concepts for multi-choice games

479 Drs. C.H. Veld
Warrant pricing: a review of theoretical and empirical research

480 E. Nijssen
De Miles and Snow-typologie: Een exploratieve studie in de meubelbranche

481 Harry G. Barkema
Are managers indeed motivated by their bonuses?

482 Jacob C. Engwerda, André C.M. Ran, Arie L. Rijkeboer
Necessary and sufficient conditions for the existence of a positive definite solution of the matrix equation $X + A^T X^{-1} A = I$

483 Peter M. Kort
A dynamic model of the firm with uncertain earnings and adjustment costs

484 Raymond H.J.M. Gradus, Peter M. Kort
Optimal taxation on profit and pollution within a macroeconomic framework

485 René van den Brink, Robert P. Gilles
Axiomatizations of the Conjunctive Permission Value for Games with Permission Structures

486 A.E. Brouwer & W.H. Haemers
The Gewirtz graph - an exercise in the theory of graph spectra

487 Pim Adang, Bertrand Melenberg
Intratemporal uncertainty in the multi-good life cycle consumption model: motivation and application

488 J.H.J. Roemen
The long term elasticity of the milk supply with respect to the milk price in the Netherlands in the period 1969-1984

489 Herbert Hamers
The Shapley-Entrance Game

490 Rezaul Kabir and Theo Vermaelen
Insider trading restrictions and the stock market

491 Piet A. Verheyen
The economic explanation of the jump of the co-state variable

492 Drs. F.L.J.W. Manders en Dr. J.A.C. de Haan
De organisatorische aspecten bij systeemontwikkeling
een beschouwing op besturing en verandering

493 Paul C. van Batenburg and J. Kriens
Applications of statistical methods and techniques to auditing and accounting

494 Ruud T. Frambach
The diffusion of innovations: the influence of supply-side factors

495 J.H.J. Roemen
A decision rule for the (des)investments in the dairy cow stock

496 Hans Kremers and Dolf Talman
An SLSPP-algorithm to compute an equilibrium in an economy with linear production technologies

497 L.W.G. Strijbosch and R.M.J. Heuts
Investigating several alternatives for estimating the compound lead time demand in an (s,Q) inventory model

498 Bert Bettonvil and Jack P.C. Kleijnen
Identifying the important factors in simulation models with many factors

499 Drs. H.C.A. Roest, Drs. F.L. Tijssen
Beheersing van het kwaliteitsperceptieproces bij diensten door middel van keurmerken

500 B.B. van der Genugten
Density of the F-statistic in the linear model with arbitrarily normal distributed errors

501 Harry Barkema and Sytse Douma
The direction, mode and location of corporate expansions

502 Gert Nieuwenhuis
Bridging the gap between a stationary point process and its Palm distribution

503 Chris Veld
Motives for the use of equity-warrants by Dutch companies

504 Pieter K. Jagersma
Een etiologie van horizontale internationale ondernemingsexpansie

505 B. Kaper
On M-functions and their application to input-output models

506 A.B.T.M. van Schaik
Produktiviteit en Arbeidsparticipatie

507 Peter Borm, Anne van den Nouweland and Stef Tijs
Cooperation and communication restrictions: a survey

508 Willy Spanjers, Robert P. Gilles, Pieter H.M. Ruys
Hierarchical trade and downstream information

509 Martijn P. Tummers
The Effect of Systematic Misperception of Income on the Subjective Poverty Line

510 A.G. de Kok
Basics of Inventory Management: Part 1
Renewal theoretic background

511 J.P.C. Blanc, F.A. van der Duyn Schouten, B. Pourbabai
Optimizing flow rates in a queueing network with side constraints

512 R. Peeters
On Coloring j-Unit Sphere Graphs

513 Drs. J. Dagevos, Drs. L. Oerlemans, Dr. F. Boekema
Regional economic policy, economic technological innovation and networks

514 Erwin van der Krabben
Het functioneren van stedelijke onroerend-goed-markten in Nederland - een theoretisch kader

515 Drs. E. Schaling
European central bank independence and inflation persistence

516 Peter M. Kort
Optimal abatement policies within a stochastic dynamic model of the firm

517 Pim Adang
Expenditure versus consumption in the multi-good life cycle consumption model

518 Pim Adang
Large, infrequent consumption in the multi-good life cycle consumption model

519 Raymond Gradus, Sjak Smulders
Pollution and Endogenous Growth

520 Raymond Gradus en Hugo Keuzenkamp
Arbeidsongeschiktheid, subjectief ziektegevoel en collectief belang

521 A.G. de Kok
Basics of inventory management: Part 2
The (R,S)-model

522 A.G. de Kok
Basics of inventory management: Part 3
The (b,Q)-model

523 A.G. de Kok
Basics of inventory management: Part 4
The (s,S)-model

524 A.G. de Kok
Basics of inventory management: Part 5
The (R,b,Q)-model

525 A.G. de Kok
Basics of inventory management: Part 6
The (R,s,S)-model

526 Rob de Groof and Martin van Tuijl
Financial integration and fiscal policy in interdependent two-sector economies with real and nominal wage rigidity

527 A.G.M. van Eijs, M.J.G. van Eijs, R.M.J. Heuts
Gecoördineerde bestelsystemen
een management-georiënteerde benadering

528 M.J.G. van Eijs
Multi-item inventory systems with joint ordering and transportation decisions

529 Stephan G. Vanneste
Maintenance optimization of a production system with buffercapacity

530 Michel R.R. van Bremen, Jeroen C.G. Zijlstra
Het stochastische variantie optiewaarderingsmodel

531 Willy Spanjers
Arbitrage and Walrasian Equilibrium in Economies with Limited Information

IN 1992 REEDS VERSCHENEN

532 F.G. van den Heuvel en M.R.M. Turlings
Privatisering van arbeidsongeschiktheidsregelingen
Refereed by Prof.Dr. H. Verbon

533 J.C. Engwerda, L.G. van Willigenburg
LQ-control of sampled continuous-time systems
Refereed by Prof.dr. J.M. Schumacher

534 J.C. Engwerda, A.C.M. Ran & A.L. Rijkeboer
Necessary and sufficient conditions for the existence of a positive definite solution of the matrix equation $X + A^*X^{-1}A = Q$.
Refereed by Prof.dr. J.M. Schumacher

535 Jacob C. Engwerda
The indefinite LQ-problem: the finite planning horizon case
Refereed by Prof.dr. J.M. Schumacher

536 Gert-Jan Otten, Peter Borm, Ton Storcken, Stef Tijs
Effectivity functions and associated claim game correspondences
Refereed by Prof.dr. P.H.M. Ruys

537 Jack P.C. Kleijnen, Gustav A. Alink
Validation of simulation models: mine-hunting case-study
Refereed by Prof.dr.ir. C.A.T. Takkenberg

538 V. Feltkamp and A. van den Nouweland
Controlled Communication Networks
Refereed by Prof.dr. S.H. Tijs

539 A. van Schaik
Productivity, Labour Force Participation and the Solow Growth Model
Refereed by Prof.dr. Th.C.M.J. van de Klundert

540 J.J.G. Lemmen and S.C.W. Eijffinger
The Degree of Financial Integration in the European Community
Refereed by Prof.dr. A.B.T.M. van Schaik

541 J. Bell, P.K. Jagersma
Internationale Joint Ventures
Refereed by Prof.dr. H.G. Barkema

542 Jack P.C. Kleijnen
Verification and validation of simulation models
Refereed by Prof.dr.ir. C.A.T. Takkenberg

543 Gert Nieuwenhuis
Uniform Approximations of the Stationary and Palm Distributions of Marked Point Processes
Refereed by Prof.dr. B.B. van der Genugten

544 R. Heuts, P. Nederstigt, W. Roebroek, W. Selen
Multi-Product Cycling with Packaging in the Process Industry
Refereed by Prof.dr. F.A. van der Duyn Schouten

545 J.C. Engwerda
Calculation of an approximate solution of the infinite time-varying LQ-problem
Refereed by Prof.dr. J.M. Schumacher

546 Raymond H.J.M. Gradus and Peter M. Kort
On time-inconsistency and pollution control: a macroeconomic approach
Refereed by Prof.dr. A.J. de Zeeuw

547 Drs. Dolph Cantrijn en Dr. Rezaul Kabir
De Invloed van de Invoering van Preferente Beschermingsaandelen op Aandelenkoersen van Nederlandse Beursgenoteerde Ondernemingen
Refereed by Prof.dr. P.W. Moerland

548 Sylvester Eijffinger and Eric Schaling
Central bank independence: criteria and indices
Refereed by Prof.dr. J.J. Sijben

549 Drs. A. Schmeits
Geïntegreerde investerings- en financieringsbeslissingen; Implicaties voor Capital Budgeting
Refereed by Prof.dr. P.W. Moerland

550 Peter M. Kort
Standards versus standards: the effects of different pollution restrictions on the firm's dynamic investment policy
Refereed by Prof.dr. F.A. van der Duyn Schouten

551 Niels G. Noorderhaven, Bart Nooteboom and Johannes Berger
Temporal, cognitive and behavioral dimensions of transaction costs; to an understanding of hybrid vertical inter-firm relations
Refereed by Prof.dr. S.W. Douma

552 Ton Storcken and Harrie de Swart
Towards an axiomatization of orderings
Refereed by Prof.dr. P.H.M. Ruys

553 J.H.J. Roemen
The derivation of a long term milk supply model from an optimization model
Refereed by Prof.dr. F.A. van der Duyn Schouten

554 Geert J. Almekinders and Sylvester C.W. Eijffinger
Daily Bundesbank and Federal Reserve Intervention and the Conditional Variance Tale in DM/$-Returns
Refereed by Prof.dr. A.B.T.M. van Schaik

555 Dr. M. Hetebrij, Drs. B.F.L. Jonker, Prof.dr. W.H.J. de Freytas
"Tussen achterstand en voorsprong" de scholings- en personeelsvoorzieningsproblematiek van bedrijven in de procesindustrie
Refereed by Prof.dr. Th.M.M. Verhallen

556 Ton Geerts
Regularity and singularity in linear-quadratic control subject to implicit continuous-time systems
Communicated by Prof.dr. J. Schumacher

557 Ton Geerts
Invariant subspaces and invertibility properties for singular systems: the general case
Communicated by Prof.dr. J. Schumacher

558 Ton Geerts
Solvability conditions, consistency and weak consistency for linear differential-algebraic equations and time-invariant singular systems: the general case
Communicated by Prof.dr. J. Schumacher

559 C. Fricker and M.R. Jaïbi
Monotonicity and stability of periodic polling models
Communicated by Prof.dr.ir. O.J. Boxma

560 Ton Geerts
Free end-point linear-quadratic control subject to implicit continuous-time systems: necessary and sufficient conditions for solvability
Communicated by Prof.dr. J. Schumacher

561 Paul G.H. Mulder and Anton L. Hempenius
Expected Utility of Life Time in the Presence of a Chronic Noncommunicable Disease State
Communicated by Prof.dr. B.B. van der Genugten

562 Jan van der Leeuw
The covariance matrix of ARMA-errors in closed form
Communicated by Dr. H.H. Tigelaar

563 J.P.C. Blanc and R.D. van der Mei
Optimization of polling systems with Bernoulli schedules
Communicated by Prof.dr.ir. O.J. Boxma

564 B.B. van der Genugten
Density of the least squares estimator in the multivariate linear model with arbitrarily normal variables
Communicated by Prof.dr. M.H.C. Paardekooper

565 René van den Brink, Robert P. Gilles
Measuring Domination in Directed Graphs
Communicated by Prof.dr. P.H.M. Ruys

566 Harry G. Barkema
The significance of work incentives from bonuses: some new evidence
Communicated by Dr. Th.E. Nijman

567 Rob de Groof and Martin van Tuijl
Commercial integration and fiscal policy in interdependent, financially integrated two-sector economies with real and nominal wage rigidity.
Communicated by Prof.dr. A.L. Bovenberg

568 F.A. van der Duyn Schouten, M.J.G. van Eijs, R.M.J. Heuts
The value of information in a fixed order quantity inventory system
Communicated by Prof.dr. A.J.J. Talman

569 E.N. Kertzman
Begrotingsnormering en EMU
Communicated by Prof.dr. J.W. van der Dussen

570 A. van den Elzen, D. Talman
Finding a Nash-equilibrium in noncooperative N-person games by solving a sequence of linear stationary point problems
Communicated by Prof.dr. S.H. Tijs

571 Jack P.C. Kleijnen
Verification and validation of models
Communicated by Prof.dr. F.A. van der Duyn Schouten